Characterization and segmentation of fetal brain structures at the mesoscopic scale using MRI

KTH Thesis Report

Angèle Denis

Authors

Angèle Denis aladenis@kth.se
Engineering Physics
KTH Royal Institute of Technology

Place for Project

Saclay, France
Neurospin, CEA-Saclay

Examiner

Mats Danielsson, Professor, KTH Royal Institute of Technology

Supervisors

Mats Persson, Assistant Professor, KTH Royal Institute of Technology

Lucie Hertz-Pannier, lucie.hertz-pannier@cea.fr

Yann Leprince, yann.leprince@cea.fr

Lucas Arcamone, lucas.arcamone@cea.fr

Abstract

Studying White Matter (WM) lesions in premature babies is a societal challenge, as premature mortality is decreasing while neurological morbidity remains the same. This master thesis is part of the p-HCP (Premature Human Connectome Project), which aims to study these lesions using extreme magnetic field MRI (11.7T) , starting with the construction of a typical ex vivo brain atlas at mesoscopic scale and comparing it, in a second phase, to an atlas of damaged brains. This work presents a segmentation strategy for the fetal brain structures to characterize them at different key stages of the development using MRI images with unprecedented resolution. We first manually segmented the structures of a 20-week gestation brain using a histological atlas, in order to explore the information contained the nineteen MRI modalities available (quantitative, weighted and diffusion images). We then segmented the fetal brain structures, without anatomical a priori, by using data engineering and automatic clustering algorithms. We have succeeded in designing a proof of concept for automatic segmentation for the fetal brain and extracting, from our MRI images, groups of structures similar in molecular composition and cytoarchitecture.

Keywords

Biomedical physics, Neuroimaging, Magnetic Resonance Imaging (MRI), Diffusion MRI, Multimodal, Segmentation, Mahalanobis distance, K-means, Fetus, Human brain development, Microstructure

Abstract

Att studera lesioner den vita substansen hos för tidigt födda barn är en samhällelig utmaning, eftersom den för tidiga dödligheten minskar medan den neurologiska sjukligheten förblir densamma. Denna masteruppsats är en del av p-HCP (Premature Human Connectome Project), som syftar till att studera dessa lesioner med hjälp av MRT med extremt magnetfält (11,7T), med början med konstruktionen av en typisk ex vivo hjärnatlas i mesoskopisk skala och jämföra den, i en andra fas, med en atlas av skadade hjärnor. I detta arbete presenteras en segmenteringsstrategi för fostrets hjärnstrukturer för att karakterisera dem vid olika viktiga utvecklingsstadier med hjälp av MR-bilder med oöverträffad upplösning. Vi segmenterade först manuellt strukturerna i en 20 veckors graviditetshjärna med hjälp av en histologisk atlas, för att utforska informationen i de nitton tillgängliga MR-modaliteterna (kvantitativa, viktade och diffusionsbilder). Vi segmenterade sedan de fetala hjärnstrukturerna, utan anatomisk a priori, genom att använda datateknik och automatiska klusteralgoritmer. Vi har lyckats utforma ett proof of concept för automatisk segmentering av fostrets hjärna och från våra MR-bilder extrahera grupper av strukturer som liknar varandra i molekylär sammansättning och cytoarkitektur.

Nyckelord

Biomedicinsk fysik, Neuroimaging, MRI, Diffusion MRI, Multimodal, Segmentering, Mahalanobisavstånd, K-means, Foster, Hjärnans utveckling, Mikrostruktur

Acknowledgements

Acronyms

CEA   French Atomic Energy and Alternative Energies Commission

MRI   Magnetic Resonance Imaging

p-HCP Premature Human Connectome Project

CODECOH Conservation d'éléments du corps humain

GW   Gestation week

3D   3-dimensional

2D   2-dimensional

NODDI Neurite Orientation Dispersion and Density Imaging

NDI   Neurite Density Index

ODI   Orientation Dispersion Index

DTI   Diffusion tensor imaging

FA   Fractional anisptropy

MD   Mean Diffusivity

WM   White Matter

NMR Nuclear Magnetic Resonance

CSF   Cerebrospinal Fluid

PBS   Phosphate Buffer Saline

KDE   Kernel Density Estimate

PCA   Principal Component Analysis

STD   Strandard Deviation

ADC   Apparent Diffusion Coefficient

GFA   Generalized Fractional Anisotropy

Contents

Chapter 1

Context and framework of the thesis

1.1 Master thesis context

1.1.1 CEA, Paris Saclay

The French Atomic Energy and Alternative Energies Commission (CEA) is a research organisation in many fields: energy, defense and security, information technologies and health technologies. The CEA was created in 1945 for research in both military and civil nuclear energy. Over the years, the CEA has become a major player in research in France, particularly for the healthcare sector. [1]

1.1.2 Neurospin

Neurospin is a research institute in neuroimaging of the CEA, in Saclay, France, that is dedicated to the study of the human brain with imaging techniques such as MRI. Thanks to the development of advanced equipment with ultra high field MRI scanners (three clinical scanners at 3,7 and 11.7T and three pre-clinical scanners at 7,11.7 and 17.2 T), researchers can visualize and analyze the structures, functions and connectivity of the human brain with really high resolution (up to 100 micrometer). Two types of research are carried out at neurospin: the more methodological part, which is at the heart of the research carried out at the CEA, and a coginitive part. The main fields of study in Neurospin include the neurological and psychiatric disorders, cerebral

development, aging, cerebral plasticity and many others. [2]

1.1.3 p-HCP

My master thesis is part of the Premature Human Connectome Project (p-HCP) which gathers clinicians from Lariboisière Hospital (Paris), Robert-Debré hospital (Paris), Sorbonne University (Paris) and Neurospin (CEA, Paris-Saclay).

The main goal of this project is to study the white matter injuries in premature babies and to find new markers of white matter injury using extreme high field ex vivo (post mortem) MR Imaging and 3D histology. The goal is to create a large atlas to describe both macro and microstructure of white matter injury of preterm newborns using MR imaging and 3D histology of injured brains and also pseudo-typical brains. Currently, at Neurospin, an ex vivo multi-modal MRI atlas of typical brains at mesoscopic scale is being created. Thanks to the pre-clinical MRI Bruker at 11.7T, anatomic, diffusion and relaxometry images of brain of different ages with a high resolution of 200 microns isotropic (and 100 microns for some modalities) are collected. In parallel, a 3-dimensional (3D) histology atlas is being built at Lariboisière Hospital on the same samples. The next step will be to create also a database at multiple fields (from 3T to 11.7T) and a 3D histology and immunohistochemical database of injured brains.

1.2 Scientific background

1.2.1 Brain anatomy

The brain is part of the central nervous system. It is made up of two distinguishable compartments : the white matter and the gray matter. The gray matter is the part of the brain containing all the neurons cell bodies such as the cortex (at the surface) and the central gray matter (basal ganglia and thalamus). The white matter is composed of neuronal fibers called axons that make up the connections between the different parts of the brain. The cerebrospinal fluid surrounds the brain and the spinal cord to give a physical and immunity protection.
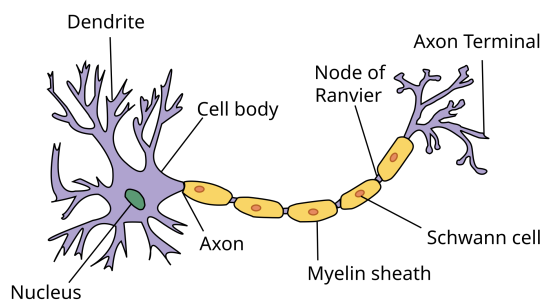
Figure 1.2.1: Neuron structure
Source: User:Dhp1080, CC BY-SA 3.0, http://creativecommons.org/licenses/by-sa/3.0/

A neuron is a cell of the nervous system that sends and receives electrical and chemical signals. The three main parts of a neuron are the dendrites, the cell body and the axon. (see Fig 1.2.1) The dendrites are branched extensions located around the cell body. They receive electrical and chemical signals from other neurons via the synapses and transmit this information to the cell body, and to other neurons through the axon.

In this project, we are looking at fetal brains in development. During fetal development, the most important neurogenetic events occur : proliferation (determination of the number of neurons), molecular specification (molecular diversity of neurons), migration (allocation in the cortex of the neurons moving from the ganglionic eminence to the cortical plate), axonogenesis (growth of axons) and synaptogenesis (formation of functional contacts).Fig 1.2.2
Around 15 Gestation week (GW), myelinization begins. Myelin is located around the axon and acts as an insulator, making it impermeable for efficient protein and electrical transport.
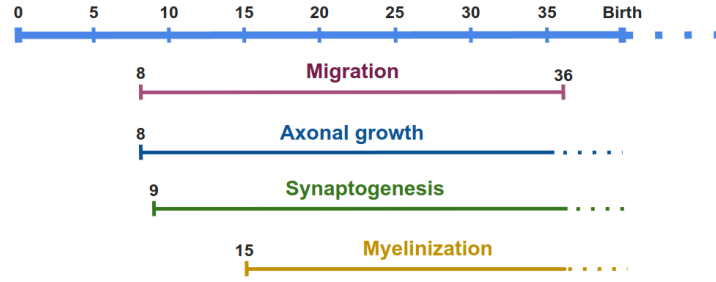
Figure 1.2.2: Development periods of main neurogenetic events

Source: Lucas Arcamone, [3]

During the development, the brain grows dramatically leading to the gyration : gyri (folds on the surface of the brain) appear, separated by sulci (see Fig. 1.2.3).
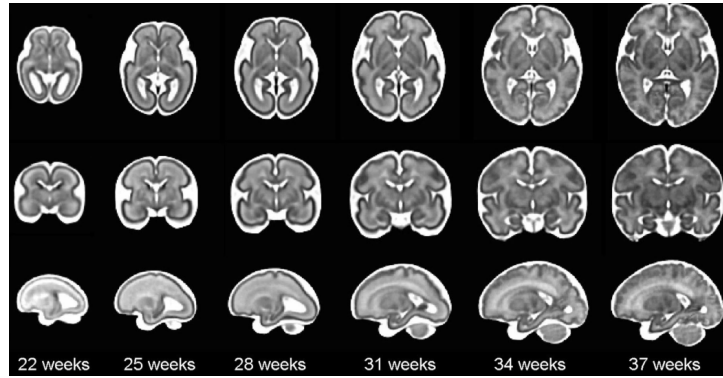
Figure 1.2.3: MRI of fetal brain at different gestational age in coronal, axial and sagittal view

Source: A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth [4]

1.2.2 MRI physics

MRI is a medical imaging technique that allows for 2-dimensional (2D) or 3D images of the body, with good contrast for soft tissues and in a non-invasive manner. MRI is usually used for diagnosis and follow-up of diseases.

MRI is based on Nuclear Magnetic Resonance (NMR) principle, using the quantum properties of the atomic nuclei. Each atom nucleus has several protons and neutrons that rotate around an axis passing through their center. This creates a spin or magnetic moment. In clinical MRI, we observe the magnetic moment of an atom, usually hydrogen, since the brain is composed mainly of water. Some MRI techniques

use other atomic nuclei such as sodium, hyperpolarized helium, phosphor, carbon, etc.

When hydrogen atoms are observed without any external magnetic field, protons are randomly oriented and the sum of all microscopic elementary magnetizations is zero. There is no macroscopic magnetization. When applying a static magnetic field, called $\vec{B}_0$, along the z-axis, all the spins align to the direction of this vector but they are not in the same orientation (up or down). The resultant of the macroscopic magnetization is then non-zero and creates a longitudinal component $\vec{M_z}0$ of the tissue vector of magnetization.

Protons do not really align along $\vec{B}_0$ but rotate around the z-axis at the Larmor frequency : $\omega_0 = \gamma B_0$ where γ is the gyro magnetic ratio, specific to each nucleus. This rotation is called precession. Since this magnetization is along the external magnetic field, the spins' orientation does not change, producing no detectable signal. To make protons "visible", another magnetic field $\vec{B}_1$, rotating at the frequency ω_1, is applied, disrupting the equilibrium. At the resonance condition, the rotation frequency of the rotating magnetic field ω_1 and the Larmor frequency of the static one ω_0 are equal. In this configuration, $\vec{M}$ is still rotating around $\vec{B}_0$ and begins to rotate around $\vec{B}_1$ so a transversal component M_{xy} appears. This state is unstable, once $\vec{B}_1$ is suppressed, the longitudinal magnetic moment $\vec{M_z}$ returns to $\vec{M_{z0}}$ -T1 relaxation- and the transversal component $\vec{M_{xy}}$ decreases quickly to zero -T2 relaxation-. The contrast in MRI images results from the differences in T1 or T2 relaxation times in various tissues.

MRI sequences acquisition parameters can be changed to obtain images of different weightings : the echo time T_E is the time between the application of the resonance frequency and the detection of an echo and the repetition time T_R is the time between two pulses. Using a short TR and TE allows the discrimination of two tissues with regard of T1 whereas long TR and TE are useful for discriminating tissues with regard to T2.

1.2.3 Diffusion

The diffusion MRI is a technique that takes benefit of the movement of water molecules. The diffusion of these molecules can be free; this is the isotropic diffusion and this is described by the Brownian motion. This results from the vibrating motion

of the water molecules, governed mainly by the temperature. The diffusion can also be anisotropic, when water molecules do not diffuse uniformly in all directions but are restricted in a specific direction (by cell membranes, for example).

There is an also diffusion technique: the Diffusion tensor imaging (DTI). The diffusion tensor is a mathematical tool commonly used to characterize the anisotropy of diffusion in anatomical tissues, by modeling the water molecule diffusion in three dimensions, giving information about the orientation and the magnitude of the net diffusion in each voxel. With the tensor we can calculate parameters such as the Fractional anisptropy (FA) and the Mean Diffusivity (MD). The FA evaluates how anisotropic the diffusion in the brain tissues is. A fraction of 0 indicates that the diffusion is totally isotropic and 1 indicates that the diffusion is fully anisotropic (all water molecules diffuse in one direction). MD measures of the average diffusion in all directions. It gives an estimation of the mobility of molecules in a voxel. A high MD is seen in the Cerebrospinal Fluid (CSF), but can also indicate a disruption in the tissue micro structure such as a loss of membrane or tissue degeneration.

The brain is composed of 80% of water molecules : some are in a configuration of isotropic motion as in CSF, and others are restricted as in axons. The myelin around axons prevents water molecules from crossing axons membranes, so they diffuse preferentially in the direction of the axon. Observing the preferred trajectories of water molecules in white matter using diffusion imaging thus provides information on the orientation of nerve fibers locally. Different models and metrics give different information about fibers. For example : the Neurite Orientation Dispersion and Density Imaging (NODDI) model. This model aims to characterize the brain microstructure giving information about the density of neurites –such as axons and dendrites- and about the orientation dispersions of those neurites in the brain. Since this is a model, it is based on simple hypotheses. It is important to note that we do not obtain real information about the neurites but a rough idea of their number and orientation. The main principle of this model is to decompose the diffusion signal in four components that represent the different types of diffusion in the brain : stationary, isotropic and extra and intra axonal diffusion. Different metrics can be computed:

- The Neurite Density Index (NDI) represents the intra axonal diffusion divided by all the other ones except the isotropic one. It indirectly reflects the density of neurons in space.

- The Orientation Dispersion Index (ODI) gives information on the arrangement of neurites in space.

To reconstruct the different modalities and to have different contrasts, it is needed to acquired different sequences using different b-values. The b-value represents the image's sensitivity to the movement of water molecules.

Chapter 2

Introduction

Throughout pregnancy and after the term, the brain develops in size, weight, shape and composition. The brain changes very intensely during the first weeks of gestation and the first postnatal months [5]. This development is based on different complex and intermingled mechanisms that lead to maturation and function specialization of the future grey matter and the white matter. Moreover, brains are made up of transient structures and evolve rapidly week by week. During brain development, neurogenetic events happen : migration, axonal growth, synaptogenesis and myelination [6]. The neuronal migration is a migration of the neurons from the subventricular zones to the cortex or subcortex. The axonal growth creates connections. Synapses, the connections between neurons in the central nervous system, are formed during synaptogenesis. Myelination (i.e. the formation of myelin around axons) takes place at the end of white matter development.

To study the brain development mechanisms at microscopic scale, 3D histology is used to describe the brain physiology with a resolution around 100 nm and 0.1 microns. It is often coupled with clinical MRI at 3T that has a macroscopic resolution around 1 mm [7]. Paralleling MRI images with histology provides a new approach to understanding the events that happen during brain development [8]. There is no study of the fetal brain at mesoscopic scale, around hundreds of microns, to make a bridge between the micro and macroscopic scale. By increasing the magnetic field intensity until 11.7 T to image fetal brains, mesoscopic resolution can be reached, offering detection and characterization of the smallest anatomical regions [9] and detailed visualization of brain fiber bundles [10]. Creating such data for the brains of premature babies is an

excellent approach for studying white matter lesions, resulting from the interruption of normal brain development in the intrauterine environment.

Premature newborns, 13 millions in the world, are a major societal challenge. Despite therapeutic advances that reduce mortality, neurodevelopmental morbidity remains high [11]. The global prevalence of mental impairment is about 13.6 % among surviving preterm infants born at 26 weeks or earlier and about 14.3 % among children weighing less than 800 g [12].

p-HCP is a project that aims to characterize the macro and microstructural white matter injuries [13] on preterm newborns using extrem high field MRI at 11.7T with a resolution of 200 microns with a 3D histological validation. The first part of this project is then to create a MRI atlas of 'pseudo-typical' brains between 20 and 41 GW correlated with a 3D histology atlas. For the moment, 3 brains are totally acquired (20 GW, 29 GW, 33 GW) and by the end of the project another will be included (41 GW) [3]. The methods of acquisition, reconstruction and registration are similar to the Chenonceau project [14]. The reconstructed and registered images must be segmented to characterize the structures and study the development of their size/shape and composition during fetal development.

This master thesis aims to present the strategy to better understand the fetal brain at different stages of development using the advances in MRI histology [15]. The study was done on the 20 GW specimen. First of all, we will use the multimodal MRI to explore the tissue composition and the cytoarchitecture of the brain [16] of fetuses. The quantitative images act as a proxy of molecule contents (water, iron, ...) and the diffusion images explain the neuronal fibers organization [17].

To characterize the tissues inside the brain using MR images, I carried out manual segmentation. The boundary of the structures was manually delineated directly on the MRI modality used for every single transversal slice [18]. The aim of this approach was to study the behavior of the different modalities in the best-known, easily-recognized structures, in order to study their behavior in all available MRI modalities. For this, I used a histological atlas and all MRI images, thus ensuring optimal segmentation of some structures of the 20 GW brain.

There are several automatic methods for brain segmentation, such as intensity-based, threshold-based or atlas-based clustering [19]. The first two methods are linked to the recognition of gray and white matter, which are indistinguishable in the fetal brain.

Indeed, until the end of neuronal migration and myelination, gray matter and white matter are still mixed in certain parts of the brain. Atlas-based segmentation cannot be applied, as there is as yet no such resolute atlas of the fetal brain. Automatic segmentation avoids introducing information that is external to the actual information (human bias) content provided by the MRI modalities. In this way, we can group structures according to their common, physically interpretable feature. We will draw a threshold-based segmentation strategy, using the multi modal MRI and guiding by the previous characterisation of the manually segmented structures.

Chapter 3

Material and Methods

3.1   Sample preparation

The brains that compose the atlas come from the foetopathology department of Robert-Debré Hospital (Paris) (Conservation d'éléments du corps humain (CODECOH), Preservation of human body parts, tissue authorization : DC-2022-5118). Firstly, they are imaged in situ (within the skull) with a clinical 3T MRI at Robert Debré Hospital to check the state of the brain, aid diagnosis and provide a clinical reference for comparing data acquired later. Then they are fixed during two months by immersion in Formalin 12% formaldehyde to avoid the natural decline in tissue quality. Then, at Neurospin, the brains are doped with Gadolinium and Phosphate Buffer Saline (PBS) to improve the contrast and reduce the relaxation times (and then reduce the acquisition time). They are first imaged on the clinical 7T MRI to have a high isotropic resolution reference, used for modeling containers and cutting dies, and for registration of tissue blocks. The 11.7T MRI has a diameter of 60 mm, the antenna used has a diameter of 50 mm and the usable diameter is 46mm. The brains in the project are (with the exception of the 20 GW) larger than this diameter, and must therefore be cut to fit the MRI antenna. After imaging is completed, they are transferred to Lariboisière Hospital where the histology is done. The specimens come from medical terminations of pregnancy or fetal/neonatal deaths, when the cause of death is not related to brain pathology. Samples are considered pseudo-typical because death often leads to brain abnormalities. We will use 3D histology as the ground truth to distinguish typical from non-typical areas. For now, we have collected brains of 20, 29 and 33 GW and a 41 GW brain will be entirely imaged by the end of this project. Those are key stages of fetal

development from middle of pregnancy to full term.

3.2 MRI protocol

Images are acquired on a BioSpec 117/16 Bruker, a preclinical MRI that reaches a field strength of 11.7T. This machine has been chosen because of its extremely high magnetic field: it allows a better signal collection to reduce the spatial resolution. Its gradients are up to 760 mT/m (compare to the clinical MRI gradient at 200mT/m). As the brains are imaged post-mortem, we can use very long acquisition times to obtain very high-resolution images. This also enables us to achieve isotropic resolution of 200 microns, or even 100 microns for certain sequences (T2-weighted and quantitative T2*) with an acquisition time of 150 hours per field of view.

Using different MRI sequences, it is possible to have weighted, quantitative and diffusion images. A weighted image is a contrasted representation based on relaxation time, providing an anatomic image with a good contrast between different brain tissues when the sequence parameters are well chosen. A quantitative image provides a measure of the tissue relaxation time in each voxel, that is correlated with the tissue composition and microstructure. Moreover, because we obtain physical values rather than reconstructed ones, it allows for direct comparison in subsequent studies, enhancing repeatability.

Modalities	Acquisition time
T2-weighted	21h
Quantitatives images (T1, T2*, T2)	26h20min
Diffusion images : DTI, AQBI, NODDI	100h
Total	150h

Table 3.2.1: Acquisition times for different modalities

By the end of the 150 hours of acquisition (Table 3.2.1), 19 images were reconstructed : quantitative T1 (longitudinal relaxation time), quantitative T2 (transversal relaxation time), quantitative T2*(transverse relaxation time, taking into account the effect of B0 inhomogeneities caused by variations in tissue molecular composition and imperfections in the external magnetic field B1 at the microscopic scale), weighted T2, NODDI (NDI and ODI) and images from the DTI model. Four metrics are calculated with the DTI model, each with the three b values $b = 1500mm/s^2$, $b = 4500mm/s^2$,

$b = 8000mm/s^2$) : the Apparent Diffusion Coefficient (ADC), the FA, the parallel diffusion that takes into account only the diffusion along the principal fibers axis, and transverse diffusion that takes into account, on the contrary, diffusion perpendicular to the main fiber. There is finally the Generalized Fractional Anisotropy (GFA).

3.3 Manual Segmentation

The manual segmentation of the present study was carried out on a 20 GW fetal brain, with the visualization software Anatomist [20] and a graphic tablet.
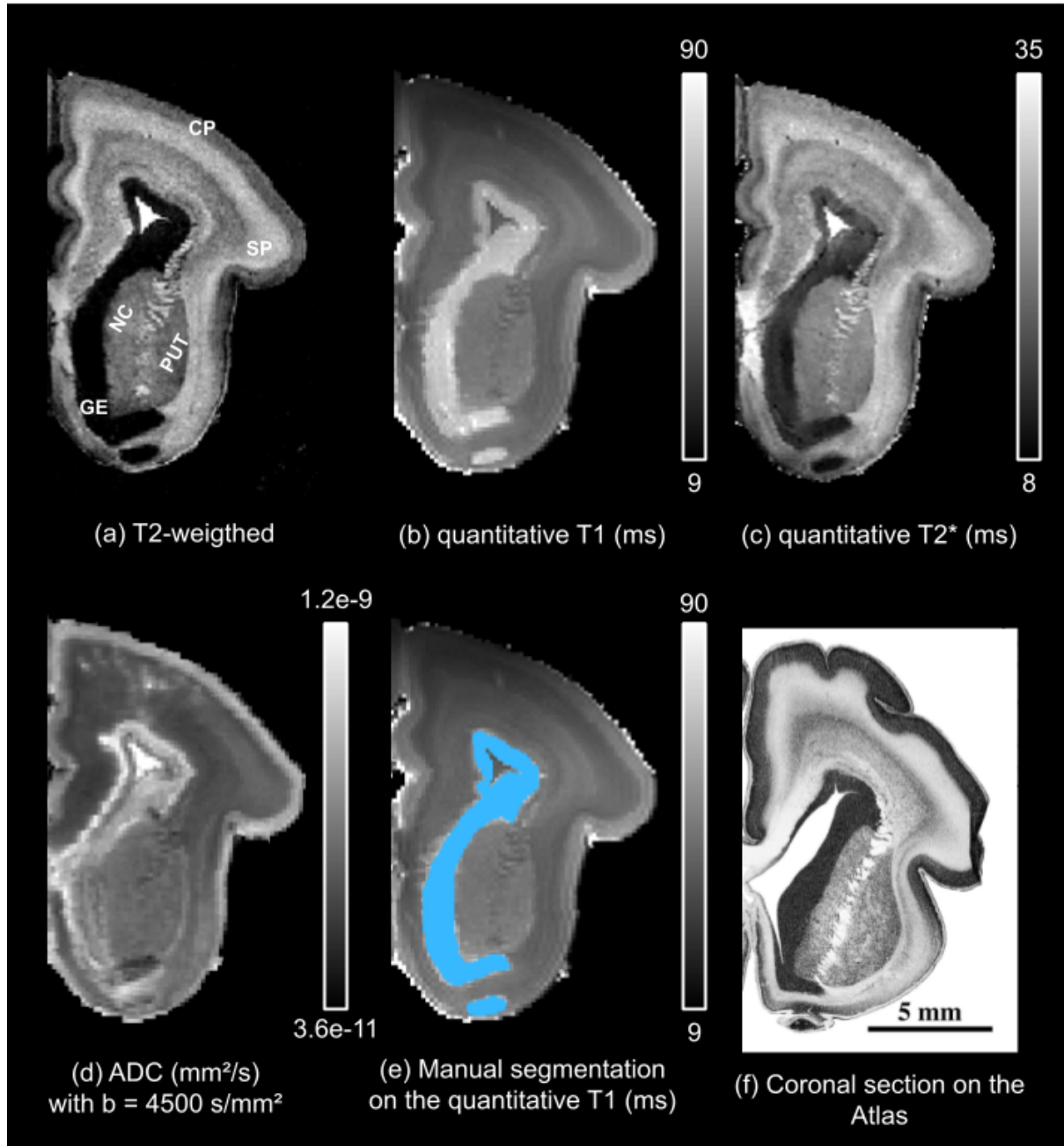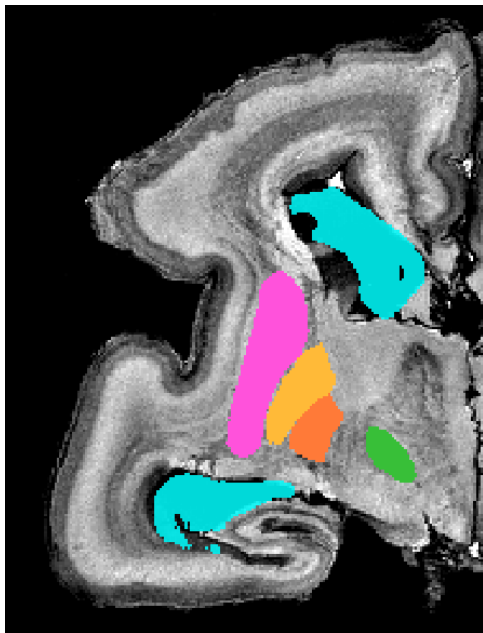
Figure 3.3.1: Manual segmentation of the ganglionic eminence, using different modalities and the Atlas [21]. (GE : ganglionic eminence, PUT : putamen, : NC : caudate nucleus, CP : cortical plate, SP : sub-plate)

To segment the regions of interest, the different structures seen on each modality of 3D magnetic resonance images were compared with those shown on the coronal and sagittal sections of same age specimen atlas [21] (see Fig 3.3.1). The advantage of having different MRI sequences was that they could be compared: some regions, such as the basal ganglia, were much better distinguished on T2-weighted images, while others, such as the globi pallidi, required diffusion images for delimitation. It was

not feasible to segment all the regions: some because they were too large, like the cortical plate (in which case only a small representative part was segmented) and others because it was not possible to delineate them in the available images.

The structures that were segmented manually include (see Fig 3.3.2) :

- The basal ganglia : caudate nuclei (dark blue), lenticular nuclei -putamen (pink), internal (orange) and external (yellow) globus pallidus-, substantia nigra (black), subthalamic nuclei (green)

- The red nuclei (red)

- The ganglionic eminences (blue)

- A small part of the cortical and subcortical plates and of some waves of neuronal migration (not visible on Fig 3.3.2)

(a) Coronal section of some manual segmentation results on T2-weighted

(b) QR codes for moving 3D visualizations

Figure 3.3.2: Results of the manual segmentation

3.4 Characterizaton of Regions of Interest

In order to obtain a "fingerprint" of each segmented structure from the manual segmentation, we aimed to better understand each segmented structure by studying the different quantitative values of relaxation times and the values of the different

diffusion modalities, in order to obtain precise information on their molecular and cellular composition, the disparity of values within the same region, and so on.

3.4.1 Histograms and KDE

To obtain information on the distribution of the data, we then displayed the histograms using the Seaborn python library [22]. An histogram is a graphic representation of the voxel intensities distribution, showing the frequency of each intensity voxel value through the three dimensions. In addition to characterizing the structure, the aim was to identify potential sub-regions within the structures. Indeed, if several peaks can be distinguished within the histogram, it either means that the structure is divisible into adjacent sub-structures, or that the structure is not homogeneous (see example Fig 3.4.1).
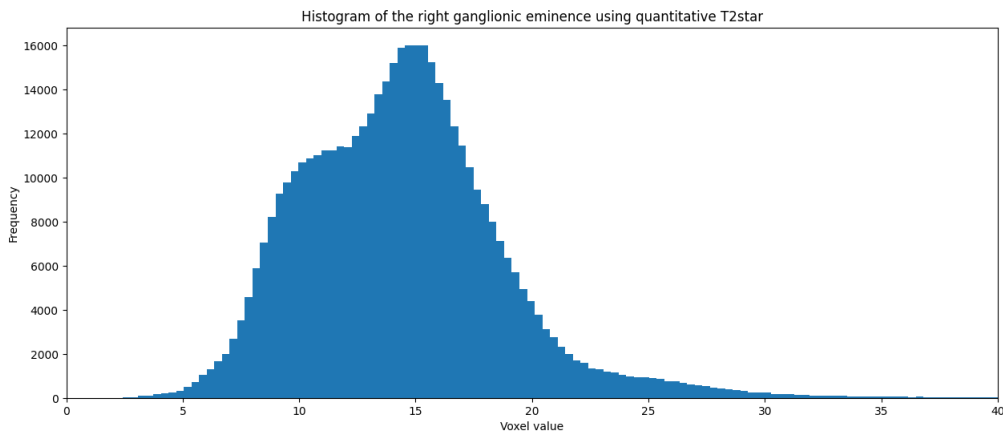

Figure 3.4.1: Histogram example on the right ganglionic eminence in q-T2star (ms)

The Kernel Density Estimate (KDE) plots represent the data using a continuous probability density curve, in order to compare different structures on the same figure for one modality(see example Fig 3.4.2). This will be useful for automatic segmentation, indicating which structure can be isolated by which modality.

3.4.2 Violin plots

The violin plot is a graph that visualizes the distribution of a set of data. It shows both summary statistics such as mean and extreme values, and the probability density of the data distribution (Fig 3.4.3). This thumbnail view was more practical given the number of histograms we needed to display.
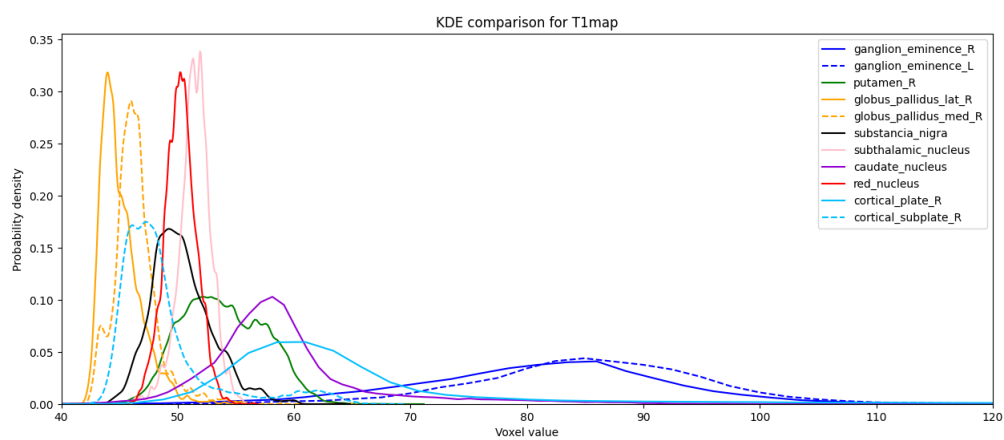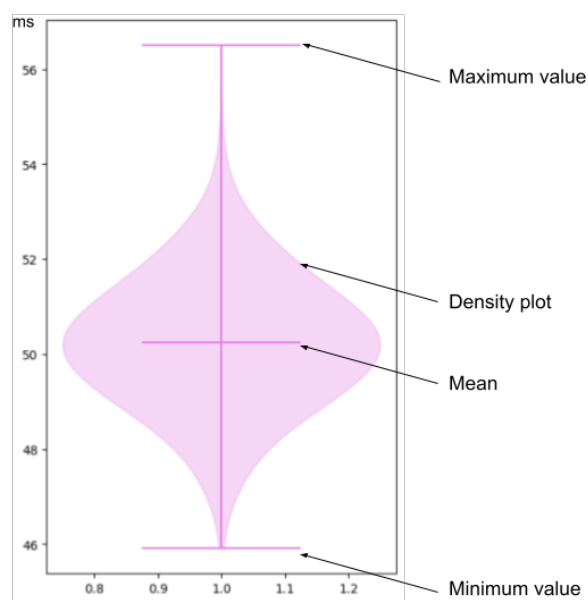
Figure 3.4.2: KDE plots example on q-T1 (ms)

Figure 3.4.3: Violin plot : example of the red nucleus on quantitative T1

3.5 Data analysis

This section presents all the explorations conducted to enhance our knowledge and understanding of this highly complex dataset (never explored before) aiming at optimally reducing the dataset.

3.5.1 Data reduction

Our dataset is composed of 19 images. Each modality gives a different contrast and therefore different information about the tissue. Using all the images makes segmentation more complicated, both in terms of data size, but also in terms of understanding and verification of each contrast. It was therefore essential to reduce the dataset.

The aim of this section is to describe how the database has been transformed from a large number of images to a smaller number, while retaining as much of the original information as possible. We subsequently explored joint histograms, correlations, and Principal Component Analysis (PCA).

Joint histograms

To reduce the data set we first considered plotting joint histograms between two modalities. The idea was that if two modalities showed sufficient colinearity, we could keep only one in our dataset. To prevent extreme values on the voxels due to PBS leftovers and fit problems during reconstruction, the same brain mask was applied to all images. To plot the histograms, python library Seaborn [22] was used. As shown on the figure 3.5.1, three diffusion modalities were essentially linear : the parallel diffusion, the transversal diffusion and the ADC metric. The ADC will be chosen, since it gives an average diffusion coefficient in every directions.

Joint histograms were usefull to reduce the dataset basing on the colinearity but most of the acquired images were not colinear.

Correlation study

To continue data set reduction, the correlation between the modalities were studied with the Pearson coefficient, a parametric correlation coefficient [23].

The correlation matrix was obtained and plotted with Seaborn [22]. (Fig 3.5.2)

(a) ADC against parallel diffusion
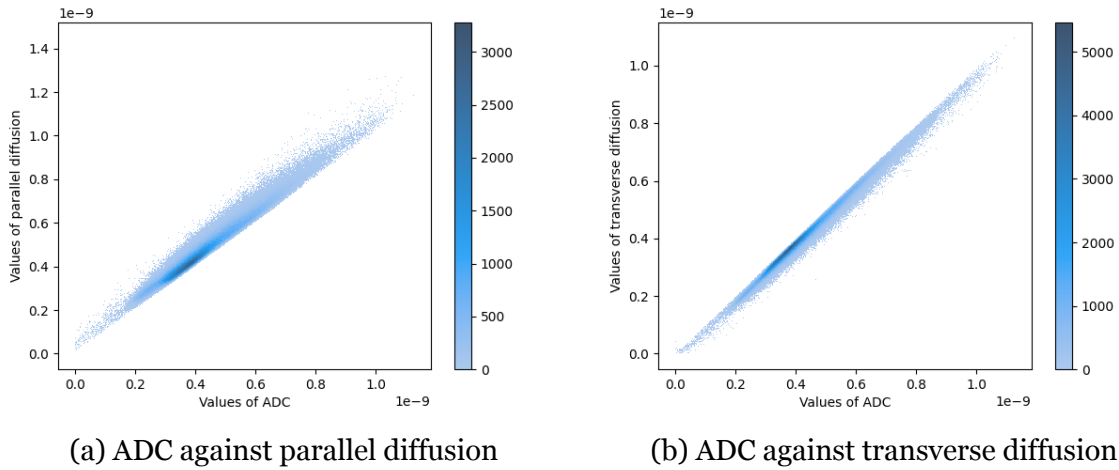
(b) ADC against transverse diffusion

Figure 3.5.1: Joint histograms showing the linearities of the diffusion modalities
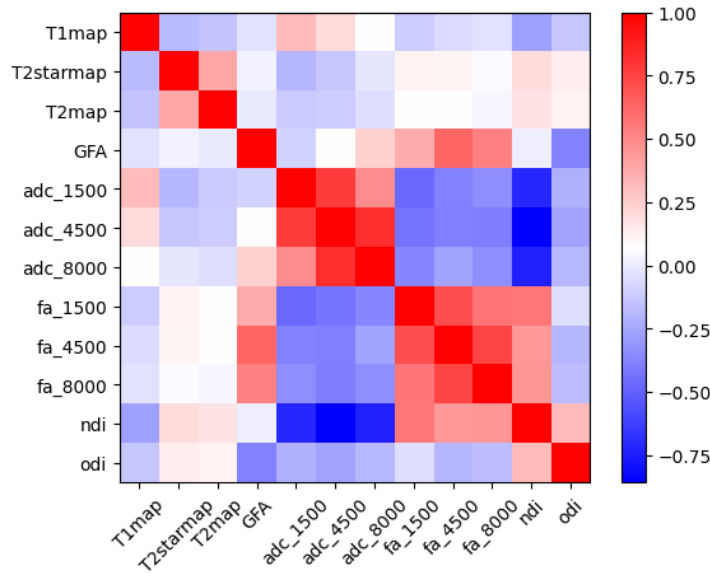
Figure 3.5.2: Pearson correlation matrix for all the modalities

Same modalities with different shells are quite correlated. The decision was taken to only keep the b-value $b = 4500s/mm^2$ (this b-value contains information on both $b = 1500s/mm^2$ and $b = 8000s/mm^2$) . The metric GFA is highly correlated to all the FA modalities. The GFA is more representative of the reality that the DTI because it takes into account all the diffusion orientations. The NDI and ODI were considered enough negatively correlated of respectively ADC and GFA to be sidelined. Finally, the five modalities kept are : quantitative T1, quantitative T2*, quantitative T2, GFA, ADC at $b = 4500s/mm^2$ (Fig 3.5.3).

(a) quantitative T1   (b) quantitative T2*   (c) quantitative T2

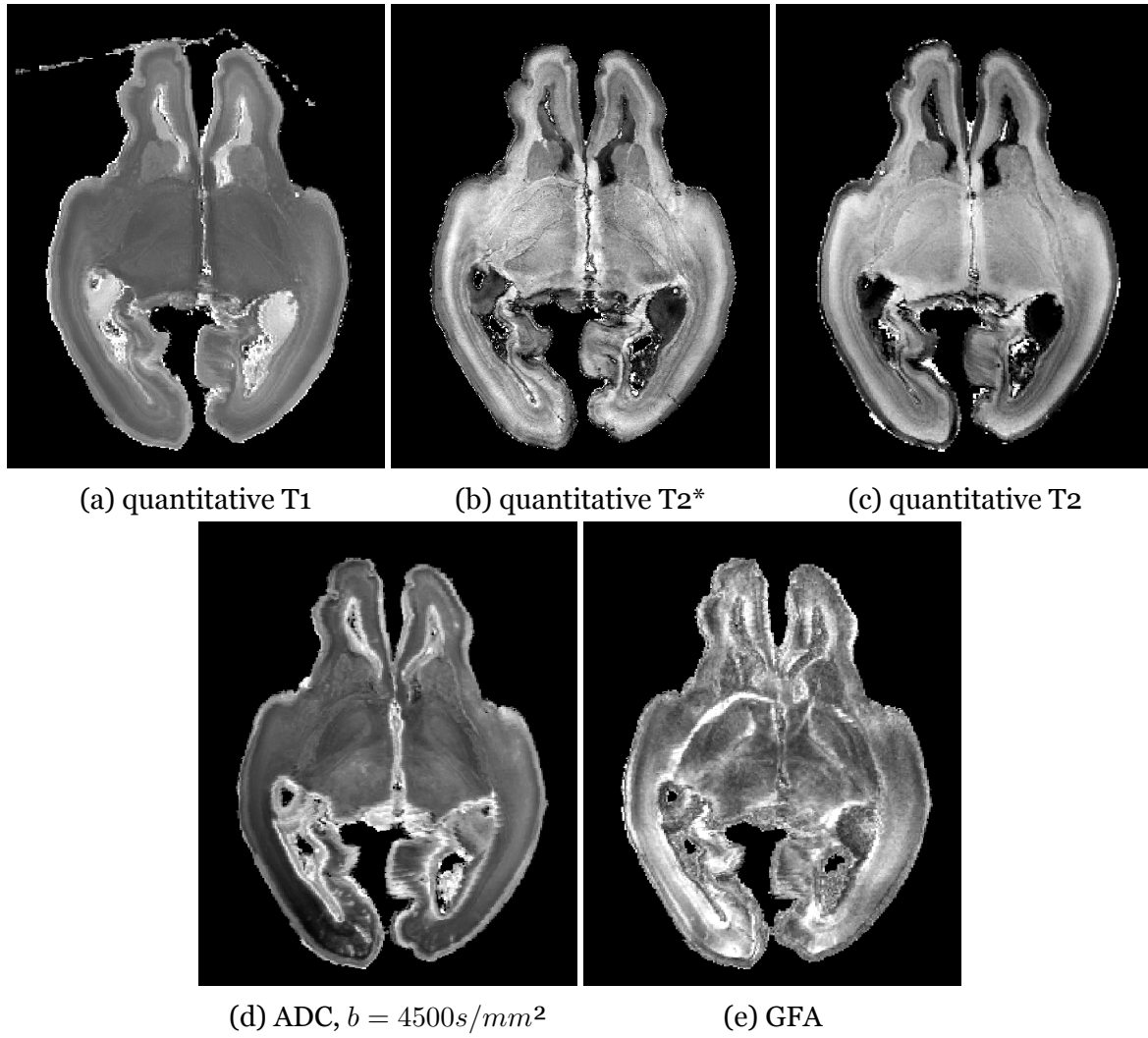(d) ADC, $b = 4500s/mm$²   (e) GFA

Figure 3.5.3: Axial sections of the 20 GW specimen

Principal components analysis

PCA is commonly used for dataset reduction. Its main goal is to create, from correlated variables, a set of uncorrelated ones. Those new variables are called principal components. Firstly, the images need to be centered and reduced to be compared. Then, it calculates the covariance matrix of the dataset and decomposes it into eigenvectors and eigenvalues. The eigenvectors represent the main directions of variation in the data, and the eigenvalues indicates the amount of variance explained by each direction. In addition to reduce the dataset by retaining only part of the explained variance, PCA is usefull for denoising.

The table 3.5.1 indicates all the explained variances and the weight values of the initial variables for each principal components.

PC	Explained Variance (%)	Quantitative T1	Quantitative T2star	Quantitative T2	GFA	ADC, $b = 4500s/mm^2$
PC1	28.4	-0.32	0.67	0.62	0.22	-0.07
PC2	24.1	0.20	0.01	-0.05	0.65	0.73
PC3	19.7	0.84	0.14	0.40	-0.32	0.08
PC4	15.0	0.32	-0.18	0.05	0.65	-0.66
PC5	12.7	-0.21	-0.70	0.66	-0.01	0.13

Table 3.5.1: Principal components explained variances and weights

The principal components can be seen on the Fig. 3.5.4. The first main component explains the anatomical metrics with high resolution to distinguish the different structures. The second is mainly explained by diffusion metrics and the fifth represents the low frequencies of the data set, showing fine details. The third and fourth principal components are mainly driven by field aberrations, allowing the other components to be corrected.
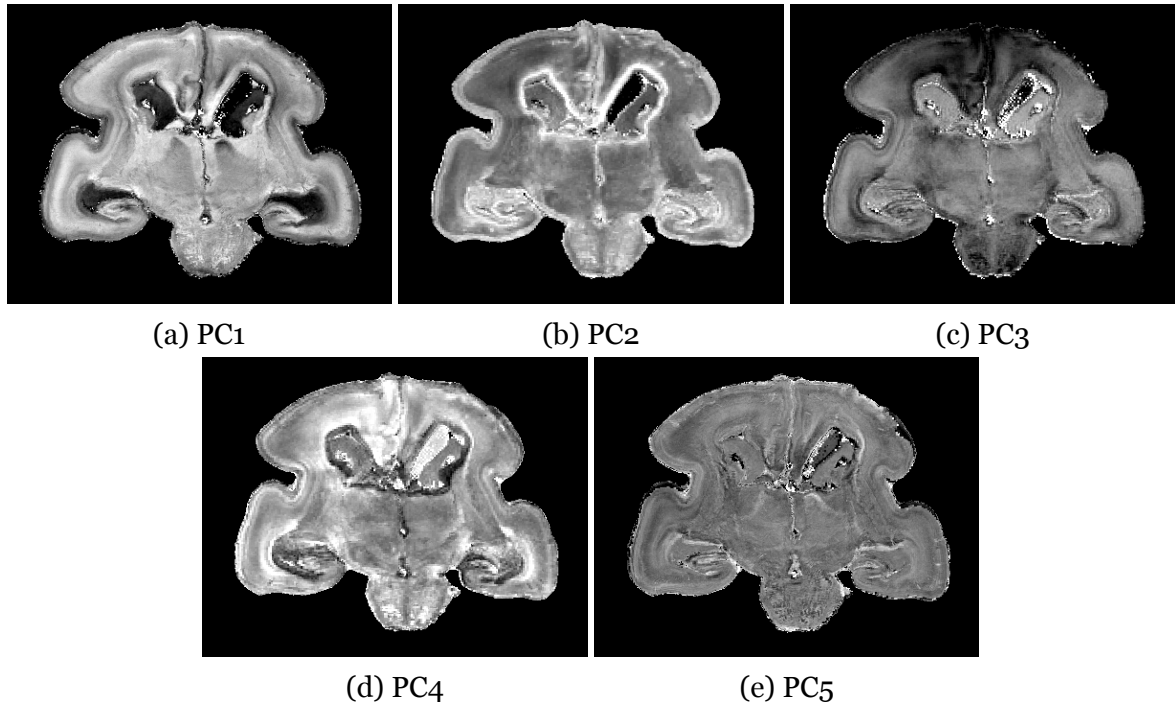

(a) PC1   (b) PC2   (c) PC3

(d) PC4   (e) PC5

Figure 3.5.4: Principal Components Analysis: coronal sections of the 20 GW specimen

3.6 Automatic segmentation

Now the dataset has a reasonable size, we can begin the automatic segmentation. This section aims to explain how the automatic segmentation was done and driven by the previous results.

3.6.1 Data-driven segmentation strategy

The data-driven segmentation is principally done using the K-meas clustering [24]. Indeed, this is the most basic segmentation tool and the easiest to implement, so it is the good first step of an automatic clustering.

The automatic segmentation needs to be separated into multiple steps. (see strategy : Fig. 3.6.1). It was not possible to obtain interpretable clusters using a single grouping. It was decided that clustering would be carried out iteratively: if groups could be distinguished from one or more images, the clustering algorithm would be applied, and this operation repeated on other images until the final group was found.
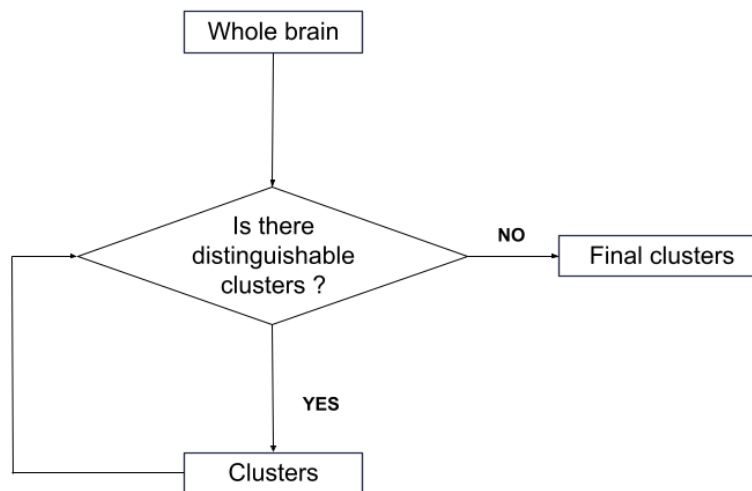

Figure 3.6.1: Segmentation strategy

The K-means algorithm decomposes a dataset into k clusters, or groups (see Fig. 3.6.2). Each cluster has a centroid which is the mean of all the values in this group. It first randomly determines k centroids on the dataset and then assigns each data

point to the group with the closest centroid. Then the new centroids of each group are re determined by taking the cluster gravitational centers and this is done as many time as decided or until the centroids do no longer change. The number of clusters can be determined using other algorithms [25] or by the user, depending on what is required.
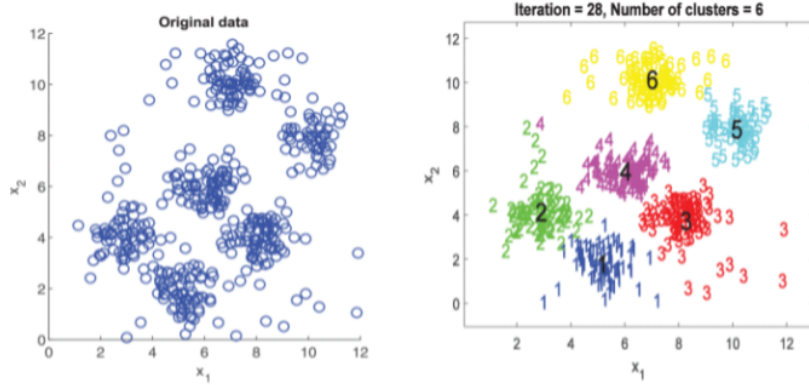

Figure 3.6.2: Kmeans illustration from "Review on determining number of Cluster in K-Means Clustering" [24]

The number of clusters was determined by starting from $n = 2$ (the image background and the brain) and adding a new cluster until we reach a result that changes very little. Indeed, once the relevant number of clusters has been reached, new labels are assigned only to individual voxels or very small and irrelevant groups of voxels. This step is arbitrary.

Identification of the structures within the regions determined by the data-driven segmentation were mainly based on the fetal brain atlas [21]. For the regions of the manual segmentations 3.3.2, an inclusion coefficient was also calculated to add a quantitative value for the verification. The inclusion coefficient used is the number of voxels present in the manually segmented and the data-driven segmented masks divided by the number of voxels in the manually segmented mask.

3.6.2 Validation

Once the clusters are determined, interpretation and performance metrics were used to validate the segmentation.

Mahalanobis distance map

To validate the automatic segmentation, I used the data from the manual segmentation. To do this, we generated Mahalanobis distance maps, enabling us to identify which structures were similar to each other. These maps also helped us to interpret the results by calculating the mean and standard deviation within these clusters, of the distances to structures whose composition was known.

The Mahalanobis distance is defined as:

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})}$$

where $(\mathbf{x}, \mathbf{y})$ are the two vectors between which the distance is calculated, $(\mathbf{x}, \mathbf{y})^T$ is the transposition of the difference vector and Σ^{-1} is the inverse of the covariance matrix of the segmented images. This metric takes into account the distribution of voxel values within the structure. In this case, x was the vector of the voxel values in the five modalities and y was the vector of the averages within the studied structure in the five modalities.

The calculated distances between the averages in the structure for each modality and the voxel values for the same modality was then injected into the final image at the same position as the initial voxel.

Those map were used calculating means and standard deviation of the distances from a manual segmentation for each clusters to understand the dispersion of the data in our five main metrics and to guide the automatic segmentation.

Structures seen within the label

The identification of the structures within the K-means clusters was firstly based on the fetal brain atlas [21] and our knowledge.

To validate the clusters, two metrics were used :

- The inclusion coefficients that represents the proportion of the manual segmentation included in the automatic segmentation.

- The DICE coefficient, an inclusion coefficient that takes into account false negatives: it can be used for segmentation, which are the only ones in an automatic cluster.

Chapter 4

Results

4.1 Manual segmentation characterization

The first objective was to characterize the fetal brain structures. We used the manual segmentation to learn more about the regions of interest. Each structure is characterized by its 'vignette': a series of violin plots in each modality (see example Fig. 4.1.1). This visualization enabled us to quickly compare the structures.
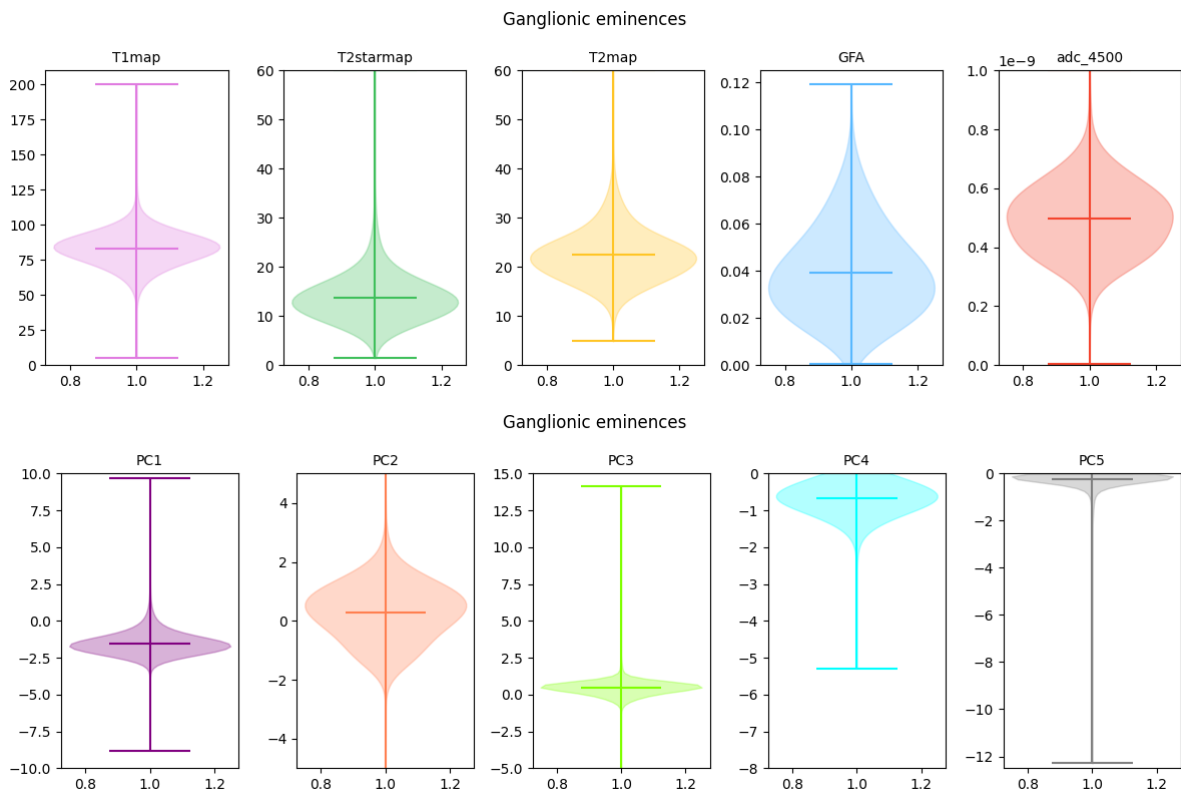

Figure 4.1.1: Example of violin plots of the ganglionic eminences

It gives information on the distribution of each structure within the different MRI parameters (in the 5 modalities and the principal components) and then on their composition. We can analyse with the figure 4.1.1 that the ganglionic eminence is characterized by a mean T1 of 80 ms while T2* and T2 are smaller (respectively 14 and 22 ms). The GFA has a mean value of 0.04 but is really dispersed, like the ADC. For the principal components, the first and fifth ones have a low standard deviation, whereas the second one is, as the diffusion metrics, quite dispersed.

4.2 Data-driven segmentation

The second objective was to implement an automatic segmentation strategy on the 20 GW fetal brain.

As a reminder, we will use the five principal components and the five modalities : quantitative-T1, quantitative-T2*, quantitative-T2, ADC, GFA.

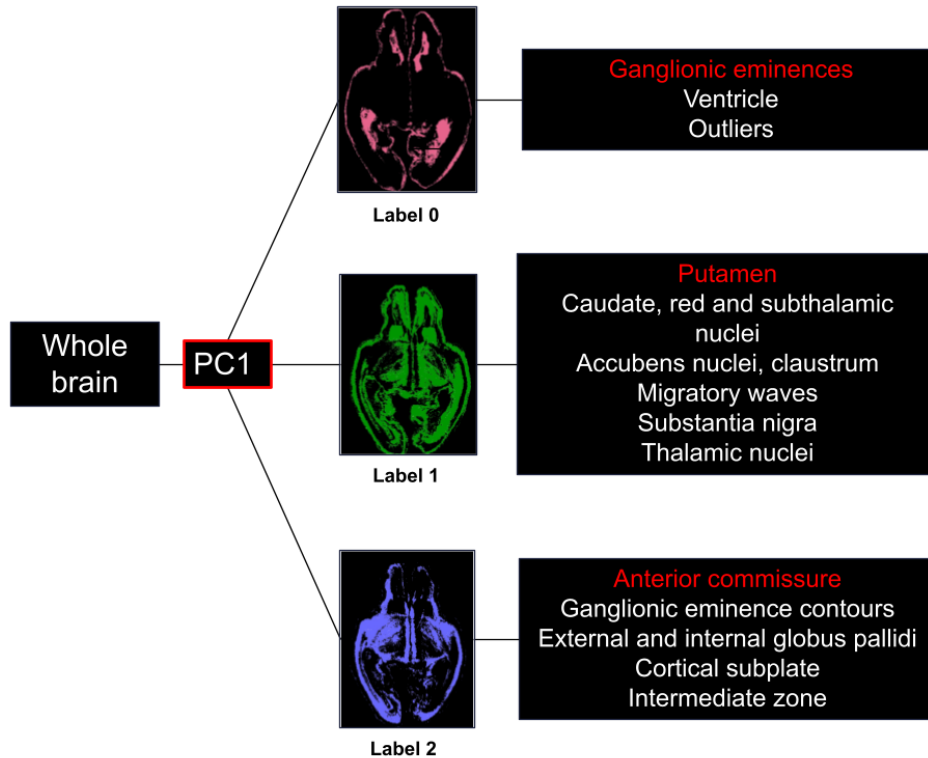### 4.2.1 First clustering

Figure 4.2.1: First clustering results

We first do a K-means on one relevant metric. PCA enabled us to obtain homogeneous images, an important criterion for the Kmeans algorithm so the first clustering was done on the first principal component (the one that explains the more the variance on the initial dataset and that has no artefact). Three clusters were extracted.

The figure 4.2.1 summarizes the first step of the K-means method and shows the first three clusters indicating the structures visible inside.

4.2.2 Characterization of the first three clusters

Once the three clusters were determined, we looked at their violin plots to guide the rest of the segmentation strategy. Label 0 or label 1 did not result in any observations that could be used to guide further segmentation.
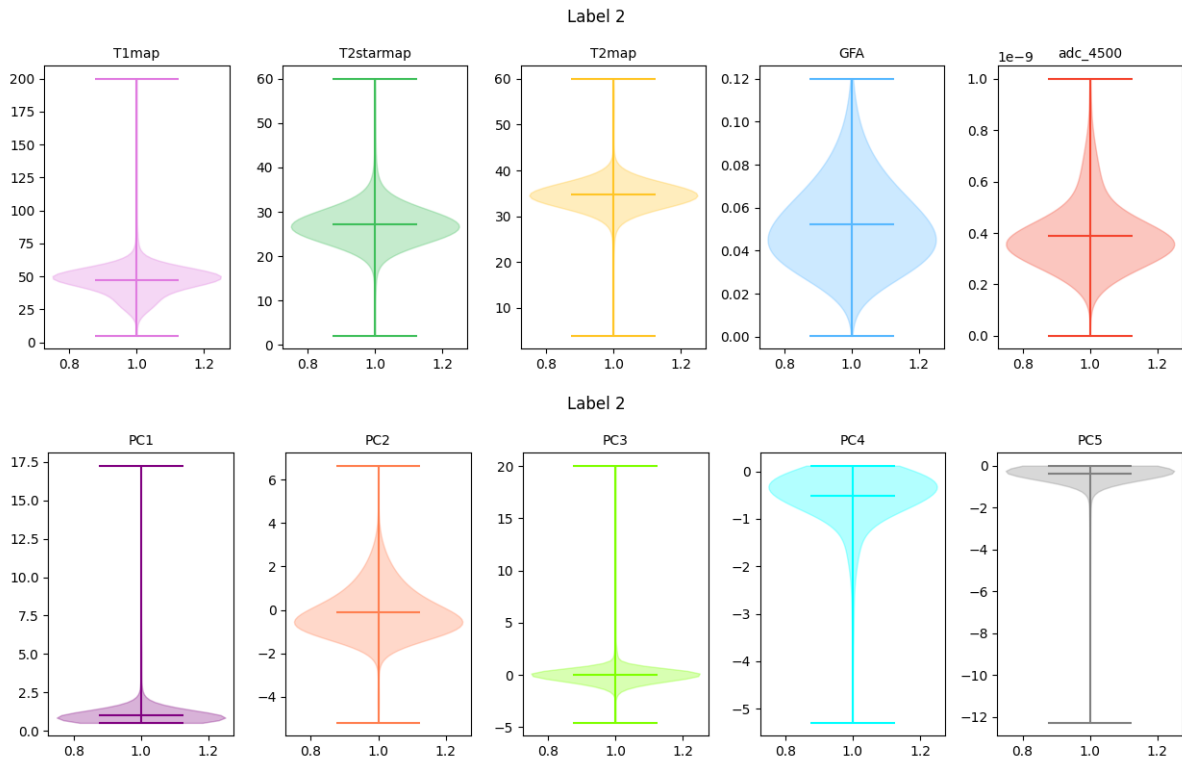
Figure 4.2.2: Example of violin plots of Label 2

The figure 4.2.2 shows an example of the violin plots of the Label 2. It shows that the GFA, ADC and then the PC2 (that is mainly explains by the diffusion metrics) are really dispersed. For the other two clusters, the violin plots did not identify any metrics that could be used for other conclusive groupings. In fact, we were unable to interpret the results of the K-means on selected modalities by looking at these violin plots.

4.2.3 Clustering of the label 2

Based on the latest observations on the violin plots, we decided to do a K-means on the principal components that allow us to do a clustering on the two diffusion metrics that have a dispersed violin plots while excluding modalities containing artifacts. That is why we chose to do the clustering based on PC1, PC2 (mostly explained by the diffusion metrics) and PC5.

We then can distinguish two clusters :

- Label 2_1 : ventricles, ganglionic eminence contours (Deep cellular layer, the first sojourn zone to appear outside the germinal matrix/ Late-forming deep layer of callosal fibers outside the germinal matrix) and anterior commissure.

- Label 2_2 : internal and external globi pallidi, cortical subplate and a part of the intermediate zone.

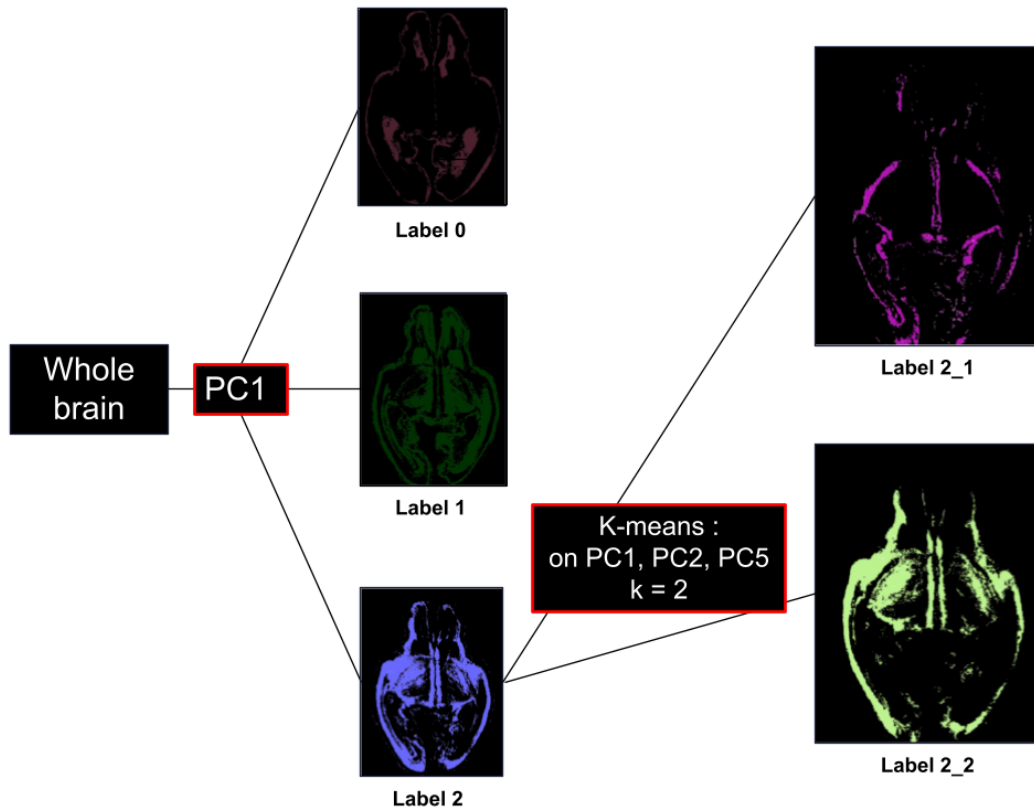The figure 4.2.3 summarizes the clustering strategy and results for the Label 2.

Figure 4.2.3: First clustering results

4.2.4   Clustering of the Label 0

To find another way of chosing metrics for the clustering, we then began to interpret the first three labels according to the biology (tissue composition).

Looking at the structures within the clusters, it seems that they have been grouped according to their density. On the label 0, the ganglionic eminences are structures with a really high concentration of progenitors cells [26]. The label 1 groups structures with an reasonable density of cells, more dispersed (the putamen is made of many cell aggregates[6]). The label 2 is on the contrary, made of structures with a low cell density (the anterior commissure is a bundle of neuronal fibers[6]).

Mahalanobis distance maps were computed from manually segmented structures on the five main modalities kept. As mentioned above, the ganglionic eminences (for which we have manual segmentation) can serve as a reference for regions of high cell density. We then looked at the Mahalanobis maps from the left ganglionic eminence and calculated, within the different labels the average of this distance and its standard deviation (see Table 4.2.1). This will give an idea of the similarity between the clusters and the ganglionic eminences and then Mahalanobis distance will act as a cell density descriptor.

Cluster	Mean	Standard deviation
Label 0	49.7	49.5
Label 1	42.41	15.3
Label 2	66.7	32.2

Table 4.2.1: Mahalanobis distance to the left ganglionic eminence : mean and standard deviation within the labels

The Strandard Deviation (STD) within the label 0 shows that the dispersion on the data on the five modalities (quantitative T1, quantitative T2*, quantitative T2, ADC, GFA) is unexpectedly high. The disparity in distances to this reference structure therefore suggests that different clusters are still separable, using the same five modalities as for the Mahalanobis distance.

Using the K-means on the five main modalities for the label 0, a mask containing the ganglionic eminences (this label will be named 'GE') was retained. The remainder of label 0 will be called 'Outliers', and contains outliers and ventricles and the part of the

ganglionic eminences not segmented by this new step of the K-means algorithm (see Fig 4.2.4).
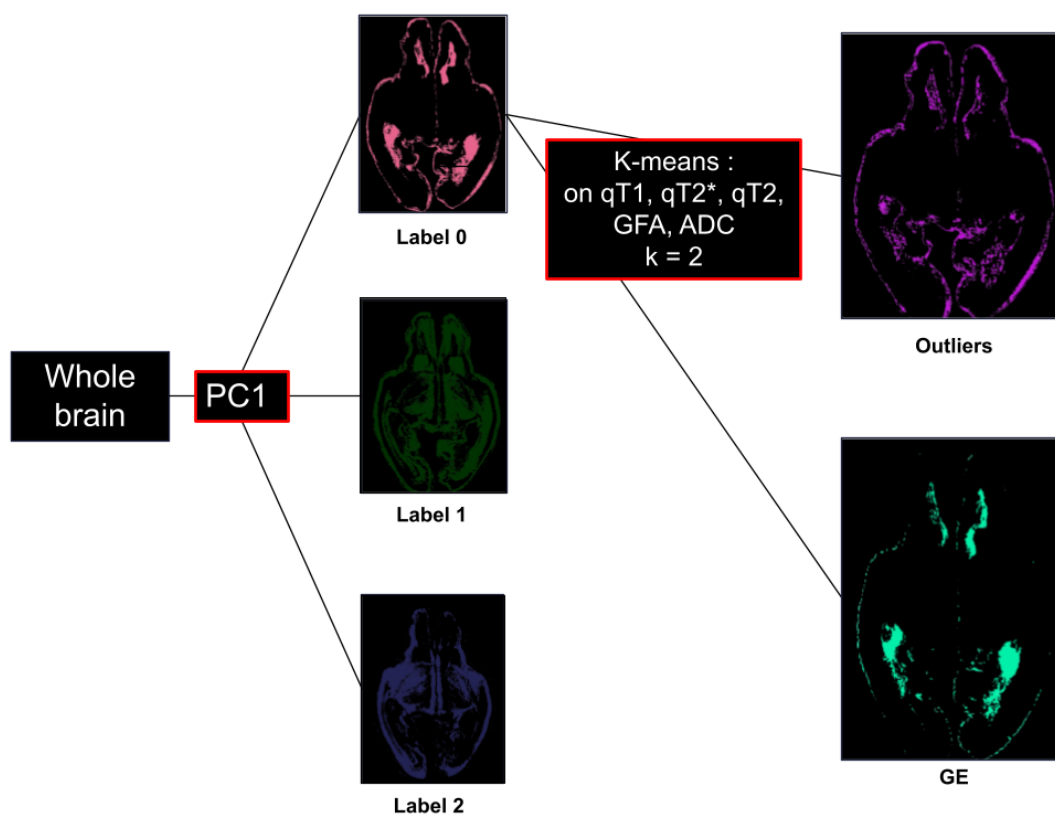

Figure 4.2.4: Results of the clustering on the label 0

To check this strategy and the two new clusters, the Mahalanobis distance averages and standard deviations within them :

Cluster	Mean	Standard deviation
GE	35.1	22.3
Outliers	57.0	57.1

Table 4.2.2: Mahalanobis distance to the left ganglionic eminence : mean and standard deviation within the new labels

The results in the table 4.2.2 for the ganglionic eminences mask GE are more expected in regard of the other labels (Table 4.2.1). The results of the outliers are inconclusive, with a high mean and high standard deviation, as expected, indicating high disparity within the cluster.

To understand why the mean and standard deviation are still high, and to have a reference, we calculated the mean and standard deviation of the Mahalanobis distance map within the manually segmented ganglion eminences versus the manually segmented ganglion eminences themselves.

Cluster	Mean	Standard deviation
Ganglionic eminences	20	11.8

Table 4.2.3: Mahalanobis distance to the left ganglionic eminence : mean and standard deviation within the manually segmented ganglionic eminences mask

If automatic segmentation were similar to manual segmentation, which we take as our reference, the mean and standard deviation results would be identical to those in the table 4.2.3. The differences are probably due to artefacts in the quantitative metrics.

4.2.5   Final clusters

The figure 4.2.5 summarizes the clustering strategy and the results of the clusters.
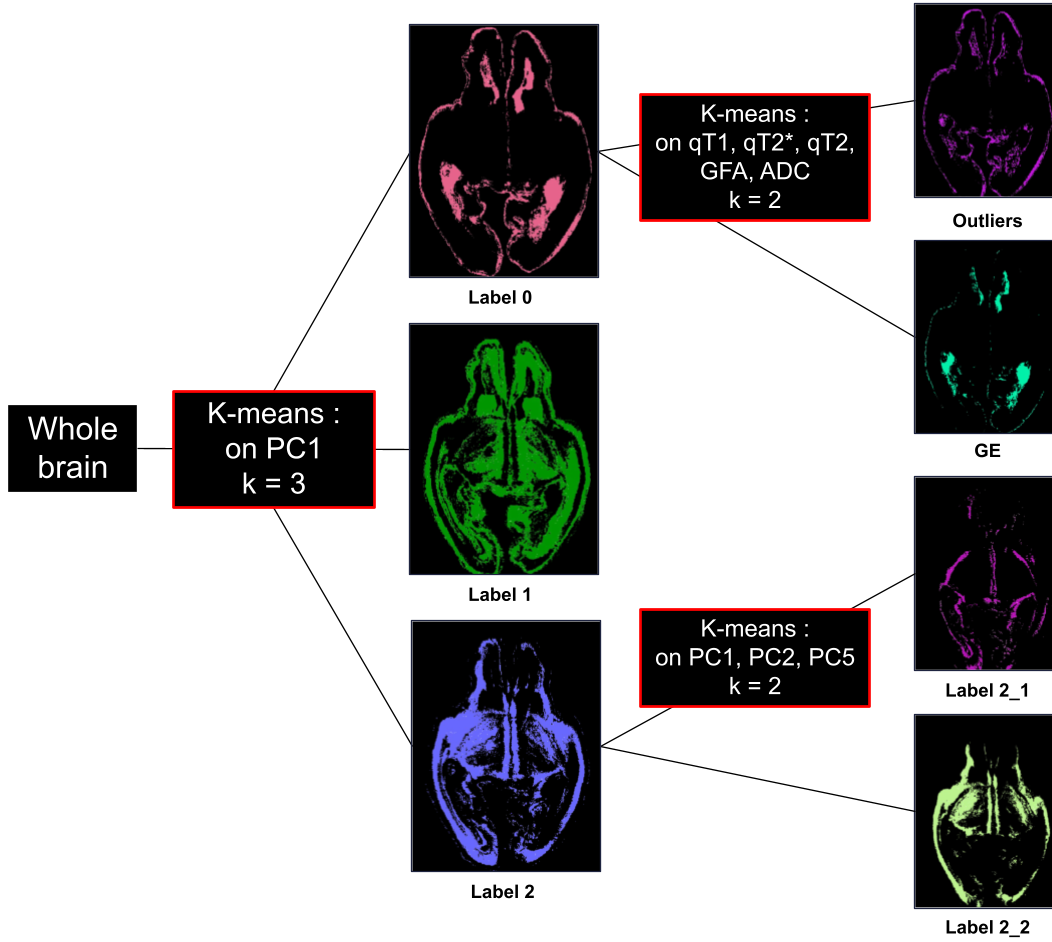
Figure 4.2.5: Strategy and clusters results

Using a K-means on the first principal component, we were able to classify the different structures primarily according to their quantitative data, and as seen previously, their cell density. The second step on Label 0 then enabled us to mask the ganglionic eminences, essential structures at this stage of brain development. Clustering on Label 2, using mainly diffusion data, enabled areas to be separated according to the direction of water diffusion within the tissue (istropic or not).

The figure 4.2.6 shows the clusters on the same image, with the colors :

- fuchsia : Outliers

- cyan : GE

- green : Label 1

- purple : Label 2_1

- light green : Label 2_2
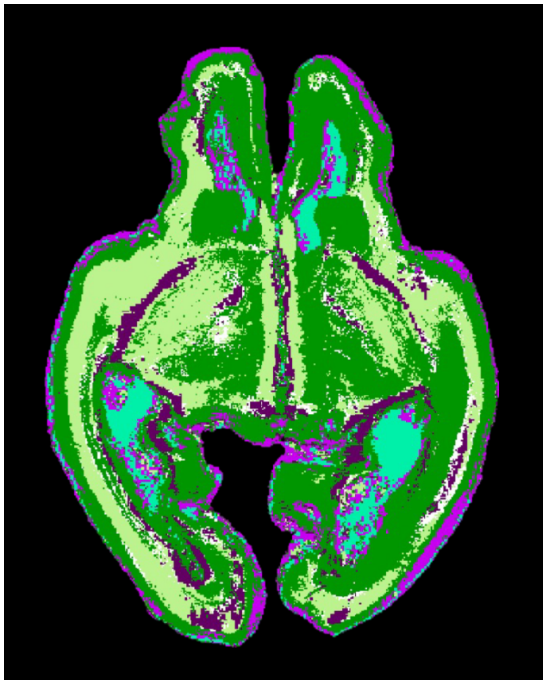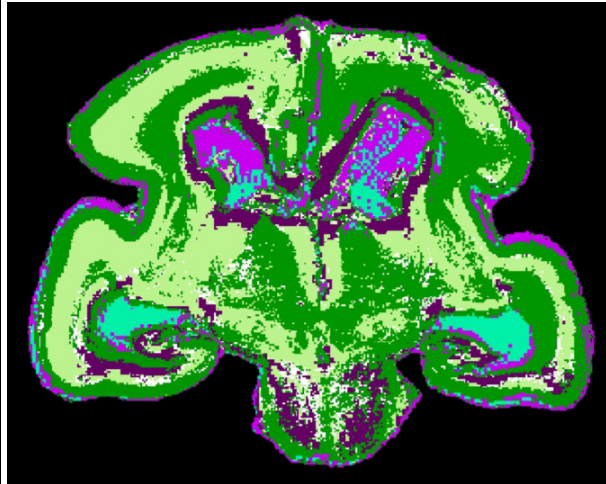
(a) Axial section   (b) Coronal section

Figure 4.2.6: Visualization of the automatic clustering

4.3   Validation of the automatic clusters using manual segmentation

Once we had the final clusters, we used manual segmentations to validate the structures present within them. Based on these manual segmentations, we used metrics to check whether the structures we thought we would find were indeed inside the clusters.

The figure 4.3.1 summarizes the results of the inclusions coefficients.
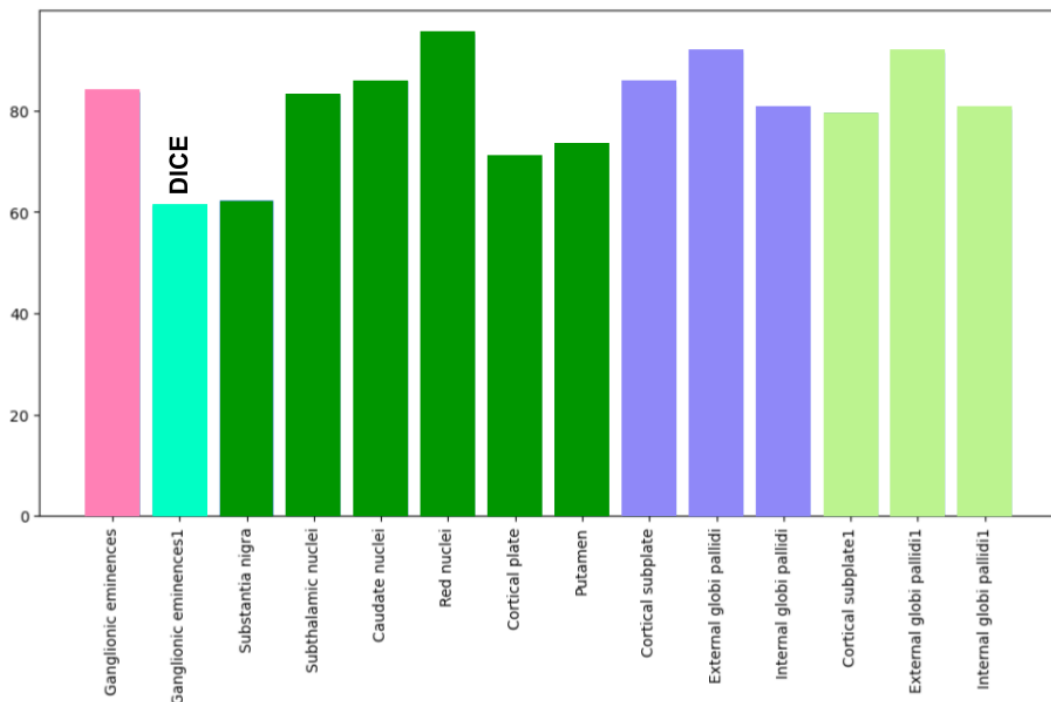
Figure 4.3.1: Inclusion and DICE coefficients results (in %)

The colors used are the same as explained for the figure 4.2.6. Each bar represents the percentage inclusion coefficient for segmented structures within the final clusters. Structures having been segmented in two different left/right clusters have been grouped together for this calculation. All coefficients found are higher than 60%, which is reasonable taking into account artifacts and potential manual segmentation issues. The structures with the lowest inclusion coefficients are not surprising: these are regions that are difficult to segment manually due to artifacts, tissue problems or data interpretation difficulties.

Chapter 5

Discussion

In this section, we will evaluate the results, discuss what could have been done differently or better, and what will be done after this master thesis.

5.1 Interpretation

5.1.1 About the MRI parameters

To understand what the different clusters represent, we first need to look at the modalities used and the information they contain. The different MRI parameters explain the composition and the structures organization within the brain tissues. The quantitative images can be considered as metrics of contents within the tissue [17]:

- quantitative T1 : explains the water and lipid contents

- quantitative T2 : explains the water content and iron accumulation

Both decrease with age, the quantitative T1 with 'pre-myelination' (membrane proliferation in the intra- and extracellular space), and the quantitative T2 with myelination. [17] As a reminder, in the brain here, aged 20 GW, myelination has not really begun. Using quantitative metrics such as T1, T2 and T2* makes also apparent the densely packed microstructures such as the neocortex [15].

The DTI metrics depend on the different maturational stages, they reflect the neuronal fibers organization as well as the pre-myelination and the myelination. [17]

The GFA give also information about the fibers organization.

5.1.2 Within the clusters

First clustering

The first K-means is based on the first principal components (PC1) that is mainly driven by quantitative-T1, quantitative-T2 and quantitative-T2*. The initial thought, that this clustering explains a gradient of cellular density, is then confirmed by the interpretation the quantitative metrics explained by PC1. T1, T2 and T2* appear here as proxy of the cell density [15].

Label 0 clustering

The Mahalanobis distance was used to separate region of ganglionic eminences (label GE) and outliers according to their local dispersion 4.2.2. This indicates that the cells of ganglionic eminences are fairly homogeneous. The results of the mean Mahalanobis distances of the Label 1 and 2, validate the cell density hypothesis for the first clustering.

Label 2 clustering

The second stage of K-means on Label 2 is based on the first, second and fifth principal components (PC1, PC2, PC5). It can be seen from the violin graphs (Fig 5.1.1) that the threshold was carried out mainly on PC2.
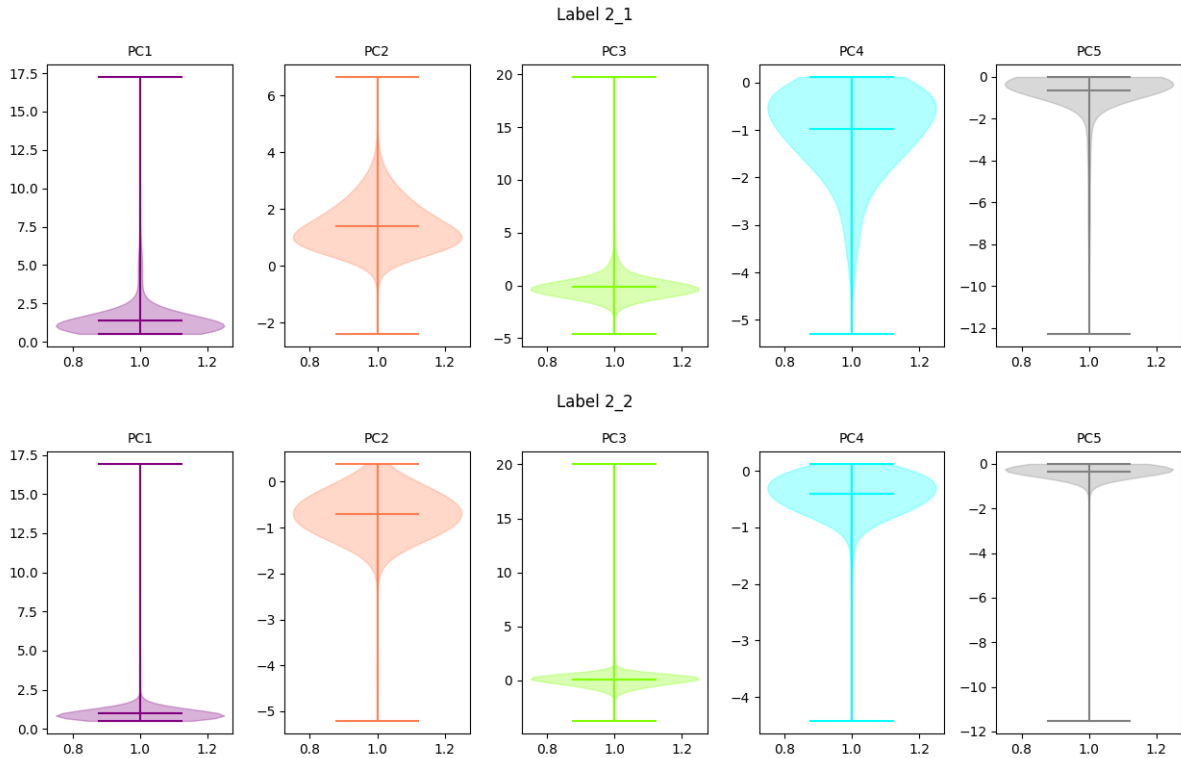
Figure 5.1.1: Violin plots of the Label 2_1 and 2_2 of the principal components

This second principal component explains the diffusion parameters. The two clusters have therefore been separated by their diffusion characteristics, as seen on Fig 5.1.2.

The first cluster has a high GFA, meaning that diffusion is fairly anisotropic (limited to one main direction) and the second has a low GFA, meaning that diffusion is more isotropic (in all directions). It would appear that this grouping has separated areas containing mainly fibers according to their dispersion.
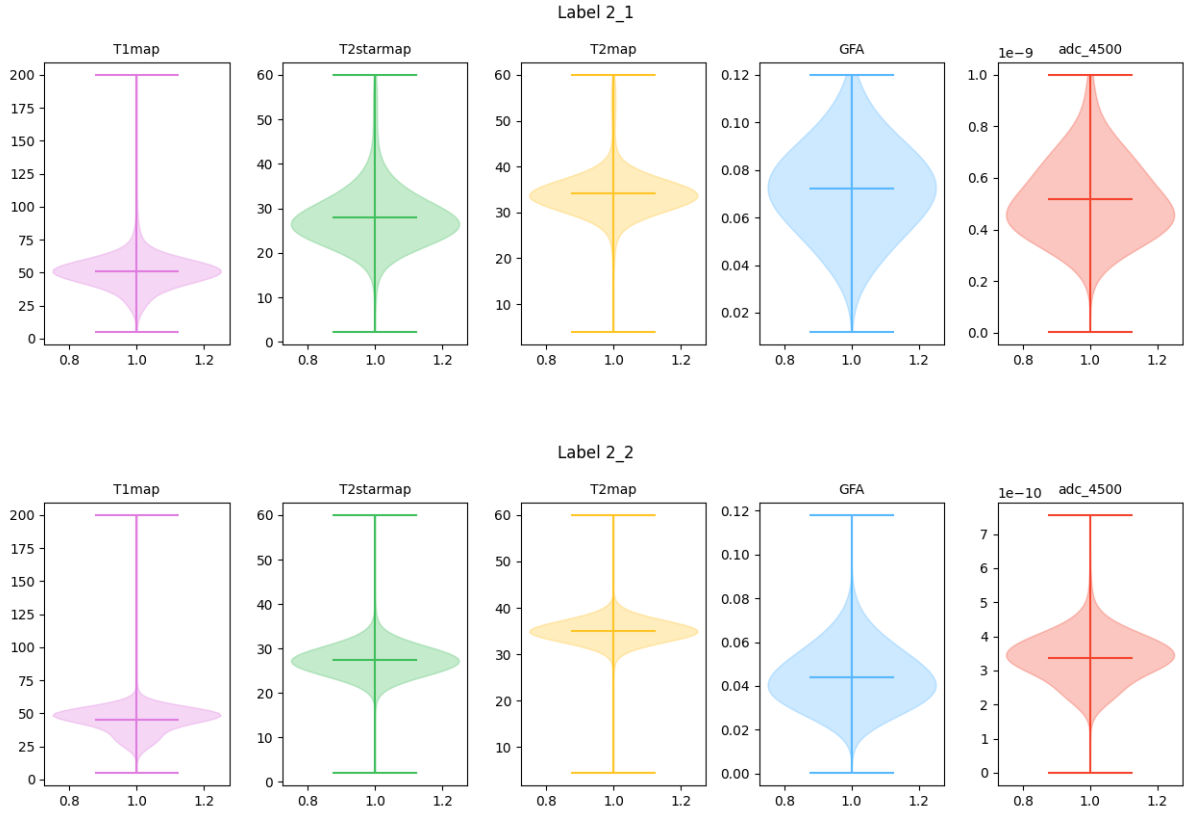
Figure 5.1.2: Violin plots of the Label 2_1 and 2_2 of five modalities

5.2 Limitations

5.2.1 MR images

Some of the MR images we used for this master thesis contained artifacts. On the quantitative images, there was an area of hypersignal that was not anatomically explained in the frontal part of the brain. This was due either to the sensitivity zone of the MRI antenna, or to the preparation of the tissue, i.e. to the presence of a gradient of the dopant within the tissue, and hence a gradient of relaxation time. For the diffusion metrics, we observed an inconsistency, probably due to saturation of one of the RF amplifiers.

5.2.2 Manual segmentation

Manual segmentation was carried out by just one person. This introduces biases into the interpretation of the tissues observed on the MRI images. To counter this, another

person, perhaps more expert in fetal anatomy, would have had to manually segment the structures, which was impossible during this master thesis. In addition, the position of the brain within the slices was not always precisely the same as in the atlas, sometimes forcing extrapolation of structures. For some structures, such as the globi pallidi, delimitations with other structures were very difficult to determine, even using all modalities, and this could result in segmentation errors.

5.2.3 Automatic segmentation

Automatic segmentation followed an exploratory method. To determine the modalities and cluster numbers for each step, arbitrary choices were made. However, K-means is a simple segmentation algorithm. It is based on the histogram of the data and performs a threshold to separate the different clusters. The histograms in our data sometimes lacked distinctive peaks enabling clear and interpretable segmentation. In the case of Label 1, for example, the results of the various K-means tried did not represent a recognizable anatomical segmentation.

5.3 Perspectives

5.3.1 To improve the work

Firstly, to improve clustering, the choice of the number of clusters in a segmentation can be automated. Indeed, there are automatic methods for determining the optimal number of clusters for a K-means [25].

Then, as K-means clustering was not sufficient to segment all structures correctly, other segmentation algorithms could be used. K-means is a partition clustering (where the cluster center is defined as the center of data points) but there are many other types of clustering algorithm [27]. For example, we can try the clustering algorithm based on distribution such as Gaussian mixture model[28]. It attributes data from the same distribution to the same clusters if there exists several distributions in the original dataset.

The data set can also be more relevant, whether by improving the data engineering (data reduction, removal or enhancement of artifacts) or by adding some features (spatial continuity between the clusters or local dispersion for example).

5.3.2 Long term perspectives

This study will be applied to the other brains of the project : 29 GW, 33 GW and 41 GW.

As fetal brain development during this period of gestation is very extensive and rapid [5], this will inform us about modality-dependent changes in the composition and shape of each structure. Myelination occurring during this period should appear longitudinally at each gestational age.

Given that the composition of the brain and its structures changes significantly between the 20th and the 41st GW due to various neurogenetic events [6], the imaging modalities used at each stage will likely differ. However, the methodological approach will remain consistent. We do not expect the same structures to be grouped within the same clusters throughout development; rather, studying these differences will provide valuable insights into the trajectories of brain development.

The segmentation strategy will also be employed to create a state-of-the-art overview of typical fetal brain development at various key stages. We anticipate that white matter injuries will exhibit distinct composition and organization compared to typical white matter. Therefore, we will utilize the same strategy to segment the injured areas of the white matter for the following of p-HCP.

Chapter 6

Conclusion

This study presents methods for characterizing the signals of fetal brain structures using various ultra-high field MRI modalities and draws interpretations about their molecular composition and spatial organization. It also demonstrates a proof of concept for automatic segmentation of these MRI images without anatomical priors.

With the manual segmentation of the 20 GW brain based on a histological atlas, we characterized several main regions in quantitative and diffusion MRI metrics. We then managed to segment regions of this brain with a similar molecular composition and cytoarchitecture. We identified clusters that differ in their cell density and the anisotropy fraction of the water scattering within them. These clusters were compared with the manually segmented regions, with reasonable success.

The iterative K-means clustering algorithm we used did not yield more accurate segmentation of the structures. Some arbitrary steps could have been made automatic, and more features combined with a different grouping method might be able to extract a different segmentation based on more fine differences within the tissues.

This master thesis presents an approach to the automatic, anatomically uninformed segmentation of very high-resolution multimodal MRI images, enabling the study of brain development from a different angle than histology.

Bibliography

[1] *CEA-saclay website*. URL: `https://www.cea.fr/english`.

[2] *Neurospin website*. URL: `https://joliot.cea.fr/drf/joliot/en/Pages/research_entities/NeuroSpin.aspx`.

[3] Lucas Arcamone, Cyril Poupon, Suonavy Khung, Marianne Alison, Homa Adle-Biassette, Lucie Hertz-Pannier, and Yann Leprince. "Ex vivo 11,7T MR mesoscopic multimodal imaging of fetal brain development". In: *ISMRM24* (2024).

[4] Ali Gholipour, Caitlin K. Rollins, Clemente Velasco-Annis, Abdelhakim Ouaalam, Alireza Akhondi-Asl, Onur Afacan, Cynthia M. Ortinau, Sean Clancy, Catherine Limperopoulos, Edward Yang, Judy A. Estroff Simon K. Warfield. "A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth". In: *Scientific reports* (2017). DOI: `10.1038/s41598-017-00525-w`.

[5] J. Dubois, G. Dehaene-Lambertz, S. Kulikova, C. Poupon, P.S. Hüppi, L. Hertz-Pannier. "The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants". In: *Neuroscience* (2014). DOI: `10.1016/j.neuroscience.2013.12.044`.

[6] M. Judas I. Kostovic G. Sedmak. "Neural histology and neurogenesis of the human fetal and infant brain". In: *Elsevier* (2019). DOI: `10.1016/j.neuroimage.2018.12.043`.

[7] Lana Vasung et. al. "Quantitative and Qualitative Analysis of Transient Fetal Compartments during Prenatal Human Brain Development". In: *Frontiers in Neuroanatomy* (2016). DOI: `10.3389/fnana.2016.00011`.

[8] Ivica Kostović, Miloš Judaš, Marko Radoš, Pero Hrabač. "Human Fetal Cerebrum Revealed by Histochemical Markers and Magnetic Resonance Imaging". In: *Cerebral Cortex* (2002). DOI: `10.1093/cercor/12.5.536`.

[9] 2 G. J. Strijkers3 J. Limpens4 R. J. Oostra1 B. S. de Bakker Y. Dawood1. "Novel imaging techniques to study postmortem human fetal anatomy: a systematic review on microfocus-CT and ultra-high-field MRI". In: *European Radiology* (2019). DOI: `10.1007/s00330-019-06543-8`.

[10] Alexandros Popov. "Global inference of the structural connectivity of white matter fiber bundles using deep learning approaches and microstructural prior knowledge". PhD thesis. Université Paris-Saclay, 2022.

[11] Steven P. Miller Stephen A. Back. "Brain Injury in Premature Neonates:A Primary Cerebral Dysmaturation Disorder". In: *American Neurological Association* (2014). DOI: `10.1002/ana.24132`.

[12] Marret, S., Marpeau, L. "Grande prématurité, risque de handicaps neuropsychiques et neuroprotection." In: *J Gynecol Obstet Biol Reprod* (2000).

[13] Steven P.Miller Jualiane Schneider. "Handbook of Clinical Neurology, Chapter 7, Preterm brain Injury: White matter injury". In: *ScienceDirect* (2019). DOI: `10.1016/B978-0-444-64029-1.00007-2`.

[14] "Chenonceau: an entire ex vivo human brain 11.7T anatomical and diffusion MRI dataset at the mesoscopic scale". In: 2021. URL: `https://archive.ismrm.org/2021/0214.html`.

[15] Nikolaus Weiskopf, Luke J. Edwards, Gunther Helms, Siawoosh Mohammadi Evgeniya Kirilina. "Quantitative magnetic resonance imaging of brain anatomy and in vivo histology". In: *Technical Review* (2021). DOI: `10.1038/s42254-021-00326-1`.

[16] Mark D. Does. "Inferring brain tissue composition and microstructure via MR relaxometry". In: *NeuroImage* (2018). DOI: `10.1016/j.neuroimage.2017.12.087`.

[17] S. Kulikiva, L. Hertz-Pannier, G. Dehaen-Lambertz, A. Buzmakov, C. Poupon, J. Dubois. "Multi-parametric evaluation of the white matter maturation". In: *Brain structure and function* (2014). DOI: `10.1007/s00429-014-0881-y`.

[18] Marko Wilke, Bianca de Haan c, Hendrik Juenger a d, Hans-Otto Karnath. "Manual, semi-automated, and automated delineation of chronic brain lesions: A comparison of methods". In: *NeuroImage* (2011).

[19] Wilfried Philips Ivana Despotović Bart Goossens. "MRI Segmentation of the Human Brain: Challenges, Methods, and Applications". In: *Computational and Mathematical Methods in Medicine* (2015). DOI: `10.1155/2015/450341`.

[20] D.Rivière, D. Geffroy, I. Denghien, N. Souedet, and Y. Cointepas. "Anatomist: a python framework for interactive 3D visualization of neuroimaging data". In: *Python in Neuroscience workshop* (2011).

[21] Joseph Altman Shirley A. Bayer. "The Human Brain During the Second Trimester (1st ed.)" In: *CRC Press.* (2005). DOI: `10.1201/9780203507483`.

[22] Michael L. Waskom. "seaborn: statistical data visualization". In: *The journal of open source software* (2021). DOI: `10.21105/joss.03021`.

[23] Jacob Benesty, Jingdong Chen, Yiteng Huang, Israel Cohen. "Pearson Correlation Coefficient". In: *Springer Topics in Signal Processing, Noise Reduction in Speech Processing* (2009). DOI: `10.1007/978-3-642-00296-0_5`.

[24] K. P. Sinaga and M. -S. Yang. "Unsupervised K-Means Clustering Algorithm". In: *IEEE Access, vol. 8* (2020). DOI: `10.1109/ACCESS.2020.2988796`.

[25] Kodinariya, T. M., Makwana, P. R. "Review on determining number of Cluster in K-Means Clustering". In: *International Journal* (2013).

[26] Vassilis Pachnis Alexandros A. Lavdas Maria Grigoriou and John G. Parnavelas. "The Medial Ganglionic Eminence Gives Rise to a Population of Early Neurons in the Developing Cerebral Cortex". In: *JNeurosci* (1999). DOI: `10.1523/JNEUROSCI.19-18-07881.1999`.

[27] Xu, D., Tian, Y. "A Comprehensive Survey of Clustering Algorithms". In: *Annotated Data* (2015). DOI: `10.1007/s40745-015-0040-1`.

[28] Carl Rasmussen. "The Infinite Gaussian Mixture Model". In: *Advances in Neural Information Processing Systems.* MIT Press, 1999. URL: `https://proceedings.neurips.cc/paper_files/paper/1999/file/97d98119037c5b8a9663cb21fb8ebf47-Paper.pdf`.