Spatio-Temporal Patch-Based Learning for Premature Neonatal Brain MRI Analysis

Mengyuan Liu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Professor Colin Studholme, Chair

Professor Paul Kinahan

Professor Chun Yuan

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Spatio-Temporal Patch-Based Learning for
Premature Neonatal Brain MRI Analysis

Mengyuan Liu

Chair of the Supervisory Committee:
Professor Colin Studholme
Professor of Bioengineering and Pediatrics

Quantitative analysis of premature neonatal brain MRI is essential for studying early human brain development, quantification of brain injury and its impact on early postnatal neuro-development. An accurate automatic delineation of the brain structures in the MRI scan remains the first step of any morphological analysis. In such studies, scans are usually acquired at any arbitrary gestational age during a rapid anatomical growth period, and with a limited imaging time. Due to the inter-subject anatomical variations and limited image quality, it is particularly challenging to accurately and automatically segment the tissue structures in such data. The objective of this work was to develop algorithmic tools that enable accurate automatic tissue segmentation and quantitative analysis of premature neonatal brain MRI scans. Multiple methods such as combining atlas-based and patch-based method in two ways for normal brain tissue segmentation, as well as combining spatial and non-spatial dictionary learning for automatic focal lesion labeling were developed and validated to show improved segmentation accuracy. The methodology developed in this work has been used for quantitative image analysis in multiple multi-site clinical studies on brain development after preterm births.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

BS: Brainstem.

BSID-III: Bayley Scales of Infant and Toddler Development - Third Edition.

CBL: Cerebellum.

DGM: Deep gray matter, including basal ganglia, thalamus, germinal matrix adjacent to the deep gray nuclei, white matter tracts between the deep gray nuclei, hypothalamus, and pituitary neurophsis.

DSC: Dice Similarity Coefficient.

DTI: Diffusion Tensor Imaging.

EM: Expectation-Maximization.

FLASH: Fast low angle shot.

GM: Grey matter.

GMM: Gaussian Mixture Model.

GW: gestational weeks.

HD: Hausdorff Distance

IVH: Intra-ventricular hemorrhage.

MRI: Magnetic Resonance Imaging.

MS: Multiple sclerosis.

PBAEM: Patch-based Augmentation of Expectation-Maximization

PMA: Postmenstrual age.

PVL: Periventricular leukomalacia.

RDF: Residual displacement field.

SCSF: Sulcal cerebral-spinal fluids.

SMRI: Structural Magnetic Resonance Imaging.

SPGR: Spoiled gradient echo.

T1W: T1-weighted.

T2W: T2-weighted.

VBM: Voxel-Based Morphometry.

VENT: Ventricles.

WM: White matter.

WMI: White matter injury.

ACKNOWLEDGMENTS

Here I would like to express my sincere gratitude towards those who have helped me during this journey. I could not have done this without any of them.

I would like to thank my advisor, Dr. Colin Studholme, for teaching me research, scientific and critical thinking, how to be a respectful researcher, how to work in the group and so many more. I will take all the lessons with me in my future endeavors.

The support and help from my supervisory committee is also indispensable during this journey. I am fortunate to have four amazing supervisory committee members: Dr. Thomas Grabowski, Dr. Paul Kinahan, Dr. Chun Yuan and Dr. Rejesh Rao, who would make time to meet with me and offered valuable advice and insights despite of their busy schedules. I truly appreciate the close mentor-mentee relationship with them and the lessons that I learnt from them.

I have also received so much help and support from the seniors and peers in our group. Sharmi, Ania, Sinchai, Xi and Averi have taught me a lot in the first a couple of years and continued to give me valuable advice. Debosmita, Thilo, Viktoria and Mary created such a happy working environment for the last few years and made working in lab much more enjoyable. Peers from Bioengineering, Jin, Haining, Kristen and so many more showed me the power of intelligence, perseverance and hard work and inspired me to keep trying when I felt like giving up. The undergraduate researchers, Averi, Lisa, Nicole, Becky, Raley, Zachary, that I have had the pleasure to work with during the past five years showed me the power of dedication and ambition.

I would also like to thank my parents, for supporting me through the whole journey, emotionally, physically and financially. I would like to thank them for teaching me not to

give up in the face of difficulties, for making me believe that hard work will pay off eventually. Even though it might be difficult to explain the challenges to them across the ocean, they kept the trust and faith in me.

Last but not least, I could not have reached this point without the support and help from my husband and dearest friend, Yang. I would like to thank him for those late nights and weekends, for sharing my ups and downs along the way. Having you and our loving dogs by my side the whole journey has been my luckiest fortune.

This work is funded by NIH/NINDS R01 NS 061957, R01 NS 055064.

DEDICATION

to my grandparents, Baofang and Jinhuan

Chapter 1

INTRODUCTION

1.1 Image Segmentation Problems in Premature Neonatal Brain Medical Image Processing and Analysis

1.1.1 Study of Human Premature Neonatal Brain

For the past decades, study of the human brain, in terms of both anatomy and function, using medical imaging has attracted enormous interest[47, 44]. Brain anatomy remains the foundation of such studies, while functional studies have revealed many interesting findings on human brain [51]. Studying early human brain development is critical for understanding of the genesis of human neural system and early detection of diseases [61]. One particular group of subjects - premature neonates - has attracted a great deal of interest both clinically and in research [61, 68, 5, 99, 100, 85]. Premature neonates are defined as babies who are born at less than 37 gestational weeks (GW) [61, 56, 107]. Even though the preterm birth rate in the United States has declined by 11% since 2006, the rate still remains high at 11.38% in 2013 [56]. Growing evidence shows that preterm infants are often at a higher risk of anatomical abnormalities and accompanying neuro-cognitive deficits, such as cerebral palsy [85], periventricular leukomalacia (PVL) [141], ventriculomegaly [64] and severe intraventricular hemorrhage (IVH) with periventricular hemorrhagic infarction (PHI) [141]. Furthermore, evidence indicates adverse neuro-developmental outcomes later in life [107], including neuromotor function [145], behavioral disorders [64], and cognitive impairment [85, 141]. Therefore, efforts are being made to better characterize early brain development [133, 61] and to diagnose neuro-developmental abnormalities as early as possible [64, 94, 73].

1.1.2 The Imaging Tool: Structural Magnetic Resonance Imaging (sMRI)

In the past decades, magnetic resonance (MR) imaging has been an essential tool to study human brain anatomy and function of all developmental stages: fetal [127, 128, 46], neonatal [61, 87, 26, 124], children [45, 78, 125], adolescents [45, 78, 125, 44] and adults [47]. MRI generally consists of two categories: structural MRI (sMRI) and functional MRI (fMRI). As the names indicate, sMRI aims to image the anatomical structures while fMRI aims to captures the functional activities in the brain. This work focuses on sMRI which will be referred to as MRI for simplicity. Compared to adults, one of the challenges in acquiring neonatal brain MR images is the unpredictable movements of the imaged neonates. In recent years, adoption of the neonatal head coil has enabled the acquisition of high-quality neonatal brain MR images with fewer motion artifacts.

1.1.3 Automatic Analysis of Neonatal Brain MRI

Most imaging studies often rely on delineation of different tissue classes to characterize the brain anatomy as a critical step before any analysis. However, manual delineation of tissues is both time consuming and difficult to reproduce due to intra-observer variation[39], especially for studies using a large cohort [5, 59, 124]. Thus, an automatic tissue labeling technique is crucial for allowing detailed and reproducible evaluation of the brain morphometry [61, 59, 82, 99]. Compared to automatic tissue segmentation of adults' MR scans, one of the biggest challenges is the anatomical development and contrast change with age. However, the age-related change also leads to interesting studies of early brain development. There is no doubt that such studies would require an even larger number of subjects compared to adult studies. Therefore, the automatic analysis procedures are even more indispensable. To the best of our knowledge, no complete automated computational analysis pipeline on premature neonatal brain MRI has been reported.

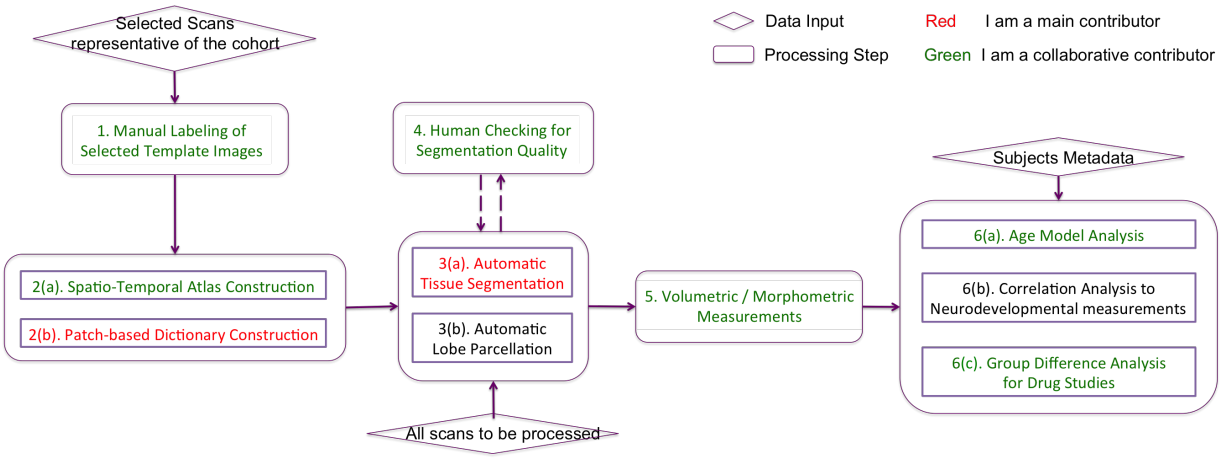

Figure 1.1: The automated computational analysis pipeline of premature neonatal brain MRIs. The whole pipeline contains 6 major steps, with 2 and 3 as the key steps. The main contribution of this work is developing novel algorithms for step 2(b) and 3(a) (as shown in red) while adapting existing methods to applications on premature neonates for step 2(a). Other collaborative contributions are shown in green.

1.2 Contributions of this Work

The main contribution of this dissertation is developing an automatic quantitative analysis pipeline of premature neonatal brain MRIs. Specifically, the work in this dissertation will enable accurate automated tissue labeling, abnormality detection, morphometric measurements extraction and outcome prediction of a premature neonatal brain. The application of methods and results presented in this dissertation can realize automatic processing of large amounts of premature neonatal brain MRI data in both research and clinical settings. In research, especially, in studies with a large cohort, this work will help to avoid any laborious manual delineation of tissue and lesions. In clinical settings, this work will enable automated detection for brain abnormalities and prediction of neuro-developmental outcome to help the doctors with diagnosis and treatment planning. I summarize the pipeline and how the contribution of this dissertation fits in the pipeline in Figure 1.1.

1.3 Dataset Used in this Work

Through this work, we used 3 main datasets consisting of T1w human premature neonatal brain MRI scans. The datasets are briefly introduced in Table 1.1 here and details will be given in appropriate sections. Note that all 3 datasets have been already collected by our group and collaborating groups before this work.

Table 1.1: Main datasets used in this work. All scans are T1w human premature neonatal brain MRI scans.

Name	Total Scans	Age Range	Manually Labeled	Processed	Comments
Vancouver	236	27 - 46 GW	36	170	main dataset
UCSF	103	30 - 37 GW	8	8	supporting dataset
PENUT	185	36 GW	4	111	7-site dataset

1.4 Structure of the Thesis

Chapter 2 provides a thorough and detailed literature review of the current field in terms of segmentation methods and their application on human brain MRI images, with a comparison between the challenges of adult and neonatal tissue segmentation.

Chapters 3, 4 and 5 will focus on introducing new segmentation methods for healthy normal premature neonatal brain MRI images and its application to a multi-site data set. Specifically, Chapter 3 describes a novel approach to automatically label different tissue classes in premature neonatal brain MR images and its validation. The described method is developed upon the spatio-temporal atlas-based EM segmentation technique proposed by Habas et al [54, 53] by introducing an additional source of tissue probability estimate using patch-based technique[30, 109]. The combining technique is a weighted fusion of two sources of tissue probability estimates. Therefore, we refer to this approach as PBAEM-WF. Chapter 4 describes the improvement made on PBAEM-WF by using a different method to

integrate the patch-based local search into an atlas-based EM framework. First, I learn the variability of anatomy at each location in the brain. Then, the patch search is made to cover the potential variability range to ensure a higher probability of finding the match. In the rest of the dissertation, this method is referred to as PBAEM using variability-constrained search (PBAEM-VCS). Chapter 5 presents validation of the automated tissue labeling techniques using multi-center datasets acquired using different scanners and different imaging protocols. Firstly, the importance and benefits as well as challenges in processing multi-center data are given in the background section. Then a preliminary study of the previously described PBAEM validated on two multi-center cross-scanner cross-protocol datasets is presented. This study will not only serve as an extensive validation of the PBAEM approach showing that comparable segmentation performances can be acquired regardless of different MR acquisitions, but also demonstrates the potential for future multi-center study using multiple datasets to study early brain development via automated tissue labeling.

Chapter 6 will focus on introducing an innovative method for full-brain tissue segmentation with IVH and VM in non-healthy premature neonatal brain MRI. IVH is apparent on MRI as bright signal pooling within the ventricular space and appears in locations within the ventricle related to the original bleed and how the blood clots pool due to gravity and the position of the babys head. This dissertation presents a novel sparse dictionary learning method to delineate the severe IVH and VM together with normal tissue classes. The dictionary approach consists of both a spatial and non-spatial component. The spatial dictionary encodes normal variation in anatomy, while we use a non-spatial dictionary to capture the shape and occurrence of IVH voxels with respect to their commonly neighboring tissues. The two dictionaries are collectively used to estimate a probability of normal and abnormal tissues of each voxel, which is then used to initiate an Expectation-Maximization based tissue labelling of the image data. Results indicate improvement in the segmentation accuracy when compared to both conventional atlas-based and spatial dictionary learning segmentation methods. This approach provides the first automatic segmentation framework for severe IVH and VM in premature neonatal brain MRIs.

Last but not least, Chapter 7 will conclude the results and contribution of this doctoral dissertation. The disadvantages of current work and future steps will also be discussed in depth.

Chapter 2

BACKGROUND

2.1 Brain Anatomy: General and Neonatal

2.1.1 General Brain Anatomy

The human brain is the central organ of the human nervous system. The brain is composed of the cerebrum, cerebellum and brainstem. Figure 2.1 shows the relative shape, size and location of the 3 parts of the brain [84].

- **Cerebrum** is the largest part of the brain and consists of left and right hemispheres. It is responsible for all higher level functions such as touch, vision, hearing, speech, fine movement, reasoning, emotions, learning etc [84].

- **Cerebellum** is located underneath the cerebral hemispheres. Its function is to coordinate muscle movements, balance and maintain posture [84].

- **Brainstem** connects the cerebrum and cerebellar to the spinal cord. It consists of three sub-structures: midbrain, pons and medulla. The brainstem is responsible for automatic functions such as breathing, heart rate, body temperature, digestion etc [84].

2.1.1.1 Structures of interest

Here we introduce the main brain partitions that we focus on in this work. The cerebrum, cerebellum and brainstem each consists of many sub-structures. However, due to the limitation of MRI image contrast, image resolution, partial voluming and artifacts, it is

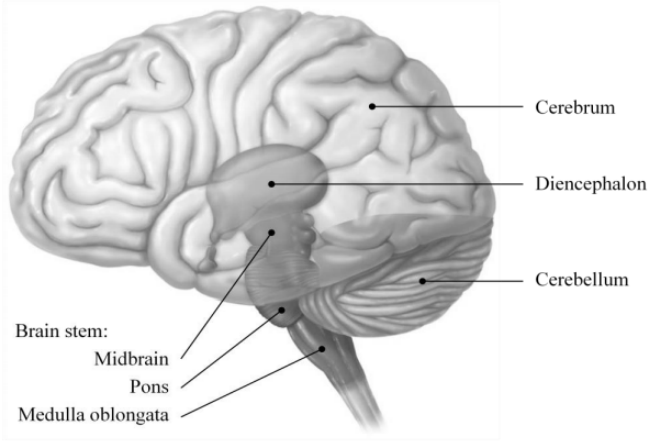

Figure 2.1: General anatomy of the human brain, The brain is composed of three parts: the cerebrum, cerebellum and brainstem [84].

extremely difficult to visualize and identify all the sub-structures. Thus, we group normal brain structures coarsely into 7 main structures of interest. We introduce the definition, including sub-structures, features and functions of these 7 structures of interest as follows. This subsection is adapted from Human Anatomy and Physiology by Marieb and Hoehm [84].

- **Grey Matter (GM)** is located at the surface of the cerebrum, named cortex. The cortical GM contains mainly neurons and few myelinated axons. The cortical GM has a folded appearance where a fold is called a gyrus and a groove is a sulcus (Figure 2.2) [28]. The folding of GM increases the surface area of the brain and enables higher functions.

- **WM Matter (WM)** is the inner region of cerebrum and consists of mainly myelinated axon tracts. Myelination contributes to the white color. An interesting structure within WM is Corpus Callosum (CC). CC is a thick axon bundle that connects the WM of left and right hemispheres [84].

- **Deep Grey Matter (DGM)** is located deep within the cerebral WM, and consists

Figure 2.2: Gyrus and sulcus of the cortical GM [28].

of a group of sub-cortical GM called basal nuclei and diencephalon. The basal nuclei is composed of the caudate nucleus, putamen and globus pallidus, and contributes to regulation of skeletal muscle movements. Major structures in diencephalon include the thalamus and hypothalamus which are responsible for sensory impulses and overall homeostasis [84]. Other small structures within DGM include pituitary gland, pineal gland and the limbic system (amygdala, hippocampus and hypothalamus).

- **Ventricles / Ventricular Cerebrospinal Fluid (VENT)** are the central hollow cavities within the brain. The hollow ventricular chambers are filled with CSF [84]. The VENT consists of inter-connecting lateral ventricles, third ventricle, cerebral aqueduct and fourth ventricle (Figure 2.3). VENT is connected to sulcal CSF (sCSF) and the spinal canal [84].

- **Brainstem (BS)** See above.

- **Cerebellem (CBL)** See above.

- **Sulcal Cerebropinal Fluid (sCSF)** represents the CSF in the subarachnoid space. Its function is to bathe and cushion the brain.

In this work, all the other structures present in the brain and non-brain region are considered as the background (BG). Typically, BG includes skull, facial tissues etc.

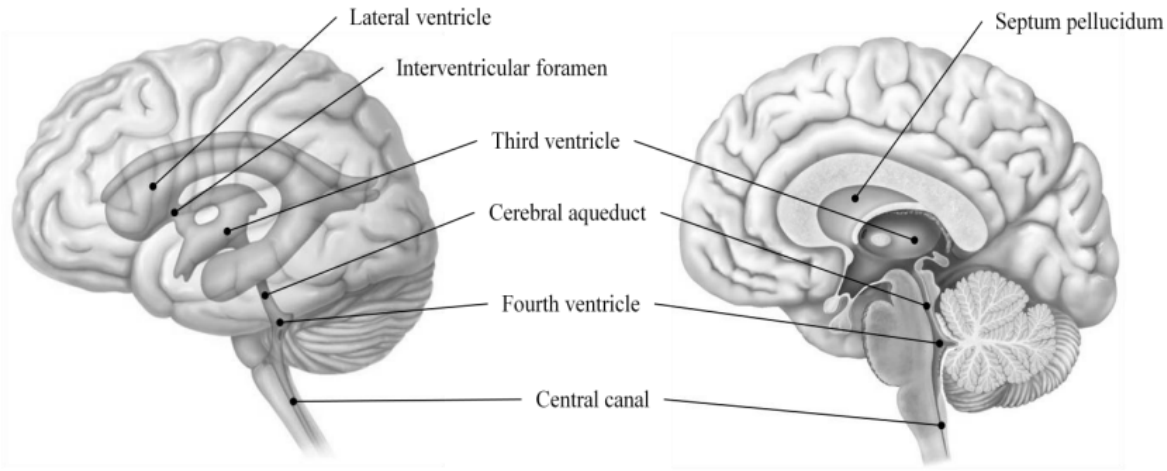

Figure 2.3: Ventricular chambers [84].

The cerebral hemispheres are divided into 4 lobes on each side by distinct fissures [28, 84]. Each lobe is responsible for some specific functions. On either left or right hemisphere, there are the frontal lobe (emotion, speech, intelligence etc), parietal lobe (language, sensation, perception etc.), occipital lobe (vision and its interpretation) and temporal lobe (memory, hearing, understanding language etc) (Figure 2.4) [84]. Because of their unique functions, examination of their own morphological measurements, such as volume and cortical surface area, can also provide valuable insights on brain development.

2.1.2 Neonatal Brain Anatomy

The human brain undergoes a complex process of morphological and functional maturation and development [86]. Compared to the previously described adult brain anatomy, the neonatal brain, and especially premature neonatal brain has many unique anatomical, morphological and functional characteristics [112]. These characteristics can certainly cause complications in the quantitative analysis of neonatal MRI images. All the unique features are the results of the continuous brain maturations that include morphological complexity,

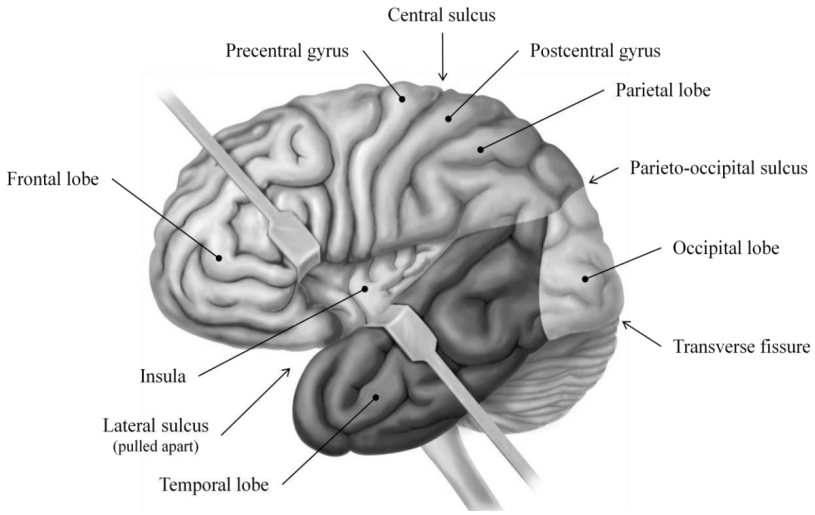

Figure 2.4: Lobes and separating fissures of the cerebral hemispheres. [84]

water content level, overall size, volume ratio across structures etc [86]. They are reflected in the MRI scan as visibly more complex structures and contrast change. I first discuss these continuous changes with ages from 24 GW to post-term in the section and later cover how these changes manifest themselves in the MRI scans.

2.1.2.1 Cortical Folding

The increases of overall size and cortical folding are the most obvious changes fetal or preterm neonatal brain after 24 weeks from histology samples [112]. While overall size is easy to understand since the rest of the body undergoes dramatic growth in size, the complication of cortical folding is unique to brain development. Cortical developments start from approximately 8 GW by replicating neurons and glial cells that migrate to the outer surface to form the cerebral cortex. Neuronal migration to the cerebral cortex is completed by 20 GW to 24 GW [112]. Therefore, the so-called GM can be shown on MRI of fetuses or neonates older than 24 GW. From 29 GW to term, cortical GM increases in volume from approximately 60ml to 160ml [61]. During the early gestational stage, the cortical layer has a smooth

appearance, described as lissencephalic. To allow further surface area increase, the growth proceeds in a convoluted pattern to form the sulci and gyri. Van der Knapp et al. [134] first studied the sulcation and gyration in infants from 30 GW to term using MRI, followed by Battin et al expanding to as early as 25 GW [13]. Battin et al. also proposed a 1-5 scoring system to describe the level of cortical folding by considering the depth, width, neighboring foldings and complexity of folding in different areas of brain [13]. Due to the time-consuming process of visual examination of images and access to subjects, the study only covered 53 scans of 17 infants. Due to the individuality of the cortical surface profile, more extensive studies are required to adequately represent the cortical folding process.

2.1.2.2 White Matter Myelination

Myelination of white matter enables faster neural impulse transmission [84]. Its progression has been extensively studied using histology. It occurs in a systematic way starting from around the end of the first trimester till adulthood, with peak myelination around the mid-20s [112]. The majority of this process happens post-term (i.e. after 40GW gestational age), with a few steps pre-term. It starts from medial longitudinal fasciculus at the end of the first trimester, then appears in the inferior cerebellar peduncles around 25 GW, followed by the inferior colliculi, posterior brain stem and ventro-lateral nuclei of thalamus [61, 25]. Onset of myelination in the posterior limb of internal capsule (PLIC) happens around 32 to 36 GW. This is the area of interest because at our gestational age of interest, myelin in PLIC shows as bright signals in T1W MRI [112]. In MRI, WM myelination can become evident in the PLIC and corona radiata at around 35 GW. It was also discovered that preterm infant show myelination within PLIC at an earlier gestational age than more mature infants [112].

2.1.2.3 Water Content

The change of water content of the human brain during growth has been investigated and reported in 1973 by Dobbing and Sands [34]. Generally speaking, the water content falls during brain development, with a rate of the reciprocal of the rising lipid concentration.

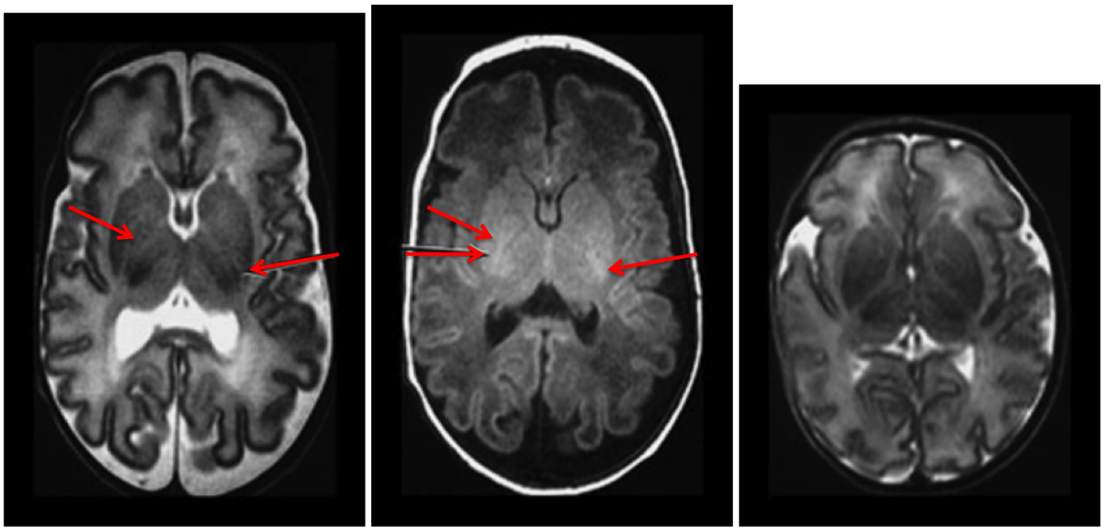

Figure 2.5: Myelination in the posterior limb of the internal capsule (PLIC). (a)(b) T2w and T1w acquisition of a preterm infant born at 30GW and imaged at 35GW. Red arrow: myelin within PLIC. (c) T2w acquisition of another preterm infant born at 34GW and imaged at 35GW, with no obvious myelin within PLIC. [112]. We see that myelination is easier to visualize in T1w scan.

Johnson et al. further reported that the neonatal brain contains a higher level of water content (92% to 95%) compared to that of 2-years-olds (70%) [66]. Its impact on choosing MRI acquisition parameters will be described later.

2.1.2.4 Prematurity

Whether prematurity has an effect on the neonatal brain development was in question for a long period of time. It was believed that the gyral development keeps a pre-set rate irregardless of being in womb or *ex utero* [88]. Huppi et al [61], however, proved that at term-equivalent age, preterm neonates have less cortical folding compared to term-born neonates, while there is no significant difference in total brain volume. These findings indicate that the brain development, especially the cortical folding process, can be disrupted by preterm birth. Since abnormalities with cortical GM development is often associated with neurobe-

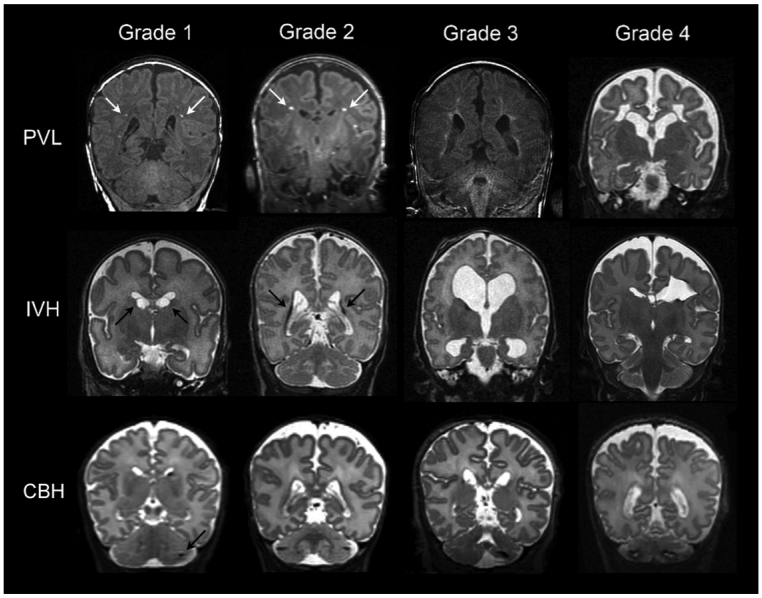

Figure 2.6: Classification of brain injury using representative T1W and T2W MRI of each grade of PVL, IVH and cerebellum hemorrhage (CBH) [72].

havioral and neurocognitive dysfunctions later in life, it becomes important to understand the effect of preterm birth on brain development. Besides disrupting the normal cortical folding rate, prematurity is often accompanied with lesions. Lesion types vary, including bleeding, excessive ventricular CSF etc. Kidokoro et al. [72] analyzed a large-scale collection of 448 premature neonatal brain images and classified and graded some of the most common abnormalities as in Figure 2.6. They concluded that both brain injury and altered development were found in the premature neonates cohorts.

2.2 MRI of Brain

2.2.1 MRI of Human Brain

MRI is an established method for visualize brain structures and diagnose abnormalities in both adults and older infants [112]. In this section, I first briefly summarize the principles of MRI and then specifically discuss the challenges of MRI on the premature neonatal brain.

2.2.1.1 Principles of MRI

MRI is an imaging modality to form images of the anatomy and physiological process of the body based on the nuclear magnetic resonance (NMR) signal from the atoms. In clinical and research MRI, hydrogen is the most commonly used atom to generate a coil-detectable radio-frequency signal, since hydrogen atoms exist naturally in humans, particularly in water and fat [41].

NMR theory Spin is a fundamental property of all nucleons. Individual unpaired nucleons possess a spin of 1/2, and can be + or -. When placed in a magnetic field of strength B, a particle with a net spin can absorb a photon of frequency ν. This frequency, which is called the Larmor frequency, depends on the gyromagnetic ratio γ of the particle. For hydrogen, γ = 42.58 MHz/T [60].

After absorbing the photon, the particle transits from the lower energy state to an upper energy state. Net magnetization is used to represent the sum vector of the magnetization vectors from all of the spin packets, where a group of spins experience the same magnetic field. At equilibrium, the net magnetization vector lies along the direction of the applied magnetic field B_o and is called the equilibrium magnetization M_o. After absorbing photons with a frequency equal to the Larmor frequency, the net magnetization in longitudinal direction M_Z is forced to be zero [14]. The time constant that describes how M_Z returns to equilibrium is called the spin lattice relaxation time (T1) [104]. The equation governing this behavior shows as follows:

$$M_Z = M_o * (1 - e^{-t/T1}) \tag{2.1}$$

After the net magnetization reduces to zero in longitudinal direction, it is only placed in the XY plane and will rotate about the Z axis at Larmor frequency. Since each spin packet experiences a slightly different magnetic field, the net magnetization begins to dephase. The time constant that describes the dephase of net magnetization in transverse direction M_{XY}

is called spin-spin relaxation time, T2, with the following governing equation [104].

$$M_{XY} = M_{XY_o} e^{-t/T2} \tag{2.2}$$

Due to the inhomogeneity of the magnetic field, a time constant that more accurately describes the decay of the transverse magnetization is $T2^*$. The relationship between T2 and $T2^*$ shows as follows [119]

$$1/T2^* = 1/T2 + 1/T2_{inhomo} \tag{2.3}$$

Different pulse sequences are applied to rotate the transverse magnetization about the Z axis, which induce a signal in the detecting coil. The amplitude of this signal depends mainly on the repetition time (TR) of the sequence, the proportionality constant k, and the density of the spins in the sample ρ. For a Spin-Echo sequence, the echo time (TE), which is the time between the 90^o and the maximum amplitude, also determines signal amplitude [119].

2.2.1.2 Contrast

To differentiate adjacent tissues, there must be a difference in detected signal amplitudes. Therefore, the contrast in MRI image mainly depends on 1) intrinsic properties of the tissue, such as T1, T2, spin density ρ; and 2)instrumental parameters, such as TR and TE etc.. Table 2.1 lists the range of T1, T2 and ρ values from brain tissue using 1.5T MRI [40]. To visualize the contrast from T1, T2 or ρ, different TR and TE are selected. For tissue with different T1, short TR with short TE is applied, which enhances the contrast predominantly caused by different T1. Three different combinations of TR and TE values that are used commonly clinically are shown in Table 2.2 .

Different TR and TE values are applied to create weighted images, thus providing better tissue contrast of the human brain. For a T1-weighted image, the CSF is dark with GM darker than WM, which is useful for visualizing normal brain anatomy. For T2-weighted image, the CSF is light with WM darker than GM. For PD (proton density) image, since CSF has a relatively high level of proton density (water), it shows bright in PD image, which is not dependent on T1 and T2 relaxation time [17].

Table 2.1: Different T1, T2 and ρ of brain tissue using 1.5T MRI [119].

Tissue	T1(s)	T2 (ms)	ρ
CSF	0.8 - 20	110 - 2000	70 - 230
WM	0.76 - 1.08	61 - 100	70 - 90
GM	1.09 - 2.15	61 - 109	85 - 125
Muscle	0.95 - 1.82	20 - 67	45 - 90

Table 2.2: Combination of TE and TR values for weighted images [119].

Weighting	TR value	TE value
T1	short	short
T2	long	long
ρ	long	Short

2.2.2 MRI of Human Neonatal Brain

The traditional way to study human neonatal brain anatomy is by histology and pathology studies [55]. However, it is impossible to study a live fetus or infant, or conduct time series study on the same subject or correlation analysis between quantitative anatomical measurements and behavioral and neurodevelopmental performance later in life. This is essential in understanding how early brain development can affect brain functions later in life. The ability of using MRI to capture the early brain anatomy is a powerful game changer, not only to study normal brain development process, but also to provide useful information on the origin of brain lesions that cause long-term adverse effect in preterm infants. Nowadays, MRI can be performed as early as 23 GW [127]. In this work, the subjects of study are mainly preterm neonates with an age range of 28 to 45 gestational age as shown in Figure 2.7.

To acquire optimal images of a preterm neonatal brain, its differences from an adult

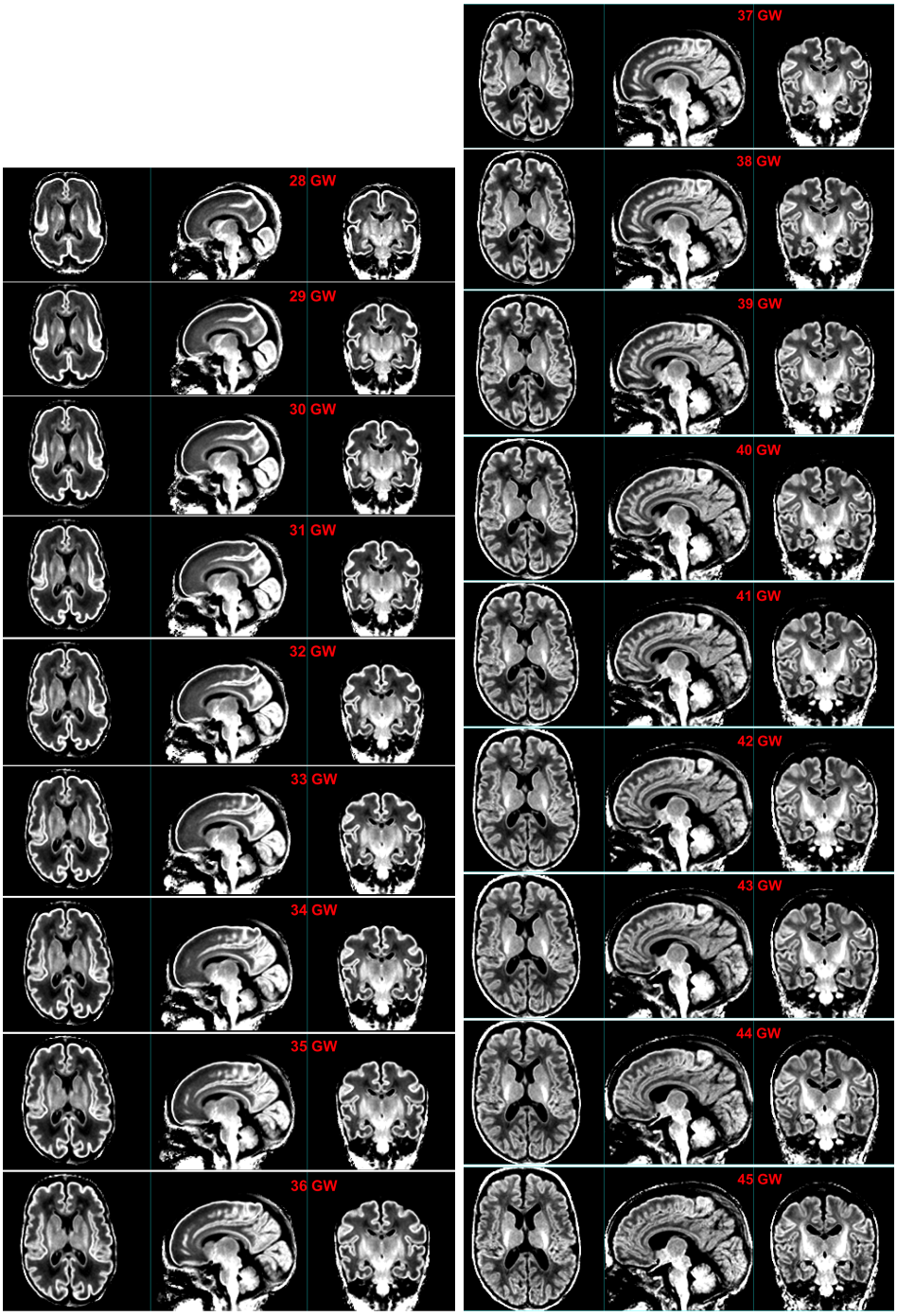

Figure 2.7: Average neonatal brain MRI over a age range of (left) 28 - 36 GW and (right) 37 - 45 GW.

brain should be properly addressed. For example, as previously described, a neonatal brain contains a higher level of water than an adult brain. This is accounted for by increasing T1 and T2 values, resulting in increased TE, TR and TI [66]. Furthermore, the use of neonatal head coils adapted especially for the small head has increased the image quality dramatically. Preterm and term neonates can be conveniently imaged using an adult knee coil or a specially designed neonatal head coil. This can greatly improve the signal to noise ratio in the MR images [112].

Sequence-wise, it becomes almost impossible to obtain multiple acquisitions using different sequences in one setting with neonates due to the special condition of this targeted cohort. In traditional MRI studies, T1 weighted conventional spin echo (CSE) images show low GM/WM contrast but superb hemorrhage contrast. T2 weighted fast spin echo (FSE), on the other hand, is fast such that motion artifacts can be reduced during neonatal imaging while showing myelination contrast [112]. In our group, premature neonatal brain MRI images were historically acquired with T1 weighted sequence, while fetal MRIs were acquired with T2 weighted technique.

During the imaging process, neonatal imaging also involves more unique elements to consider compared to adult imaging. These include access to the infant while scanning in case of emergency, ventilation, temperature maintenance, especially in clinical cases, and avoidance of sedation [112]. Even though these elements may not impact MRI image quality under an established study protocol, they are of utmost importance to ensure a successful study.

2.2.2.1 Tissue Contrast

Change with Age As previously described, human brains in gestation contain a higher water content percentage. Therefore, a longer T1 and T2 can be observed compared to adult brains. As a result, in T1 weighted MRI, WM shows a lower signal and appears darker than GM, which is opposite of adult MRI; in T2 weighted, *vice versa*. As for the DGM structures such as the thalamus and basal bangalia, they are identifiable at 25 GW as low signal on

T2 weighted images, and then become isointense after 24 GW on both T1 and T2 weighted MRIs. At approximately 35 GW, they can only be identified as one union of high signal in T1 weighed MRI. This contrast change with age adds another reason why fetal or premature neonatal MRI can not be generalized to one single template.

WM Myelination in MRI As previously introduced, WM myelination is one of the most significant features of fetal and neonatal brain maturation [84]. We have discussed its physiology background. Here I present how WM myelination shows up in T1W premature neonatal brain MRI scans in Figure 2.8. As expected, WM myelination does not exist until around 37 GW. In the 38.4 GW scan, we can clearly see the myelination already started and a part of the Globus Pallius is already myelinated. Through 38 to 45 GW, the progression of WM myelination in the posterior limb of internal capsule (PLIC) can be clearly visualized. Even though it is reassuring to visualize the expected changes with age in the MRI, it can cause challenges in both manual delineation of the brain structures as well as automatic image processing pipeline. Apparently, this violates the assumption that the MR intensities of one structure is uniform throughout the image if we consider it as part of WM. If the myelinated WM is considered as a separate structure apart from WM, then a manual labeling protocol needs to be first developed. This can also be quite challenging due to its less clear boundary. Therefore, the effect of WM myelination on image processing pipeline is still to be studied.

2.2.2.2 Imaging Artifacts

Most common imaging artifacts in neonatal brain MRI that are different from adult brain imaging include bias artifacts, motion artifacts, partial voluming etc. These artifacts mainly result from either lack of dedicated neonatal imaging techniques and equipment, or constraints posed by neonates as subjects. Here we look at two major artifacts which cause challenges in quantitative analysis of premature neonatal MRI.

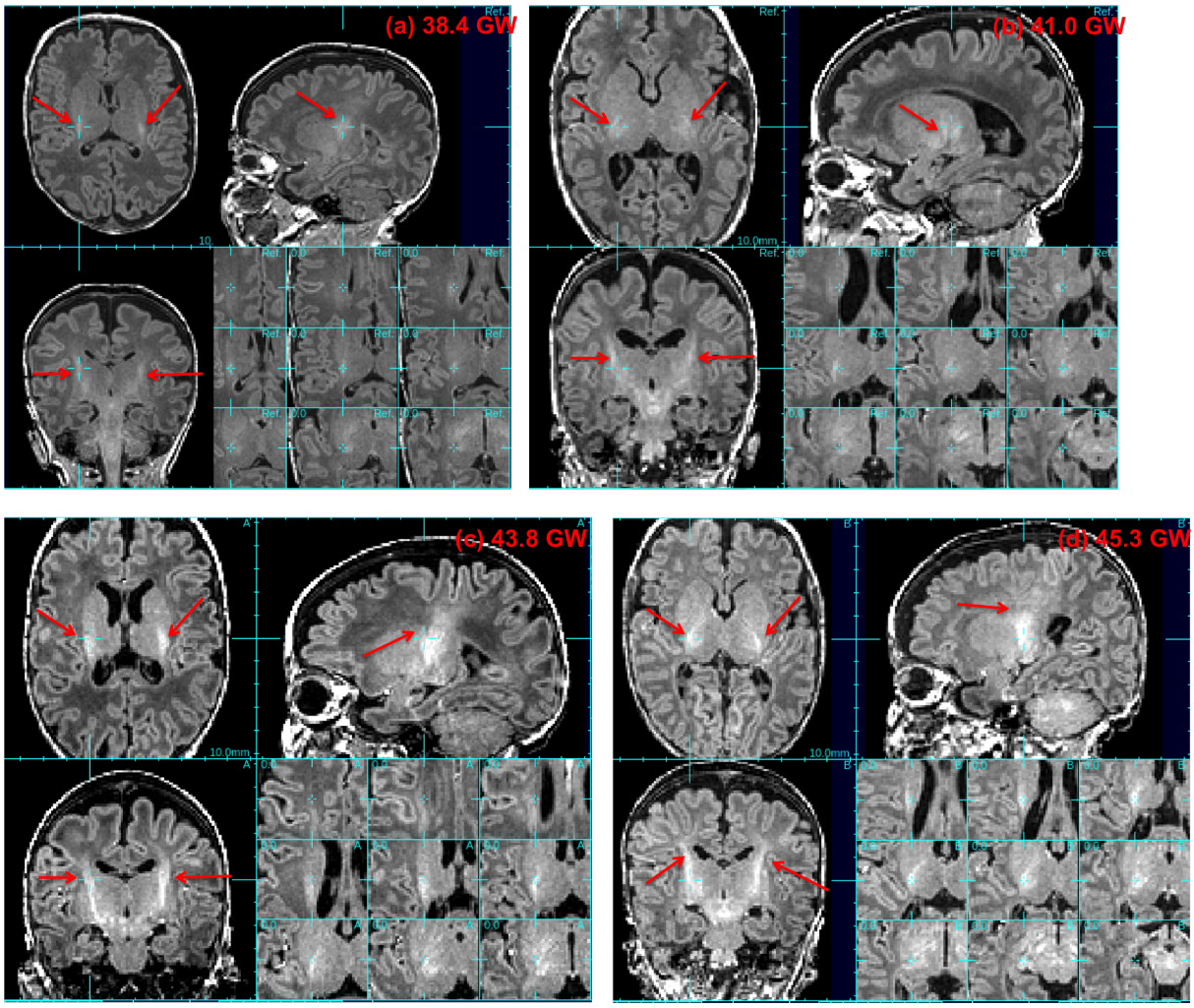

Figure 2.8: WM myelination progress shown in T1W MRI from Vancouver dataset. Gestational ages are marked on each scan. Red arrows: WM myelination and its progress.

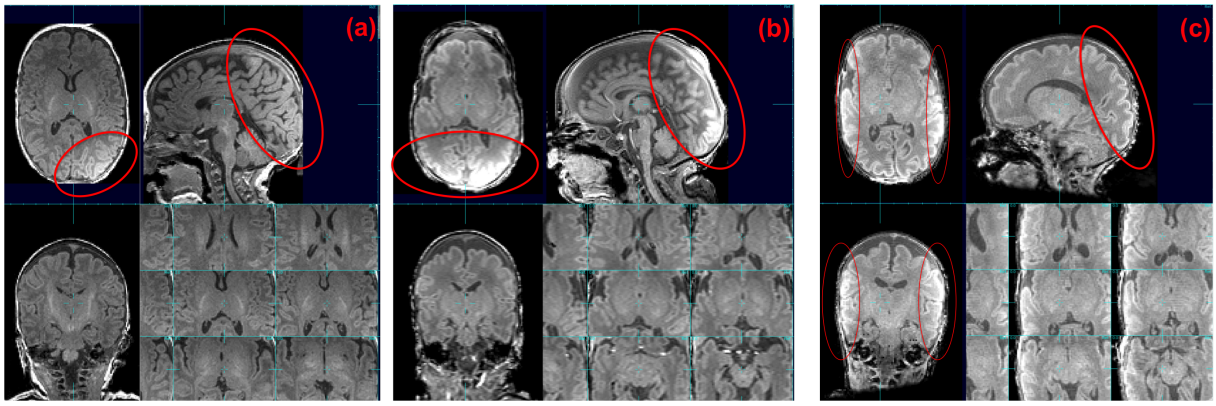

Figure 2.9: Three examples of bias-corrupted premature neonatal MRI with different levels and features of bias artifact. Red circle: abnormally high intensity caused by bias artifact.

Intensity Inhomogeneity The intensity inhomogeneity in MRI is also referred to as "bias artifact" [50]. Bias artifact is a result of uneven coil sensitivity and patient position, rather than anatomical differences. Locations closer to the coil will have higher intensity than locations further away from the coil. Bias artifact creates challenges for most image processing tasks. Most image analysis methodologies assume that the MR intensities of a given structure are relatively uniform throughout the image. Therefore, bias artifacts will result in confusion with intensity level of different structures. For example, in Figure 2.9(b), WM in within the red circle has similar intensity to GM outside the circle, which can easily cause confusion when the algorithm determines the structure based on its MR intensity.

Bias field is referred to a map of the continuously varying spatial signal change. Bias correction is the process to estimate the bias field and correct for it. Due to its importance yet relative simpleness, bias correction has been extensively investigated and is often an extra step in the image processing pipeline. Details will be discussed in later sections.

Motion Artifact In neonatal studies, especially premature neonates andor pathological cases, due to safety concerns, sedation is always avoided. Due to inability to control their

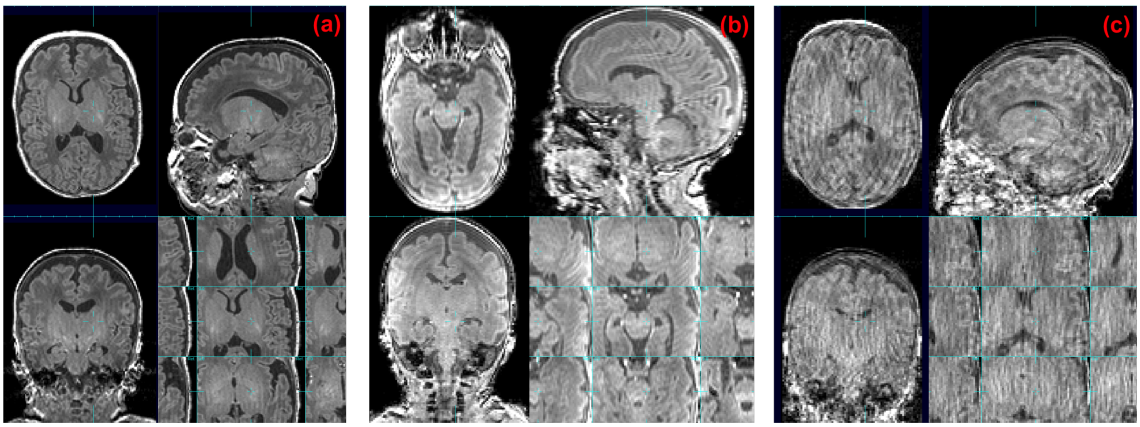

Figure 2.10: Examples of premature neonatal brain MRIs with (a) mild, (b) moderate and (c) severe motion artifact. Generally speaking, images like (a) and (b) have a higher probability to be correctly processed than (c).

head movement as adults, it is almost impossible to have a premature neonate subject remain still during the acquisition. Head movements can be, to some extent, restricted by coils with limited space. By using a dedicated neonatal head coil, motion artifacts can be effectively reduced.

However, motion corruption remains one of the most common image artifacts in neonatal imaging. Motion artifacts show in the image as circular bands of similar curve with the shape of the brain. Figure 2.10 shows 3 examples of motion artifacts at different severity levels. As one would expect, the more severe the motion artifact, the more challenging it is to be accurately processed. Generally speaking, given reasonable tissue contrast, images with mild to moderate motion artifacts can be processed as normal, while severe motion artifact can render the image unusable. In special studies, such as studies of WM sub-layers, motion artifacts can be mis-identified as sub-layers and thus pose an even larger challenge.

Wrap-around Artifact Another artifact that we encounter is the warp-around artifact, as shown in Figure 2.11. Wrapping artifact is a result of non-optimal field-of-view (FOV) setting

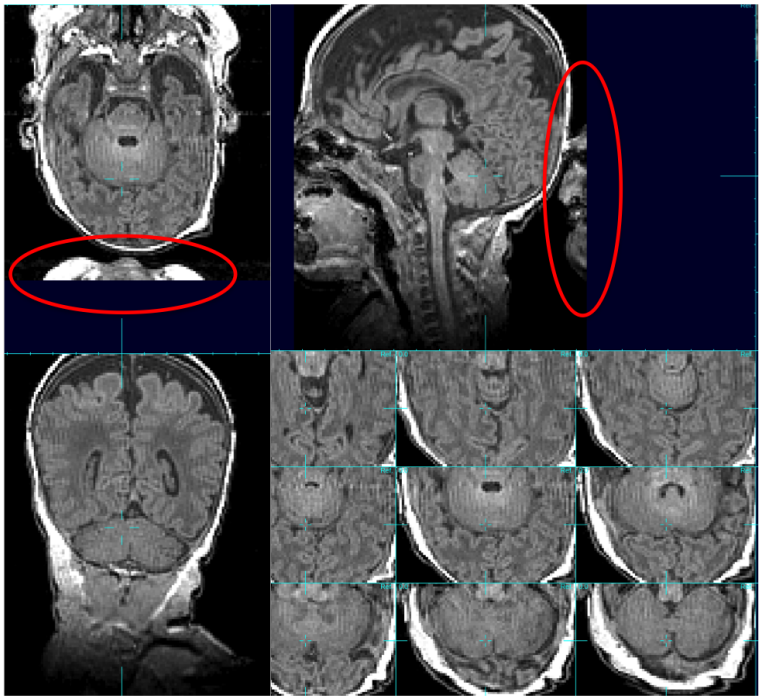

Figure 2.11: An example of wrap-around artifact from Vancouver dataset. The face of the subjects is wrapping posterior to the back of the head in MRI.

during acquisition. It happens when the dimensions of the imaged object exceeds the defined FOV. This artifact is easy to recognize and can be corrected in time during acquisition. Therefore, we do not see any severe wrap-around artifacts that affect the region-of-interest. For example, in Figure 2.11, the wrap-around face does not affect any quantitative analysis of brain structures.

2.2.2.3 Lesions in MRI

As discussed above, preterm births are often accompanied by brain lesions. These lesions show up in scans right after the preterm births, as opposed to later in life such as term-equivalent ages. With respect to the challenges they pose on the automatic segmentation and quantitative analysis, we divide them into two categories: focal and diffuse lesions.

Focal Lesions Focal lesions are referred to as an abnormal structure that is spatially limited and dense, with a relatively clear boundary. Focal lesions are generally easy to visualize by medical imaging techniques and relatively easy to manually delineate the boundary. Therefore, they remain one of the areas of focus in quantitative studies. For example, Intraventricular hemorrhage (IVH) (Figure 2.12(c)) is one type of focal lesion and can often be accompanied by an enlargement of cerebral VENT, termed Ventriculomegaly (VM) [139]. When these blood clots die away, these space-occupying lesions will commonly cause void of normal tissue structures (Figure 2.12(d)), leading to affected or missing neurobehavioral and neurocognitive functions. However, not all focal lesions can be so easy to recognize. In Figure 2.12(a), an example with peri-ventricular leukomalacia (PVL) in WM is given. In such cases, the lesions are even difficult to be recognized by human eyes. We can conclude that focal lesions can be an existing of abnormal structures or an absence of normal structures.Therefore, it is important, but at the same time very challenging, to quantify these focal lesions and understand their effects on brain development.

Diffuse Lesions Another type of lesion that we often observe in premature neonatal brain MRI scans is referred to as "diffuse lesion". Diffuse lesions are abnormalities without a clear boundary, and thus difficult to quantify. A common example is another type of white matter injury (WMI). Different from Figure 2.12(b), diffuse WMI often shows as a region within WM with abnormally low MR intensity. It is considered as abnormal tissue instead of missing tissue. As we can see in Figure 2.13, this type of lesion is extremely difficult to delineate, even with human eye. Different adjustments of the greyscale window can be misleading to the human eye when determining the boundary. Therefore, this remains one of the least studied areas in quantitative neonatal brain studies. In this work, we did not directly study this type of lesion. But it remains a significant part in developing robust automatic tissue segmentation algorithms because they pose significant challenges to the performance of automatic delineation of tissue structures in the MRI. Because of their unexpected intensity, it is very difficult for the algorithm to find a correct match from the template. We will

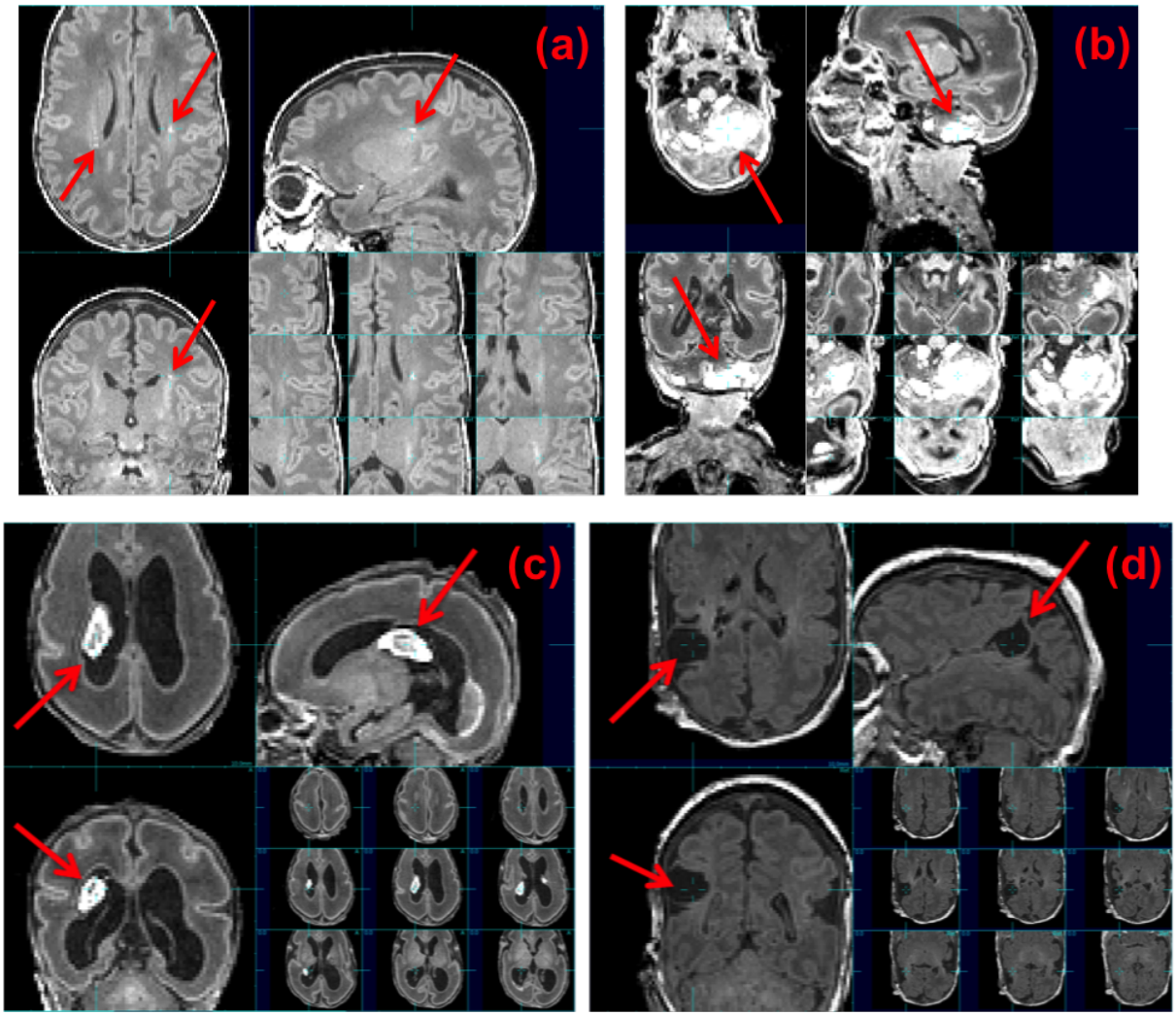

Figure 2.12: Four examples of focal lesions. (a) Peri-ventricular Leukomalacia (PVL); (b) Blood clot in CBL; (c) Intra-ventricular hemorrhage (IVH) with ventricumegaly (VM); (d) Cystic white matter injury (WML).

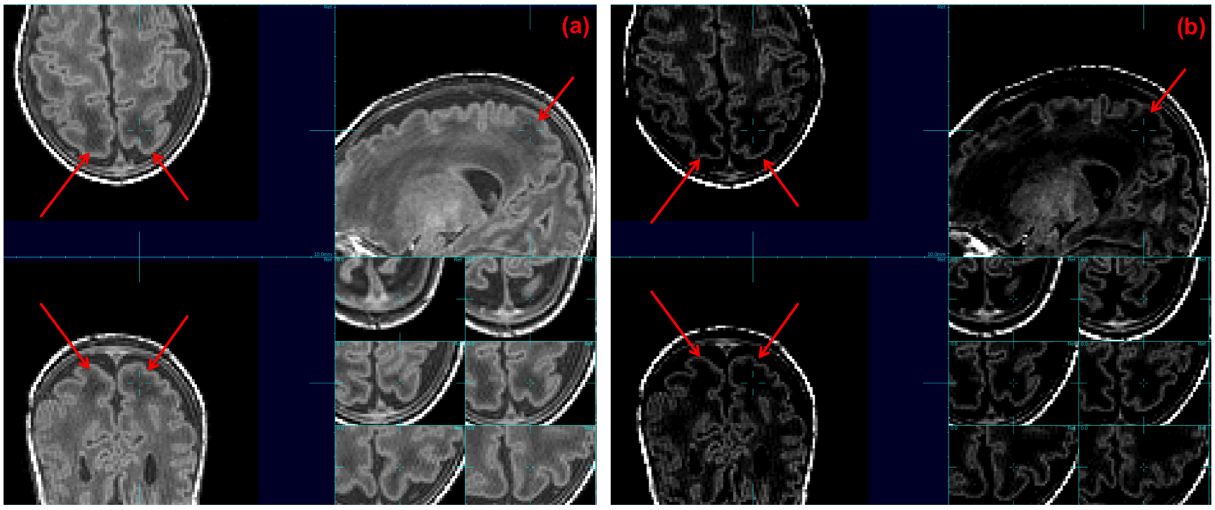

Figure 2.13: An example of diffuse WMI with different grayscale adjustment. (a) A good grayscale setting can visually show the diffuse WML; (b) An extreme example of bad grayscale setting where WML is invisible. Red arrow: diffuse WML.

further discuss this in the following sections.

2.3 Automatic Brain MRI Tissue Segmentation

2.3.1 Problem Statement

Automatic brain MRI tissue segmentation describes a task to automatically identify and delineate the different anatomical structures in a brain MRI scan using algorithms and automated image processing pipeline. In a brain MRI scan, as we discussed above, some different structures with different water content show different level of grey scale value, i.e. the MRI intensity. Naturally, the identification and delineation of these structures will depend highly on their unique intensity level. These structures include GM, WM, fluids including VENT and sCSF. Nonetheless, some structures share similar intensity level. But due to its unique location in the brain, it is easy to correctly identify and delineate them based on both its location and intensity. These structures include DGM, CBL and BS which share similar intensity level with WM, and VENT vs sCSF which share the same intensity

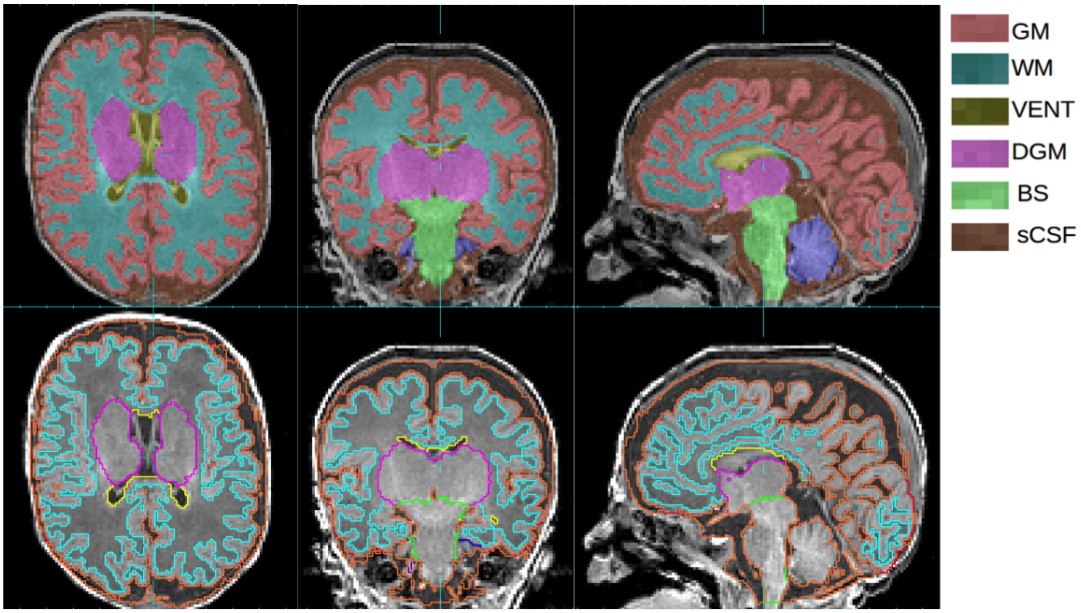

Figure 2.14: A manual labeling of seven tissue structures demonstrating the expected result of an automatic brain MRI tissue segmentation pipeline.

with each other. The problem of automatic brain MRI tissue segmentation is to develop automated algorithms and processing pipeline that take advantages of the MRI intensity and the location information to correctly identify and delineate the anatomical structure in the scan.

2.3.2 Methods

In the past decades, various approaches have been proposed for automated tissue labeling of brain MR images [148, 95, 152, 49]. Significant efforts have been made in developing an atlas-based approach [29, 99, 100]. In such methods, an atlas of neonatal brain anatomy is generated from manual or semi-automated labeling and used as prior knowledge for automated tissue labeling. However, errors in aligning new images to the atlas can induce errors that may be correlated with changes of interest in the anatomies being studied. To reduce such errors, multi-atlas label fusion algorithms were proposed [4, 8, 113]. Label fusion strate-

gies such as Majority Voting (MV) and the STAPLE algorithm [143] have been extensively investigated. However, such approaches still rely heavily on having manually delineated atlas subjects that match the subjects being studied. These methods can be limited in cases of more extreme developmental abnormalities and often pose challenging registration problems between individual and atlas. Related to multi-atlas methods, developments in image denoising prompted a patch-based approach for automated labeling [30, 109, 38]. This method does not require an accurate nonrigid registration and aggregates multiple label candidates in a neighborhood from the textbook images based on local similarities.

Studies also aimed to achieve a more detailed automatic labeling of brain tissues and regions. Gousia et al[48] evaluated and compared an atlas-based and a label fusion-based approach to automatically segment neonatal brain MR images at term age into 50 regions. Makropoulos et al[83] further pushed ahead to a wider age range using an Expectation-Maximization (EM)-based framework and showed promising results. However, both T1w and T2w images were used and only relatively healthy subject scans were shown.

2.3.2.1 Atlas-based EM Methodology

One of the most commonly used approaches for brain MRI tissue segmentation is the atlas-based EM tissue labeling method [144, 77, 9, 18]. In such approaches, a spatial atlas was constructed from a set of manually delineated MR scans and used to provide the essential spatial information in an Expectation-Maximization (EM) algorithm [32] based on the voxel intensity clustering of MRI[144, 77, 9]. The conventional EM-based tissue labeling approach is based on a Gaussian Mixture Model (GMM) of the observed MRI intensities, where the intensity values of each tissue class are assumed to have a Gaussian distribution, whose parameters can be estimated based on the Maximum Likelihood Estimator (MLE) using the EM algorithm [123]. The EM algorithm is dependent on *a priori* information from a spatial prior, due to overlap in MR intensity ranges of different tissue classes [18]. One unique example in T1-weighted premature neonatal brain MRI is the similar intensity level exhibited by the myelinated white matter and gray matter. Here, without a spatial prior,

the EM algorithm will be unable to label the tissues based on MR intensity alone. In cases of rapidly growing anatomies, such as in fetuses [53, 54] and premature neonates[81, 82], an additional dimension of time is incorporated into the atlas and a spatio-temporal atlas can be constructed from a set of manually delineated MRIs covering the age range of interest. For each new MR scan to be segmented, an age-specific anatomy that corresponds to the new subject's age is synthesized from the spatio-temporal atlas and used to provide the essential spatial prior in the EM algorithm.

2.3.2.2 Patch-based Dictionary Learning Technique

Introducing a patch-based approach into other segmentation frameworks, for example, probabilistic approaches, has also been proposed and studied. Dictionary learning refers to the process of learning a dictionary of patch items that can represent the general anatomy to label, and finding the matching items from this dictionary for an unseen patch to label. A probabilistic patch-based label fusion model for multi-atlas segmentation was first proposed by Bai et al[12] and evaluated on cardiac MR images. Wu et al[147] and Asman et al[11] further developed such an approach for the adult brain MRI and thyroid segmentation task. Initial attempts to apply such method on neonatal brain MRIs have been made. Li et al[142] evaluated a patch-driven level sets approach on normal neonatal T1w images. However, the tissue labeling task was not detailed enough since only GM and WM were labeled, and the age range was limited to around term age (41.5 ± 1.7 GWs).

2.3.3 Evaluation

The accuracy of automated segmentation is most commonly evaluated by comparing against a ground truth, the manual labeling. Dice Similarity Coefficient (DSC) [33] is often used to quantify the similarity between the manual and automatic segmentation. The DSC between region E and F is defines as [33]:

$$DSC = \frac{2 \times |E \cap F|}{|E| + |F|}. \tag{2.4}$$

Other quantitative measurements include Hausdorff Distance (HD) and Mean Distance (MD). HD is defined as:

$$HD(E,F) = max\{sup_{e\in E} inf_{f\in F} dist(e,f), sup_{f\in F} inf_{e\in E} dist(e,f)\}. \quad (2.5)$$

where dist(e,f) represents the distance between point e and point f. MD is define as the mean of pairwise distance on the surface of E and F.

In addition to HD and MD between two unit tissue structures, it is also interesting to quantitatively evaluate the accuracy of the cortical surface by calculating the HD and MD between two GM outer surfaces. As introduced above, each hemisphere can be divided into 4 lobes. With this division, quantitative evaluation of the accuracy of the lobe surface can also be performed to analyze if any part of the cortical regions are particularly difficult for the algorithm to label.

2.4 Challenges in Premature Neonatal Brain MRI Tissue Segmentation

One of the key challenges in automated analysis of premature neonatal MR data is the rapid anatomical changes with age [82, 99]. The age-dependent differences that can be observed in MR scans include the size and shape of the brain due to brain growth, as well as the changes in MR intensity contrast caused by brain maturation [53].

Another key challenge in automated delineation of brain tissues in premature neonatal MRIs is the large inter-subject variation or, in extreme cases, the occurrence of abnormalities. Although the addition of a parametrized spatio-temporal atlas can provide a more age-specific and thus accurate tissue prior [53], the method can still fail to accurately segment the MRI anatomies that exhibit large anatomical variations that are not well captured by an average tissue atlas derived from a training dataset [123]. Especially in clinical studies of brain injuries occurring from preterm birth, it is a challenge to completely capture a model of all possible variations within a new subject using a finite training dataset. If the individual MRI to be segmented differs from all atlas subjects, then the accuracy of the prior is dependent on the estimation of a non-rigid mapping to align the new individual anatomy to the atlas

prior [82]. However, this problem cannot be simply solved using a finer scale alignment of the subject and atlas MRI intensities. A very fine scale mapping may resolve normal geometric differences between the average prior and the subject, but may also incorrectly adapt and remove fundamental pathological differences that should be preserved and labeled. For example, when regional tissue contrast is modified by brain injury, fine scale registration may simply remove or shrink these differences so that they better match the atlas prior anatomy, rather than preserving the geometry of the differences, such as lesions, damaged tissues or regions of high contrast blood clot and their location in the surrounding normal anatomy. This problem becomes particularly challenging in brain tissue segmentation in the premature neonates, due to both extreme shape variation in normal tissue boundaries (e.g. ventricle size and shape as in Fig. 2.15) to be aligned to the statistical average, and the significant variation in local tissue integrity and intensity (e.g. abnormal white matter intensity as in Fig. 2.15) that need to be preserved. Cardoso et al. proposed an adaptive MAP-EM-based segmentation algorithm (AdaPT) especially for such cases in preterm neonates and showed satisfactory results in cases of ventricumegaly[19]. However, the age range covered by their dataset was limited to 35.7 to 44.3 GWs (40.4 $\pm$ 1.74 GWs), and thus failed to include evaluation on cases of more variations shown at younger ages.

Last but not least, the location, shape and size of focal lesions can be highly variable. Thus, it is not possible to approach the problem using whole brain learning from large sets of example images, as enough images to cover all possible configurations of ventricle and IVH location and extent are not available.

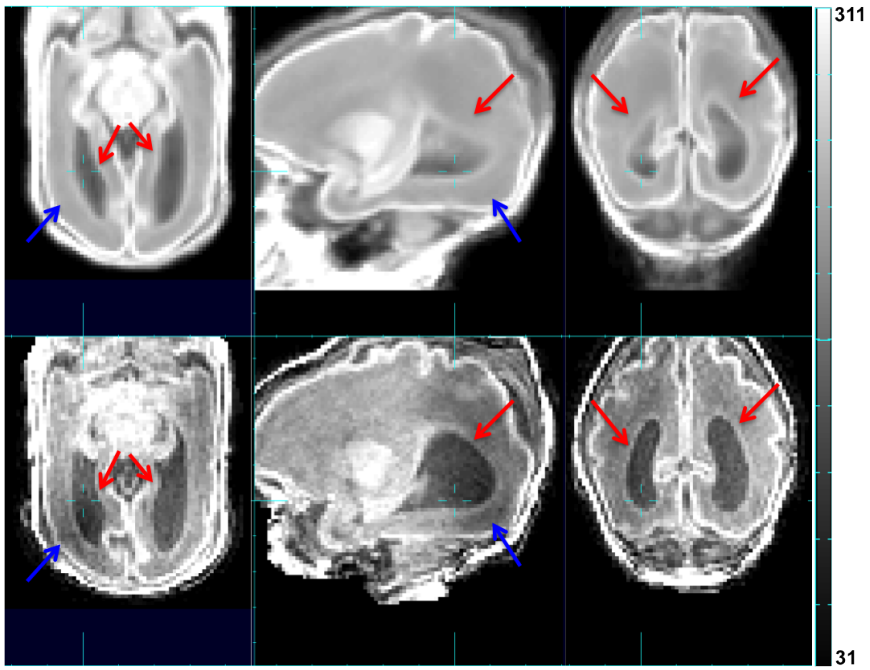

Figure 2.15: Comparison between the age-specific average image warped into subject space (upper row) and the subject MR image (lower row). Red arrow: difference in ventricle size and shape. Blue arrow: abnormal white matter intensity.

Chapter 3

PBAEM USING WEIGHTED FUSION (PBAEM-WF) FOR HEALTHY PREMATURE NEONATAL BRAIN MRI TISSUE SEGMENTATION

3.1 Introduction

In this work, we propose an approach that makes the best use of both an accurate parameterized atlas prior and a patch-based prior in the EM segmentation framework. To address this goal we derive a method that i) learns the location of greatest errors in the atlas-based tissue segmentation based on training data, and ii) adapts the priors in these locations to make use of priors derived from a local patch-based dictionary search of the atlas data. This hybrid approach should then provide a more robust prior that is derived from both an anatomical atlas average and a local tissue match for the EM segmentation. This allows a balance between the accuracy of mapping between subject anatomy and atlas, and the intensity similarity assumptions for patch searches in different regions of the brain. The proposed approach can be generalized as a patch-based augmentation of EM (PBAEM) with specific application to parametric atlas-based segmentation. We show that such method can produce more accurate automated tissue labeling, especially in cases of large inter-subject anatomical variation, compared to either the conventional atlas-based and patch-based approaches. In the next sections, we describe the approach in detail and its validation on a set of premature neonatal brain MR images with a high level of anatomical variation.

The contributions of this work are five-fold.

- We proposed a novel method to achieve accurate automatic segmentation of premature neonatal MRIs at any given age, instead of any pre-set targeted age, especially before

term age. This is especially important in the clinical settings since clinical neonatal MR image may be acquired at any random time point based on diagnostic needs.

- We proposed a learning approach for adapting tissue probability from two sources in a probabilistic framework.

- The proposed model improves the automatic tissue segmentation of our previous study and outperforms current state-of-the-art techniques in the field.

- To the best of our knowledge, the premature neonatal brain MRI atlas constructed in this work covers the widest age range from 27.29 to 46.43 GWs.

- Our algorithm allows for the first time accurate brain T1-weighted (T1w) MRI segmentation into 5 tissue classes in premature neonates with a wide age range. Especially, the proposed approach shows superiority when dealing with large inter-subject variation in young premature neonates.

3.2 Methodology

3.2.1 Overview

In this section, we first briefly review the EM-based automatic tissue segmentation framework and define the "running prior" which is the focus of our work. Next, we introduce the process to synthesize atlas-based tissue probability from the spatio-temporal atlas, followed by the computation of patch-based tissue probability. Then we demonstrate how we spatially adapt the tissue probability prior using both the atlas- and patch-based tissue probability. Finally, we summarize the entire PBAEM-WF framework.

3.2.2 Preliminaries & EM-based Tissue Segmentation Algorithm

Let i be the voxel index of the image to be segmented, and y_i be the logarithm of the MR intensity value at voxel i. K is the number of tissue classes we aim to segment the image

into, and $k = 1 : K$ is the current tissue class. μ_k and σ_k^2 are the mean and variance of tissue class k in the GMM model. $G_{\sigma_k^2}(y_i - \mu_k)$ is the Gaussian probability density at voxel i given mean μ_k and variance σ_k^2, defined as

$$G_{\sigma_k^2}(y_i - \mu_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} \tag{3.1}$$

. $P(k)$ stands for the prior tissue probability of class k and $p(y_i \mid k)$ is the posterior tissue probability of class k given log-transformed voxel intensity y_i. The EM algorithm consists of an initialization step and an iterative process to estimate the GMM parameters [32]. To distinguish, the initialization prior is denoted as $P_{init}(k)$. The prior which is used in the iterations is referred to as the **"running prior"** and denoted as $P_{run}(k)$. The superscript of (0) indicates the variable values in the initialization step, and superscript (t) indicates the variable values at the t-th iteration of EM.

In addition, to account for MRI signal variations (i.e. bias fields), a polynomial model of the bias field is integrated in the EM iterative framework [54, 138]. The degree of the polynomial model gradually increases from zero to three upon EM convergence [54]. We assume a standard multiplicative bias model [138] and correct for the bias field based on current estimate of the GMM in each iteration [54, 138]. The log-transformed bias field at t-th iteration at voxel i is denoted as $bias_i$. Thus, the bias-corrected log-transformed MR intensity y_i^{bc} at voxel i is given by:

$$y_i^{bc} = y_i - bias_i \tag{3.2}$$

Moreover, following [54], a neighborhood dependency further is incorporated into model. A voxel's label is also dependent on its neighborhood: it should be more probable to be the same with its neighboring voxels than different. This is to eliminate the potential isolated voxel label produced by conventional EM segmentation scheme [54]. The neighborhood constraint term at voxel i is denoted as $P_n(k \mid y_i)$. This neighborhood-based prior can be fused with the combined tissue probability to form a spatially varying prior in the EM algorithm.

Given the definitions above, the EM-based tissue segmentation algorithm can be expressed by the following equations [32, 54, 138]:

Initialization:

$$P_{init}(k) = Prior(k), k = 1\colon K \tag{3.3}$$

$$y_i^{c^{(0)}} = y_i \tag{3.4}$$

$$p^{(0)}(k \mid y_i) = P_{init}(k), k = 1\colon K \tag{3.5}$$

$$\mu_k^{(0)} = \frac{\sum_i p^{(0)}(k \mid y_i) * y_i^{c^{(0)}}}{\sum_i p^{(0)}(k \mid y_i)}, k = 1\colon K \tag{3.6}$$

$$\sigma_k^{2^{(0)}} = \frac{\sum_i p^{(0)}(k \mid y_i)) * (y_i^{c^{(0)}} - \mu_k^{(0)})^2}{\sum_i p^{(0)}(k \mid y_i)}, k = 1\colon K \tag{3.7}$$

Iterations:

$$e - step\colon \qquad P_{run}^{(t)}(k) = Prior(k), k = 1\colon K \tag{3.8}$$

$$g^{(t)}(y_i \mid k) = G_{\sigma_k^{2(t-1)}}(y_i^{c^{(t-1)}} - \mu_k^{(t-1)}), k = 1\colon K \tag{3.9}$$

$$p^{(t)}(k \mid y_i) = \frac{g^{(t)}(y_i \mid k) * P_{run}^{(t)}(k)}{\sum_k g^{(t)}(y_i \mid k) * P_{run}^{(t)}(k)}, k = 1\colon K \tag{3.10}$$

$$m - step\colon \qquad \mu_k^{(t)} = \frac{\sum_i p^{(t)}(k \mid y_i) * y_i^{c^{(t-1)}}}{\sum_i p^{(t)}(k \mid y_i)}, k = 1\colon K \tag{3.11}$$

$$\sigma_k^{2^{(t)}} = \frac{\sum_i p^{(t)}(k \mid y_i) * (y_i^{c^{(t-1)}} - \mu_k^{(t)})^2}{\sum_i p^{(t)}(k \mid y_i)}, k = 1\colon K \tag{3.12}$$

where $Prior(k)$ generally represents the tissue probability prior of tissue class k from any source. In this work, we omit the neighborhood constraint on the $P_{run}^{(t)}(k)$ to obtain superior performance in older subjects. The initialization prior and running prior are the focus of our work and are explained in details in the following subsections.

3.2.3 Atlas-based Estimation of Tissue Probability

In this subsection, we first explain the construction of spatio-temporal atlas, followed by the steps to synthesize age-specific atlas-based tissue probability. In the conventional task

of tissue labeling for adult brain images, a spatial atlas is constructed from multiple manually traced training datasets and provides essential spatial information to distinguish tissue classes with similar intensity levels. However, at this stage of life, premature neonates experience rapid anatomical changes due to brain development. Therefore, the morphometry and location of certain tissue classes are dependent on age. To incorporate the temporal information in the atlas, we used a parameterized spatio-temporal atlas approach as proposed by [53]. Instead of a quadratic model used previously [53, 82], a cubic model was experimentally chosen due to the extended age range that is covered by the premature neonatal imaging atlas. The spatio-temporal atlas is a third-order temporal polynomial model of MRI intensity change with age, a third-order temporal polynomial model of linear size change, a third-order temporal polynomial model of local displacement and a third-order temporal polynomial model of tissue probability changes.

To obtain the atlas-based estimation of the tissue probability for a new subject MR scan, we first synthesize an age-matched MR intensity template (A-MR) and a tissue probability map (A-TPM) from spatio-temporal atlas to correspond to the new scan. The MR image of the new subject is firstly linearly registered to the age-specific MR using 12 parameters including translation, rotation, scaling and skewness. In this step, a mask is used to exclude the non-brain portions in the MR scan such as neck, nose etc. Then the MR image is non-rigidly registered to the age-specific MR template as in previous fetal brain segmentation[120, 53]. Based on the inverse of the estimated spatial transformation, we map the age-specific MR template and age-specific tissue probability map into the space of the new subject MRI. We define this as the subject-space age-specific MR template (SA-MR) and the subject-space age-specific tissue probability map (SA-TPM) respectively. The SA-TPM is the resulting atlas-based estimation of tissue probability for the new subject scan. In the atlas-based segmentation approach [81, 54], SA-TPM is used as a source of spatially varying priors in the EM algorithm.

In this paper, we denote the atlas-based tissue probability of class k given log-transformed voxel intensity y_i at voxel i as $P_a(k \mid y_i)$. Thus, in the EM segmentation algorithm with only

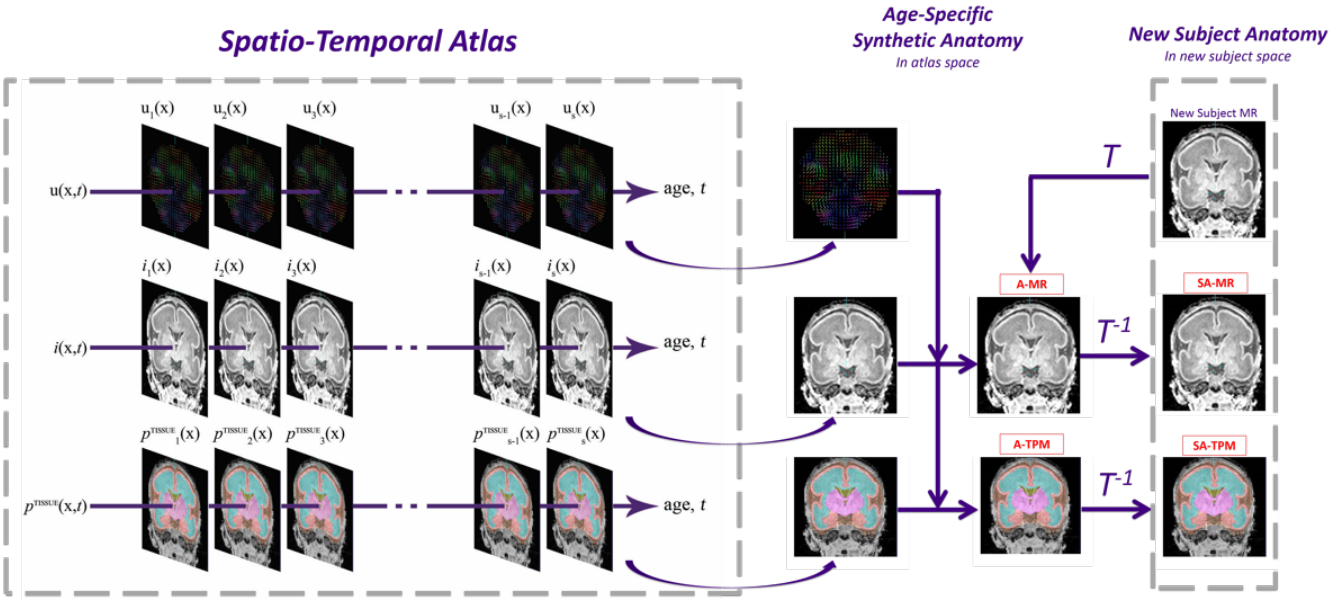

Figure 3.1: Illustration of synthesizing atlas-based tissue probability from the spatio-temporal atlas.

an atlas-based prior, Eqs. (3.2) should be replaced with:

$$P_{init}(k) = P_a(k \mid y_i), k = 1 \colon K \tag{3.13}$$

and Eqs. (3.7) is finally replaced with:

$$P^{(t)}_{run}(k) = P_a(k \mid y_i), k = 1 \colon K \tag{3.14}$$

Patch-based Estimation of Tissue Probability

Following the framework of previous patch-based segmentation approaches [109, 38, 30], for a given location in the subject anatomy, we carry out a local search of a reference dictionary image for feasible matches to the observed region of subject anatomy to be labeled. Unlike patch-based segmentation of adults [109, 38, 30], we do not simply search a single labelled subject anatomy textbook, but choose to search an age-specific estimate for that scan. This

is important in studies of rapidly developing brain anatomy where a single age representative matched training subject may not exist for a new MRI study, and it is more appropriate to search a model based representative average template synthesized for that age. This search is carried out across the aligned age-specific MR template (i.e. SA-MR) within a given search distance of the voxel to be labelled. The search range is assigned as a ratio of the total tissue volume, which is estimated according to the atlas-based TPM. This is to avoid any bias in search neighborhood size due to brain volume differences across ages. In order to eliminate differences in global contrast between the subject MRI and the SA-MR, we run a first-round atlas-based age-specific EM segmentation to obtain the bias-corrected MR image. We further scale its intensity so that the mean intensities within the brain between SA-MR and the new subject MRI are the same. A weight $w(i,j)$ between the patch to be labeled $H_I(i)$ centered at voxel i in the MR image to be segmented (denoted as I) and each patch $H_T(j)$ in the search neighborhood $N(i)$ centered at voxel j in SA-MR template (denoted as T) is computed as [109, 38, 30]:

$$w(i,j) = exp(-\frac{\sum_{i' \in H_I(i), j' \in H_T(j)} (y_{i',I} - y_{j',T})^2}{2p^3 \phi \varepsilon^2}) \tag{3.15}$$

where $y_{i',I}$ is intensity of voxel i' in image I; p^3 is the number of voxels in a 3-D patch with p being the 1-D patch size; ε is the standard deviation of the noise and ϕ is the smoothing parameter. In this manner, the computed weight of each sample is only driven by the similarity of the intensity between the two patches in the MR to be segmented and the SA-MR. Unlike a direct patch labeling scheme as in previous patch-based approaches[109, 38, 30], the proposed approach extracts the best tissue probability estimate for each class k from the dictionary image SA-TPM instead of the best label estimate for use in the EM optimization. A patch-based tissue probability for each possible patch is computed by a fusion of weighted tissue probabilities of patches within the defined search neighborhood[109, 38, 30]:

$$\widetilde{P_p}(k \mid y_i) = \frac{\sum_{j \in N_i} w(i,j) * P_a(k \mid y_j)}{\sum_{i \in N_i} w(i,j)} \tag{3.16}$$

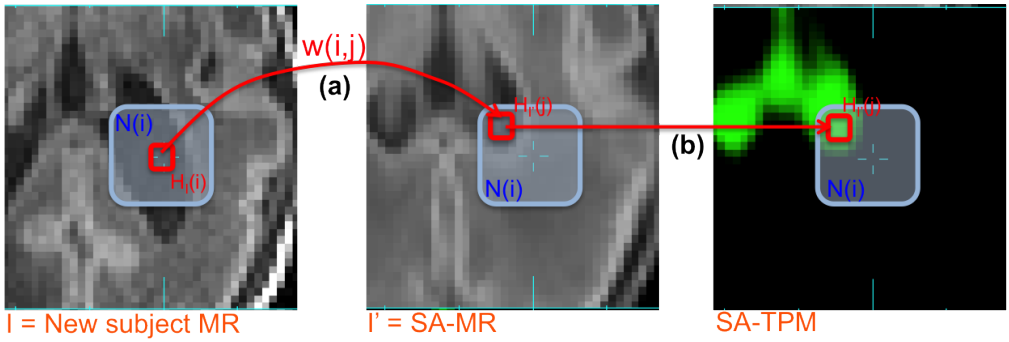

Figure 3.2: Illustration of patch-based local search adapted for using with spatio-temporal atlas. (a) For the patch $H_I(i)$ centered at voxel i, the local search is conducted within the search neighborhood $N(i)$ in SA-MR (I') and weight $w(i,j)$ between each possible pair of patches is computed. (b) Tissue probability of each possible patch within $N(i)$ is extracted from the SA-TP and used to compute the patch-based tissue probability of voxel i.

where $\widetilde{P_p}(k \mid y_i)$ is the computed patch-based tissue probability (unnormalized) of class k given voxel y_i; $P_a(k \mid y_j)$ is the atlas prior of voxel i's neighbor j in the corresponding SA-TPM. Figure 3.2 illustrates the patch-based local search.

In order to ensure we have a probability estimate at each voxel, such that $\sum_k \widetilde{P_p}(k \mid y_i) = 1$, we normalize the estimates at each voxel:

$$P_p(k \mid y_i) = \frac{\widetilde{P_p}(k \mid y_i)}{\sum_k \widetilde{P_p}(k \mid y_i)} \tag{3.17}$$

where $P_p(k \mid y_i)$ is the normalized patch-based tissue probability of class k given voxel y_i. Using this definition, voxels with similar surrounding neighborhoods are considered to have similar tissue probabilities [30].

To reduce computational time required for voxel by voxel patch evaluation during larger searches, we exclude patches centered outside the brain. Pre-computation of local image statistics in the reference SA-MR and subject MRI is carried out to allow a more efficient and accurate level of pre-exclusion of regions [30] of patch matches. In our study, this pre-selection procedure is based on the structural similarity measure (SSIM)[30, 38]:

$$ss = \frac{2m_{i,I}m_{j,I'}}{m_{i,I}^2 + m_{j,I'}^2} * \frac{2s_{i,I}s_{j,I'}}{s_{i,I}^2 + s_{j,I'}^2} \tag{3.18}$$

where $m_{i,I}$ is the mean and $s_{i,I}$ is the standard deviation of the voxel intensities within the patch centered at voxel i in image I.

In the EM segmentation algorithm using only an patch-based prior, Eqs. (2) should be replaced with:

$$P_{init}(k) = P_p(k \mid y_i), k = 1 \colon K \tag{3.19}$$

and Eqs. (7) is finally replaced with:

$$P^{(t)}_{run}(k) = P_p(k \mid y_i), k = 1 \colon K \tag{3.20}$$

3.2.4 Spatially Adapted Tissue Probability Prior

Using the framework described above, we now have two alternative probabilities for tissues at each location in the MRI scan to be segmented. In PBAEM-WF, we combine the two estimates into a single weighted tissue probability for each voxel to provide a better initial estimate for use as a running prior in the EM algorithm. For the weighting to be optimal, we account for the "Patch Contribution" (PC) to quantify how much new information is introduced by adding the patch-based estimates of tissue probability as well as its accuracy and reliability. Also, we implement a weighting for "Voxel Label Accuracy" (VLA) to quantify the overall performance of each tissue prior within the EM algorithm. To capture the age-specific feature of the spatially varying PC and VLA map, we adopted same approach from the spatio-temporal atlas to construct a spatio-temporal model of both PC and VLA maps, and synthesize age-specific PC and VLA maps for each new subject to be segmented. In the following paragraphs, details of learning these weightings are given.

Age-Specific Patch Contribution (PC) Map The contribution of the patch-based approach into the combined tissue probability depends on the intensity uniqueness on two levels: intensity structure within a patch itself and the patch similarity within the search neighborhood. If a voxel has a similar MR intensity to its neighbors, then the patch centered

at that voxel cannot uniquely define the local anatomy. Similarly, if a patch has a similar structure to its neighbors, then the search neighborhood cannot provide any new information. Therefore, measurements that characterize intensity uniqueness within patch and search neighborhood are essential to quantify the actual contribution of patch-based local search. We propose to measure the patch contribution per voxel in two parts: patch uniqueness (PU) and neighborhood uniqueness (NU). PU measures the intensity variance given image noise within the patch under consideration, while NU describes how much on average the intensities are different between the patch under consideration and all patches within its search neighborhood. Here we derive PU as an example.

The measured MR intensity of a voxel can be seen as its true value plus noise. The probability of two voxels' (i and i') actual intensities $\bar{y_i}$ and $\bar{y_{i'}}$ being the same given their measured intensities y_i and $y_{i'}$ and image noise variance σ^2 is:

$$P(\bar{y_i} = \bar{y_{i'}} \mid y_i, y_{i'}, \sigma^2) = exp\left(-\frac{(y_i - y_{i'})^2}{2 * \sigma^2}\right) \tag{3.21}$$

The probability of n voxels in a patch $H_I(i)$ centered at voxel i in image I having the same true intensities as voxel i is:

$$P(\bar{y_i} = \bar{y_{i'}}, i' \in H_I(i) \mid y_i, y_{i'}, i' \in H_I(i), \sigma^2) = \sqrt[n]{\prod_{i' \in H_I(i)} exp\left(-\frac{(y_i - y_{i'})^2}{2 * \sigma^2}\right)} \tag{3.22}$$

$$= \exp\left(-\frac{\sum_{i' \in H_I(i)} (y_i - y_{i'})^2}{2 * \sigma^2 * n}\right) \tag{3.23}$$

The n-th root is taken to eliminate any bias caused by the different number of voxels in a patch, especially in patches at brain boundary. Therefore, the patch uniqueness, which is the probability of n voxels in a patch having different intensities as the center voxel, is defined as:

$$PU(i) = 1 - \exp\left(-\frac{\sum_{i' \in H_I(i)} (y_i - y_{i'})^2}{2 * \sigma^2 * n}\right) \tag{3.24}$$

In the same manner, we derive NU as follows:

$$NU(i) = 1 - \exp\left(-\frac{\sum_{j \in N(i)} \sum_{i' \in H_I(i), j' \in H_I(j)} (y_{i'} - y_{j'})^2}{2 * \sigma^2 * n * N}\right) \tag{3.25}$$

where N is the number of patches within the search neighborhood of voxel i.

The patch contribution (PC) at each voxel is computed as the product of PU and NU:

$$PC(i) = PU(i) * NU(i) \tag{3.26}$$

To eliminate the intensity variation across the image as well as between subjects, all computations is conducted on bias-corrected MRIs with intensity normalized with respect to atlas. To construct a spatio-temporal model of PC maps across the age range, we follow the approach used for temporal modeling of tissue probabilities [53]. LogOdds representation of probabilities [53, 96] is adopted to ensure that the resulting model estimates of PC values are in the range of valid probabilities. The LogOdds, $\mathcal{O}(\bullet)$, of probability $PC(i)$ is defined as:

$$\mathcal{O}(PC(i)) = log\left(\frac{PC(i)}{1 - PC(i)}\right) = l(i) \tag{3.27}$$

The probabilities are then modeled in the LogOdds space using a polynomial model of designated degree [53]. To synthesize the age-specific PC map, the inverse LogOdds is computed as:

$$\mathcal{O}^{-1}(l(i)) = \frac{exp(l(i))}{1 + exp(l(i))} \tag{3.28}$$

Age-Specific Voxel Label Accuracy (VLA) Map From the perspective of the overall performances of the two tissue probabilities used as running priors in the EM labeling, we use our training data to learn where each estimate provides a more useful EM prior. To do this we evaluate the results of EM segmentation using the two different priors on the training data. Then we evaluate overall performance of each at every voxel using a leave-one-out cross-validation against the manual tracing for each subject in the training dataset. From this we create an overall Voxel Label Accuracy (VLA) map for both the atlas- and patch-based tissue probabilities. The VLA of each voxel is defined as follows:

$$VLA_a(i) = \frac{CL_a(i)}{TL(i)}, VLA_p(i) = \frac{CL_p(i)}{TL(i)} \tag{3.29}$$

where $VLA_a(i)$ and $VLA_p(i)$ are the VLA of voxel i of atlas- and patch- based tissue probability respectively; $CL_a(i)$ and $CL_p(i)$ are the number of scans that are labeled correctly via EM segmentation using atlas- and patch-based priors respectively; $TL(i)$ is the total number of scans segmented at voxel i. We follow the same approach to construct a temporal model of VLA maps in the LogOdds space and synthesize age-specific VLA maps for each new subject scan. These two maps allow us to quantify of the accuracy level of two tissue probabilities for each voxel in the MR image to be segmented.

Spatially Weighted Tissue Probability For tissue class k given voxel intensity y_i for voxel i, the combined tissue probability is the weighted average of atlas-based tissue probability $P_a(k \mid y_i)$ and the patch-based tissue probability $P_p(k \mid y_i)$:

$$P_w(k \mid y_i) = \frac{VLA_a(i) * P_a(k \mid y_i) + VLA_p(i) * PC(i) * P_p(k \mid y_i)}{\sum_k [VLA_a(i) * P_a(k \mid y_i) + VLA_p(i) * PC(i) * P_p(k \mid y_i)]}, k = 1 \colon K \quad (3.30)$$

3.2.5 Patch-based Augmentation of EM using Weighted Fusion(PBAEM-WF) Framework

In the proposed approach, we combine the desirable properties of both atlas- and patch-based approaches to provide an improved prior with which to initiate EM labeling of a new MRI scan. Experimentally, we found that best automated tissue labeling performance was achieved using patch-based tissue probability as the initialization prior and the combined tissue probability as the running prior in the EM algorithm. Therefore, in the initialization step, Eqs. (2) should be replaced with:

$$P_{init}(k \mid y_i) = P_p(k \mid y_i), k = 1 \colon K \quad (3.31)$$

In the iterations, Eqs. (7) should be replaced with:

$$P_{run}^{(t)}(k) = P_w(k \mid y_i), k = 1 \colon K \quad (3.32)$$

3.3 Experiments and Validation

3.3.1 MR Imaging Protocol & Image Preprocessing

Premature neonatal imaging was performed by our collaborators at University of British Columbia in Vancouver, Canada using a 1.5T Siemens scanner with a neonatal head coil. For each scan, a T1-weighted image was acquired using a 3D FLASH sequence with imaging parameters TR = 36ms, TE = 9.2ms and a voxel resolution of $1.04 \times 1.04 \times 1mm^3$. Then each MR image was manually segmented into six cerebral tissue regions: gray matter (GM), white matter (WM), ventricles (VENT), deep gray matter (DGM), sulcal CSF (sCSF) and non-brain background region (BG). The resulting tissue label maps were used to create the spatio-temporal atlas and also served as reference segmentation for accuracy mapping and validation.

3.3.2 Population

The test dataset used for validation consists of 32 T1-weighted brain MR scans of premature neonatal brains. The birth ages range from 24.86 to 31.43 GWs (27.92 ± 2.26 GWs) , and the ages at scan range from 27.29 to 46.43 GWs (35.04 ± 5.59 GWs). The male to female ratio of 32 training subjects is 16/16. Our dataset is one of the most long-standing manually-traced premature neonatal brain MRI dataset dating back to 2008 [107].

3.3.3 Parameter and Model Selection

We experimentally determined the best parameters for the patch search and algorithmic models. First, the temporal model was evaluated and selected. A cubic temporal model of local deformation, MR intensities and tissue probabilities showed superior performance over quadratic model. While, quadratic and cubic temporal models of PC and VLA maps showed comparable results. For the sake of consistency, we adopted the cubic model for all further experiments. In this work, we aimed to use the same patch search parameters for all tissue classes for the patch-based local search. To select the optimal parameters, a patch size of

$3 \times 3 \times 3$ voxels and $5 \times 5 \times 5$ voxels, and a search neighborhood size ratio from 0.0005 to 0.015 of total brain volume were tested. Results will be shown in the following section.

3.3.4 Validation

To evaluate the overall performance of the segmentation approach against the conventional age-specific EM, experiments were performed using a leave-one-out cross-validation due to the limited number of 30 manually traced scans. Each scan was segmented with the spatio-temporal atlas from the other 31 subject scans. The oldest scan (46.43 GW) was excluded from segmentation performance evaluation because its age lay beyond the maximum age of its correspondent atlas and thus, its anatomy could deviate greatly from the younger atlas due to the rapid anatomical change at its age. However, in the younger age range, we have adequate scans to allow us to test the youngest scan. To avoid excessive computations, the PC and VLA models were constructed from all training scans. Since the PC map does not depend on the spatio-temporal atlas, the PC model used all 30 training scans. However, the VLA model could only be constructed from 31 training scans due to the exclusion of the oldest scan. The comparison of overall performance between the conventional atlas-based segmentation approach and PBAEM-WF was made using 31 sets of automated and manual tissue labeling. To quantify the performance of the automatic segmentation, the Dice Similarity Coefficient (DSC) between manual and automatic tissue labellings was used.

3.4 Results

3.4.1 Patch Search Parameter Selection

We experimented with different patch search parameters to achieve the optimal performance of the PBAEM-WF approach. To determine parameter settings to optimize performance, we assessed the average DSC of PBAEM-WF automatic labeling of all 31 testing scans using patch-based tissue probabilities computed with different sets of parameters. A patch size of $3 \times 3 \times 3$ voxels and $5 \times 5 \times 5$ voxels, and a search neighborhood size ratio from 0.0005

to 0.015 of total brain volume were explored experimentally. Figure 3.3 shows the average DSC obtained using these sets of parameters. The average DSC stabilizes around a search neighborhood ratio of 0.003, regardless of the patch size. We chose the optimal search range ratio of 0.0025 allowing for optimal performance in cortical regions (e.g. GM, sCSF) and WM, while limiting the decrease of DSC in DGM. A patch size of $3 \times 3 \times 3$ voxels or $5 \times 5 \times 5$ voxels showed comparable results in GM, WM, VENT and DGM at search neighborhood ratio of 0.0025. However, given the different performances on the sCSF class, we chose a patch size of $3 \times 3 \times 3$ voxels to allow more improvement in sCSF. Therefore, the final parameters for the following experiments and results are a patch size of $3 \times 3 \times 3$ voxels and a search neighborhood ratio of 0.0025.

3.4.2 Atlas- and Patch-based Estimation of Tissue Probability

For each manually segmented subject, we constructed a corresponding spatio-temporal atlas using the other 31 scans in the dataset and synthesized A-TPM. Next, we transformed the A-TPM into the space of the scan to be segmented to obtain the SA-TPM. The first row in Figure 3.4 gives an example of SA-TPMs of 5 tissue classes for one scan. The patch-based tissue probability was computed using the following parameters: a patch size of $3 \times 3 \times 3$ voxels and a search neighborhood ratio of 0.0025. The second row of Figure 3.4 presents an example of patch-based tissue probability map of 5 tissue classes for one scan. Compared to the atlas-based tissue probabilities, the less defined tissue boundaries of patch-based tissue probabilities would allow the EM labeling more freedom to adapt to unseen anatomies in the new subject to be segmented. Figures 3.5, 3.6 and 3.7 illustrate the complementary strengths of these two tissue probabilities. The patch-based prior has superior performance over the atlas-based prior in cases of normal inter-subject variations in GM as shown in Figure 3.5. This is because each subject's unique cortical folding in GM cannot be fully captured by the atlas constructed from only a limited set of scans. Figure 3.6 shows a scan of a subject with ventriculomegaly, where the abnormally enlarged ventricles were not captured in the atlas constructed from healthy subjects. Therefore, the atlas-based approach fails to accurately

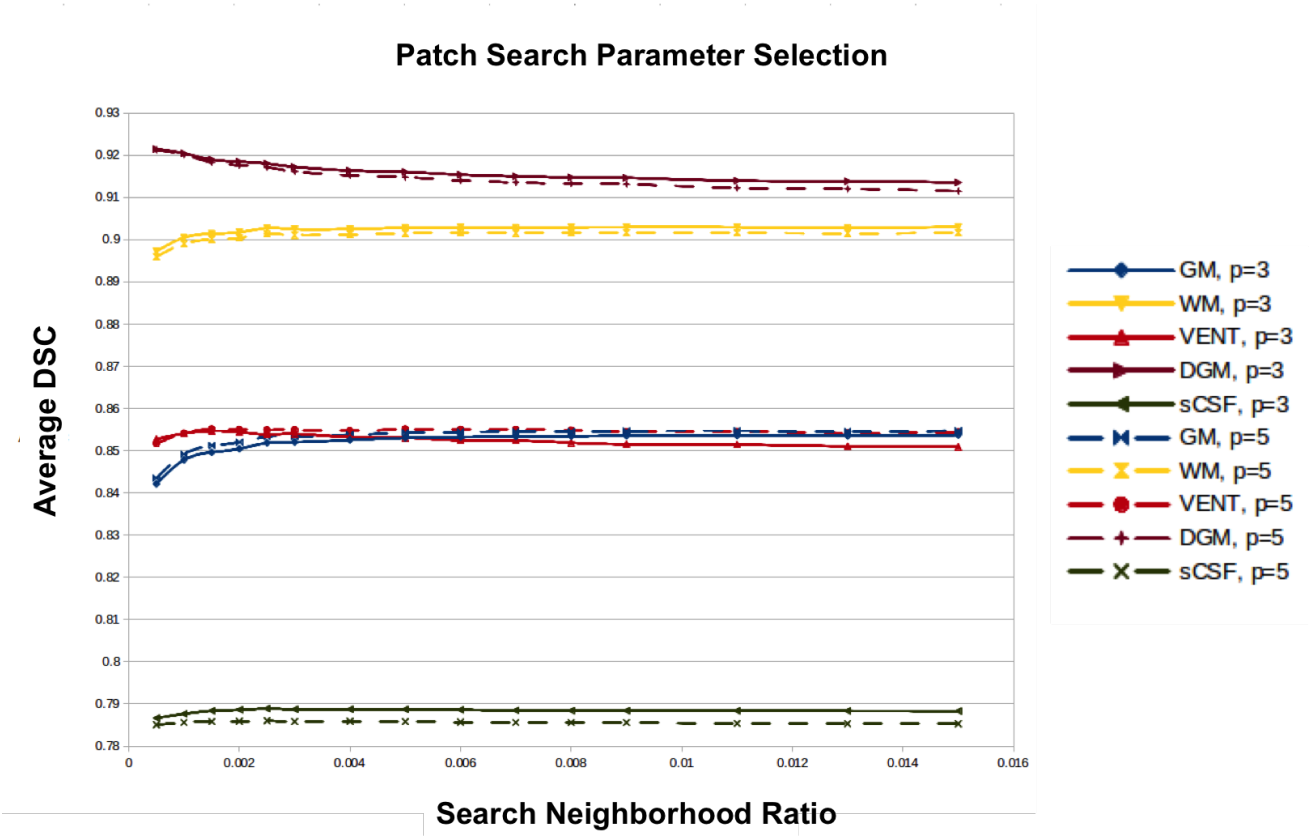

Figure 3.3: Average DSC of PBAEM-WF automated segmentation obtained using different sets of patch search parameters. Patch size of $3 \times 3 \times 3$ voxels or $5 \times 5 \times 5$ voxels is indicated by solid or dashed line.

estimate the tissue probability of the ventricular regions. However, the clear boundary of the ventricles allowed the patch-based approach to generate a more anatomically viable tissue probability estimation in this case. In tissue classes without clear boundaries or less variant inter-subject intensity differences, such as DGM as shown in Figure 3.7, atlas-based tissue probability estimation demonstrated higher accuracy than patch-based tissue probability.

3.4.3 Spatially Adapted Tissue Probability Prior

Age-Specific Spatially Varying PC Map The PC maps in Figure 3.8 demonstrates the amount of additional information that is introduced by the patch-based search. In cortical regions (e.g. sCSF/GM boundaries and GM/WM boundaries), the uniqueness of MR intensities of different tissue regions is high. Therefore, the intensity-based local patch search can better distinguish between tissues. This is reflected in the high PC values in these regions. In contrast, at WM/DGM boundaries where there is low MR intensity contrast, the patch based search could not reliably separate WM and DGM. The lower PC values here reflects the decreased reliability and utility of the patch-based search. The PC has its lowest value inside uniform regions, such as inside WM and DGM, where the patch-based tissue probabilities, by definition, are the same as the atlas-based tissue probabilities. Therefore, little additional information can be introduced by incorporating the patch-based search.

Age-Specific Spatially Varying VLA Map The VLA maps in Figure 3.9 illustrate the complementary strengths of atlas- and patch-based tissue probability estimates. We can see that at GM-WM boundaries, the patch-based tissue probabilities are more accurate (shown as higher intensity in the VLA map) compared to the atlas-based tissue probabilities, while at DGM-WM boundaries, they are less accurate. The patch-based estimates tend to be more accurate where there is a clear tissue boundary or the inter-subject variation is large, such as GM, WM, VENT and sCSF. In tissue classes like DGM whose boundary does not have high intensity contrast, or where little inter-subject variation is present, atlas-based estimation of tissue probability tends to be more accurate. VLA maps succeed in measuring

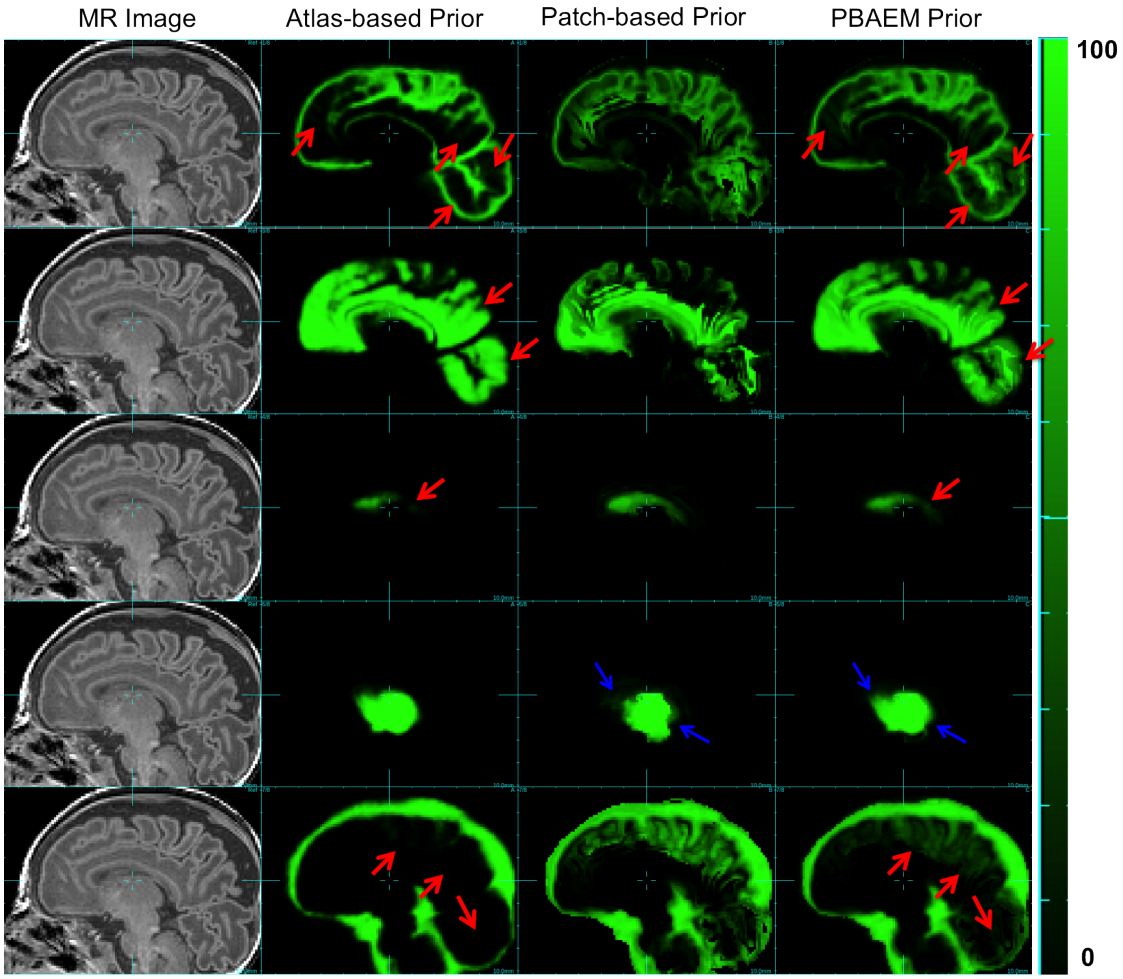

Figure 3.4: Examples of running priors. From left to right: Raw MR image (column 1), atlas-based (SA-TPM) (column 2), patch-based (column 3) and PBAEM-WF (column 4) tissue probability map of, from top to bottom, GM, WM, VENT, DGM and sCSF of one subject as an example. Tissue probability on a scale 0 - 100. Red arrows: PBAEM-WF tissue probability is more accurate than the atlas-based one; Blue arrows: PBAEM-WF tissue probability is more accurate than the patch-based one.

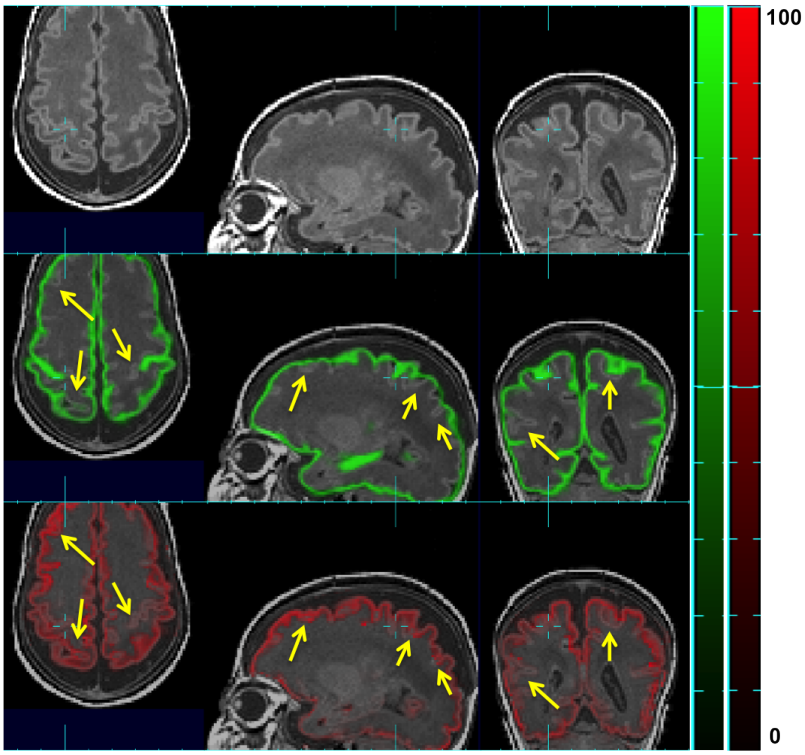

Figure 3.5: Raw MR image overlaid (top row) with atlas-based tissue probability map (SA-TPM) (middle row) and patch-based (bottom row) tissue probability map of GM in one subject. Tissue probability is on a scale of 0 - 100. Yellow arrows point where patch-based TP is more accurate than atlas-based one.

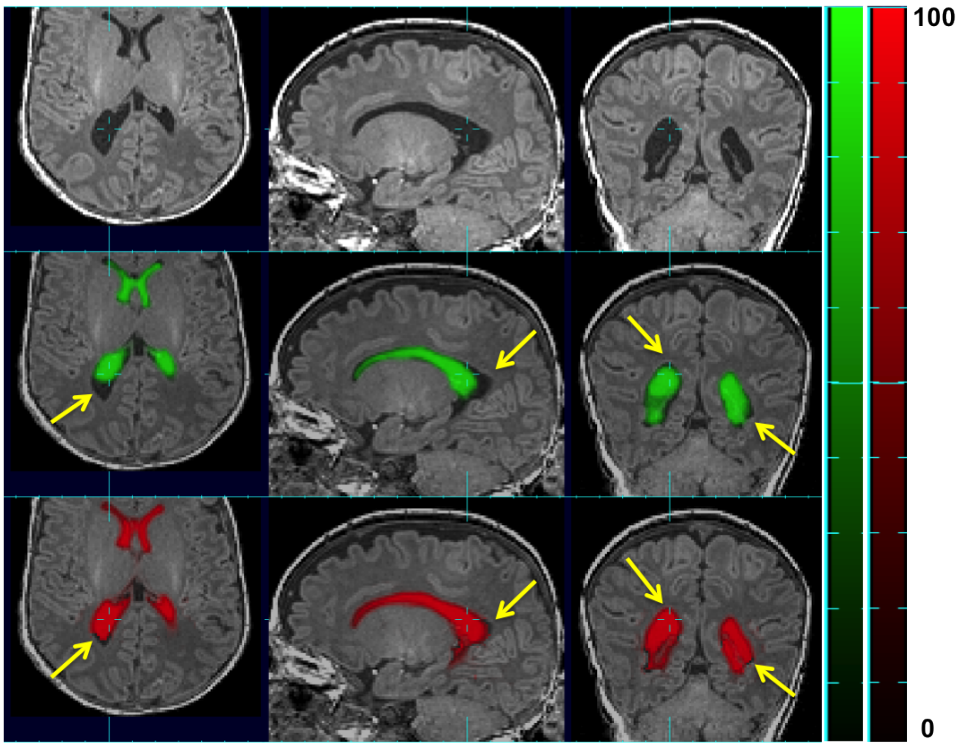

Figure 3.6: Raw MR image overlaid (top row) with atlas-based tissue probability map (SA-TPM) (middle row) and patch-based (bottom row) tissue probability map of VENT in one subject with ventriculomegaly. Tissue probability is on a scale of 0 - 100. Yellow arrows point where patch-based TP is more accurate than atlas-based one.

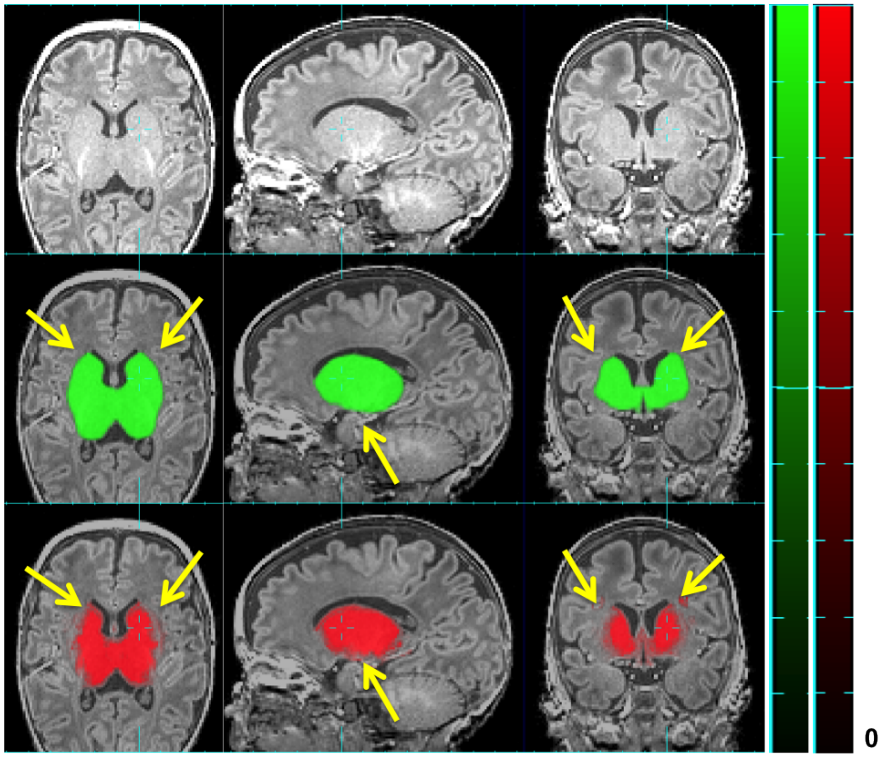

Figure 3.7: Raw MR image overlaid (top row) with atlas-based tissue probability map (SA-TPM) (middle row) and patch-based (bottom row) tissue probability map of DGM in one subject. Tissue probability is on a scale of 0 - 100. Yellow arrows point where patch-based TP is less accurate than atlas-based one.

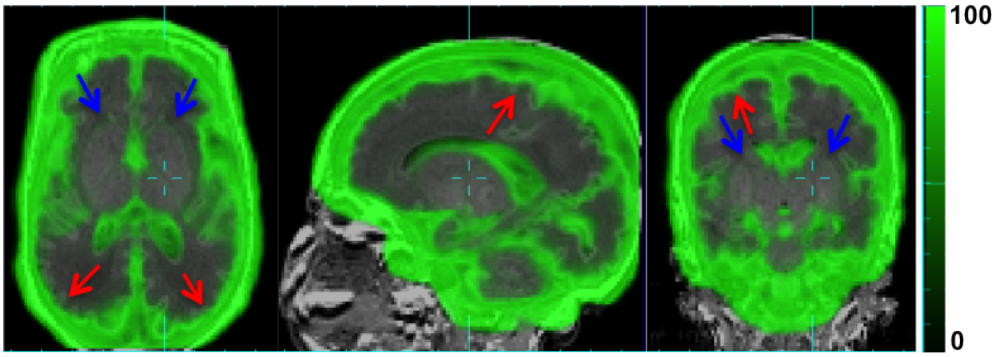

Figure 3.8: Age-specific Patch Contribution (PC) map (green) overlaid over raw MR image (grayscale). PC values are on a scale of 0 - 100. Red arrows: high PC values in cortical regions at sCSF/GM and GM/WM boundaries. Blue arrow: low PC values at WM/DGM boundary and inside WM, DGM.

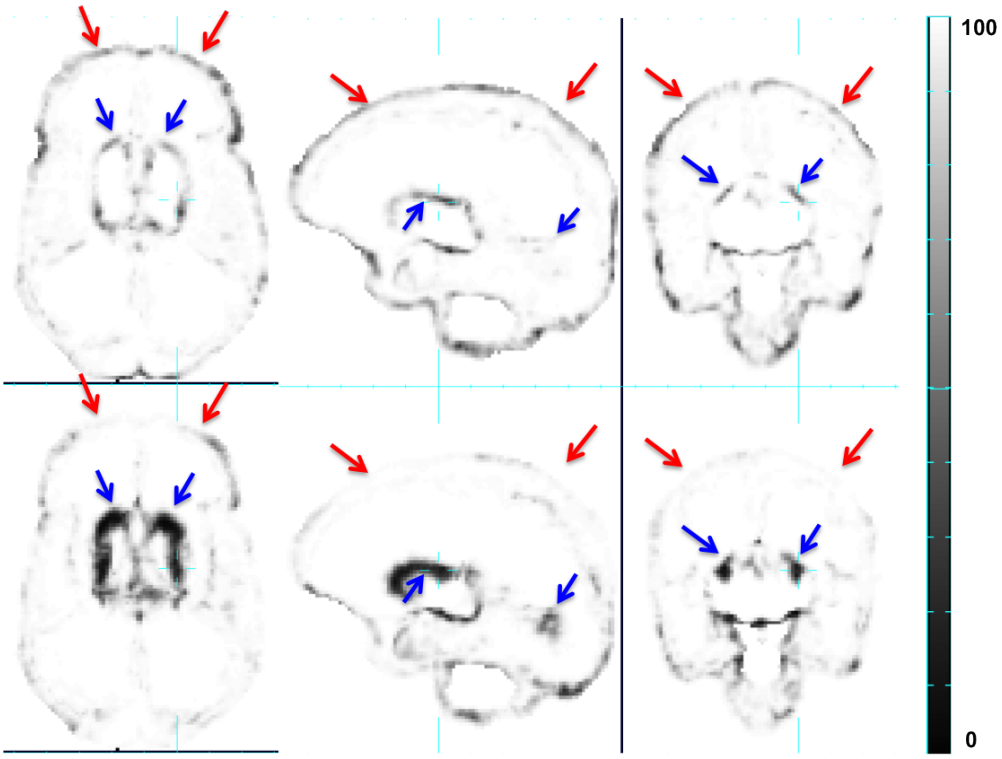

Figure 3.9: Comparison between Voxel Label Accuracy (VLA) maps resulting from EM segmentation using atlas-based (top row) and patch-based (bottom row) tissue probabilities as running priors. VLA values are on a scale of 0 - 100. Red arrows: patch-based tissue probabilities is more accurate at sCSF/GM and GM/WM boundaries; Blue arrows: atlas-based tissue probabilities is more accurate at DGM/WM boundaries.

these differences in accuracy levels on a voxel basis, allowing us to optimally combine the two tissue probabilities at a voxel level.

Patch-based Augmented Prior Figure 3.4 compares the spatially adapted tissue probability (bottom row) with the atlas- (top row) and patch-based (middle row) tissue probability map of 5 tissue classes of one subject. We show that the patch-based augmentation of atlas-based tissue probabilities takes advantage of the complimentary strengths of both the atlas-based and the patch-based methods, improving accuracy in GM, WM, VENT and sCSF compared to atlas-based TP and in DGM compared to patch-based TP.

3.4.4 Comparison with Conventional Age-specific EM Automated Segmentation

We compared the performance of PBAEM-WF with the baseline algorithm: conventional atlas-based EM tissue segmentation using spatio-temporal atlas[54]. Average DSCs for 5 tissue classes of the 31 scans were computed and compared among using atlas-, patch-based and PBAEM-WF prior in the EM algorithm. Table 3.1 shows the comparison of average DSCs between the conventional atlas-based EM segmentation and the PBAEM-WF approach. On average, the PBAEM-WF approach significantly ($p < 0.05$) improved the segmentation accuracy in GM, WM, VENT and sCSF compared to the conventional atlas-based EM segmentation. Although we see a significant ($p < 0.05$) decrease in accuracy in the DGM, the amount of decrease is less that 1%. This was due to the fact that our PBAEM-WF-based method improves the automated segmentation for tissues with high boundary contrast and large inter-subject intensity variation. However, the boundaries of DGM are less distinguishable and inter-subject intensities less variant than GM, WM, VENT and sCSF.

Figure 3.10 plots individual DSCs with age of five tissue classes. For most subjects, we can see an significant improvement in DSC in GM (26/31), WM (15/31), VENT (23/31) and sCSF (31/31). In addition, a trend of decreasing segmentation accuracy with age is shown in WM, while for GM, VENT, sCSF and DGM, comparable segmentation accuracy is obtained regardless of age. This is due to the fact that white matter myelineation and increasing level of cortical folding renders automatic segmentation increasingly difficult over developmental age.

Figure 3.11 gives an example of the improved automated labeling at GM-WM boundaries compared to the conventional atlas-based EM method. Figure 3.12 illustrates that our proposed PBAEM-WF approach produced a more anatomically correct automated labeling for the ventricular regions compared to the atlas-based EM segmentation approach. Table 3.2 shows the comparison between using EM with patch-based prior alone and with the fully optimized PBAEM-WF prior. It is clear that the segmentation accuracy improved in all 5 tissue classes using PBAEM-WF compared to only using the patch-based prior. Table 3.3

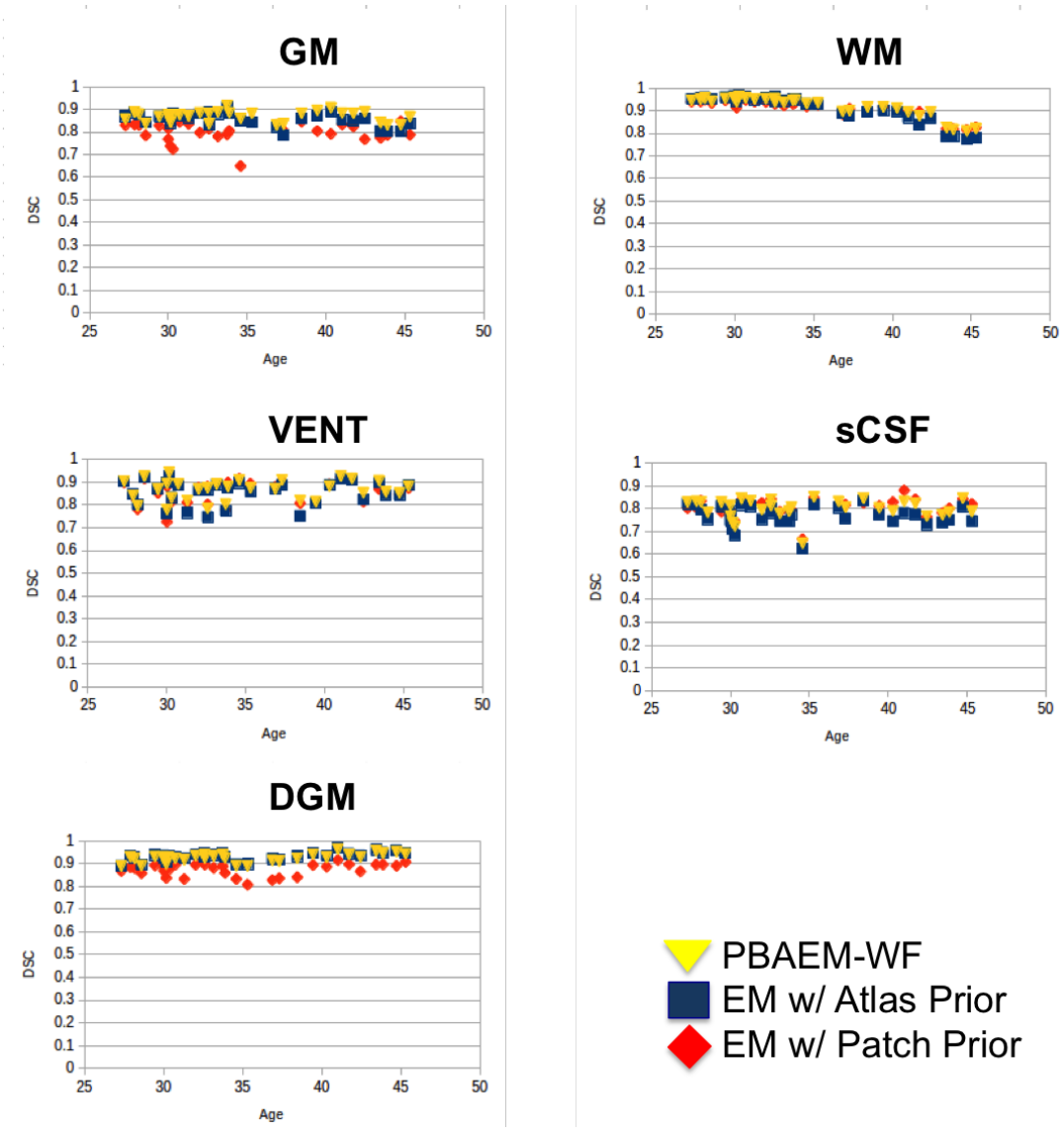

Figure 3.10: DSCs of five tissue classes of 31 individual scans plotted with age.

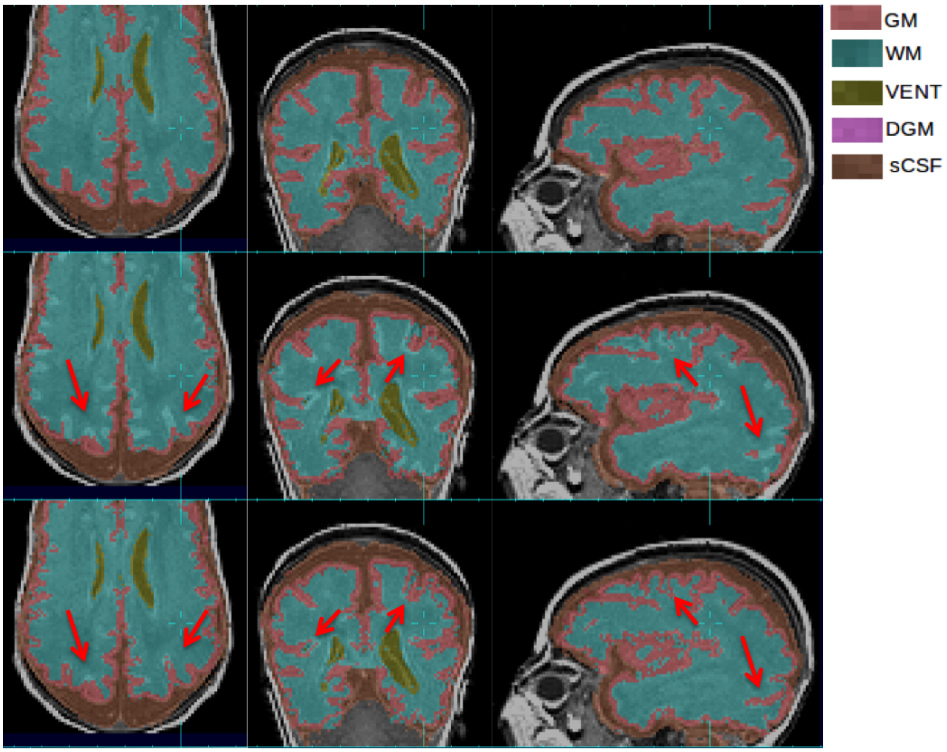

Figure 3.11: Improvements of GM and WM segmentation in a subject where the cortex is significantly more folded than the age-specific average template. Top row: manual segmentation; Middle row: atlas-based automatic segmentation (DSC: GM 0.7580, WM 0.8882); Bottom row: PBAEM-WF (DSC: GM 0.8477 with improvement of 0.0897, WM 0.9151 with improvement of 0.0269). Red arrows: GM-WM boundaries where PBAEM-WF was proved to generate more accurate labeling than atlas-based approach.

lists the number of scans that obtained higher DSC using PBAEM-WF prior compared to using either atlas- or patch-based prior alone. We can see that the segmentation performance improved in a majority of subjects using PBAEM-WF prior compared to only using either atlas- or patch-based prior.

Furthermore, we computed the VLA maps of the atlas-, patch-based and PBAEM-WF prior using leave-one-out cross-validation and show comparison in Figure 3.13. We can see that at GM-WM boundaries and VENT boundaries, the PBAEM-WF prior was more accurate (shown as brighter in the VLA map) compared to using atlas-based prior alone,

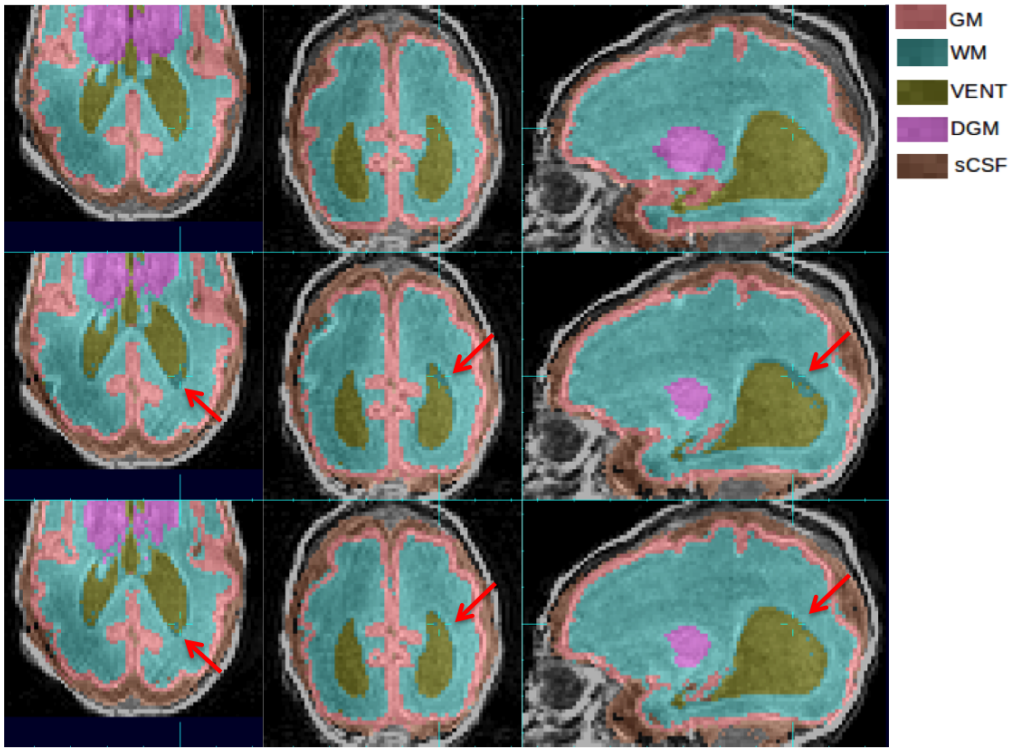

Figure 3.12: Improvements of VENT segmentation in a subject where VENT is significantly larger than the age-specific average template. Top row: manual segmentation; Middle row: atlas-based automatic segmentation (DSC: VENT: 0.9191); Bottom row: PBAEM-WF (DSC: VENT 0.9364 with improvement of 0.0173).Red arrows: VENT boundaries where PBAEM-WF was proved to generate more accurate labeling than atlas-based approach.

Table 3.1: Comparison of average DSCs of 31 subjects between using EM algorithm with atlas-based prior and proposed PBAEM-WF approach. Tissues with improvements in DSC are highlighted.* indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test.

Tissue Class	EM w/ Atlas-based Prior	PBAEM-WF	Difference in Average	Difference in StdDev
GM	**0.8558 ± 0.0298**	**0.8686 ± 0.0236**	**+0.0128***	**-0.0062**
WM	**0.9105 ± 0.0611**	**0.9184 ± 0.0452**	**+0.0079***	**-0.0159**
VENT	**0.8555 ± 0.0525**	**0.8673 ± 0.0438**	**+0.0118***	**-0.0087**
sCSF	**0.7692 ± 0.0451**	**0.8014 ± 0.0428**	**+0.0322***	**-0.0023**
DGM	0.9318 ± 0.0191	0.9260 ± 0.0192	-0.0058*	+0.0001

Table 3.2: Comparison of average DSCs of 31 subjects between EM algorithm with patch-based prior and proposed PBAEM-WF approach. Tissues with improvements in DSC are highlighted.* indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test.

Tissue Class	EM w/ Patch-based Prior	PBAEM-WF	Difference in Average	Difference in StdDev
GM	**0.8014 ± 0.0428**	**0.8686 ± 0.0236**	**+0.0672***	**-0.0192**
WM	**0.9109 ± 0.0405**	**0.9184 ± 0.0452**	**+0.0075***	**+0.0047**
VENT	**0.8591 ± 0.0491**	**0.8673 ± 0.0438**	**+0.0082***	**-0.0053**
sCSF	**0.8029 ± 0.0416**	**0.8014 ± 0.0428**	**-0.0015**	**+0.0012**
DGM	**0.8742 ± 0.0280**	**0.9260 ± 0.0192**	**+0.0518***	**-0.0088**

Table 3.3: Number of scans out of 29 with improved DSC (difference more than 0.0005) using PBAEM approach compared to using atlas- or patch-based prior .

Comparison	GM	WM	VENT	sCSF
PBAEM compared to Atlas-based	24/29	20/29	23/29	20/29
PBAEM compared to Patch-based	26/29	29/29	24/29	28/29

while at DGM-WM boundaries, the PBAEM-WF prior proved to be more accurate compared to using patch-based prior alone.

3.4.5 Comparison with NeoBrainS12 Challenge Result on Scans with Similar Ages

The NeoBrainS12 dataset is one of the most established public dataset of neonatal brain MR images. Automatic segmentation performance has been reported using various methods in the NeoBrainS12 challenge. In this section, we aim to compare our results to the published ones from the challenge to show the superiority of the proposed PBAEM-WF.

However, we could not test directly on the NeoBrainS12 dat set due to two reasons. On one hand, the dataset only contains brain MR scans at two time points: 30 and 40 GW. Thus, the spatio-temporal atlas that is required in our proposed method could not be constructed from the provided dataset. On the other hand, we could not use the spatio-temporal atlas constructed from our dataset to automatically segment the NeoBrainS12 images because their manual labeling followed different protocols. Thus, our atlas cannot represent the tissure features of the NeoBrainS12 dataset. As an alternative, we compute the average DSCs of scans from our dataset around 30 and 40 GW respectively, and compared them with the DSCs published in the challenge as in Table 3.4 and 3.5. Note that, for our DGM, we compared with the basal ganglia and thalami (BGT) in the challenge as an approximation. We can see that at age of 30 GW, PBAEM-WF demonstrates superiority over all methods in GM, WM and DGM, and over most methods in VENT and sCSF. At age of 40 GW, PBAEM-WF

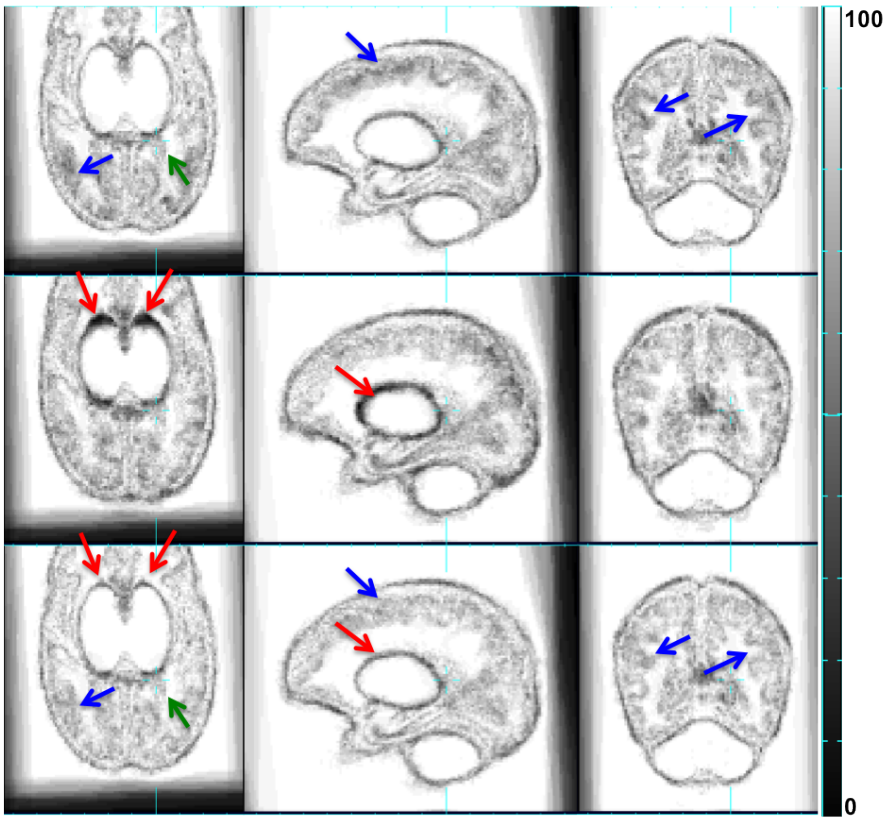

Figure 3.13: Comparison between VLA maps of atlas-based prior (top row), patch-based prior (middle row) and PBAEM-WF prior (bottom row). Red arrows: PBAEM-WF prior is more accurate than patch-based one at DGM boundaries; Blue arrows: PBAEM prior is more accurate than atlas-based one at GM-WM boundaries; Green arrow: PBAEM prior is more accurate than atlas-based one at VENT boundaries.

Table 3.4: Comparison between the 1st place on the 30 GW coronal MRI in NeoBrainS12 Challenge and average performance of PBAEM-WF on 5 scans around 30 GWs ($30.10 \pm 0.42 GWs$). For a fair comparison, we rounded to the hundredth place.

		GM	WM	VENT	sCSF	DGM
DSC	1st Place	0.71	0.90	0.88	0.83	0.84
	PBAEM-WF	0.86	0.95	0.86	0.77	0.91
	Rank	> 1st place	> 1st place	3rd place	2nd place	> 1st place

Table 3.5: Comparison between the 1st place on the 40 GW coronal MRI in NeoBrainS12 Challenge and average performance of PBAEM-WF on 2 scans around 40 GWs ($40.21 \pm 1.12 GWs$).

		GM	WM	VENT	sCSF	DGM
DSC	1st Place	0.77	0.84	0.84	0.77	0.88
	PBAEM-WF	0.86	0.88	0.86	0.79	0.95
	Rank	> 1st place	> 1st place	> 1st place	> 1st place	> 1st place

achieved higher segmentation accuracy over all published methods in all 5 tissue classes.

3.4.6 Application to Severe Abnormalities

To further test the robustness of our proposed algorithm when applied to unseen abnormal cases that are significantly different from those in the atlas, we constructed a spatio-temporal atlas using all 30 atlas subjects and applied our method on a non-atlas subject with severe ventriculomegaly and Grade 2 intraventricular hemorrhage (IVH). Figure 3.14 illustrates the comparison of the performances between PBAEM-WF and the conventional atlas-based approach. Because this scan's anatomy drastically varied from the spatio-temporal atlas constructed from healthy subjects, the atlas-based approach failed to generate a valid auto-

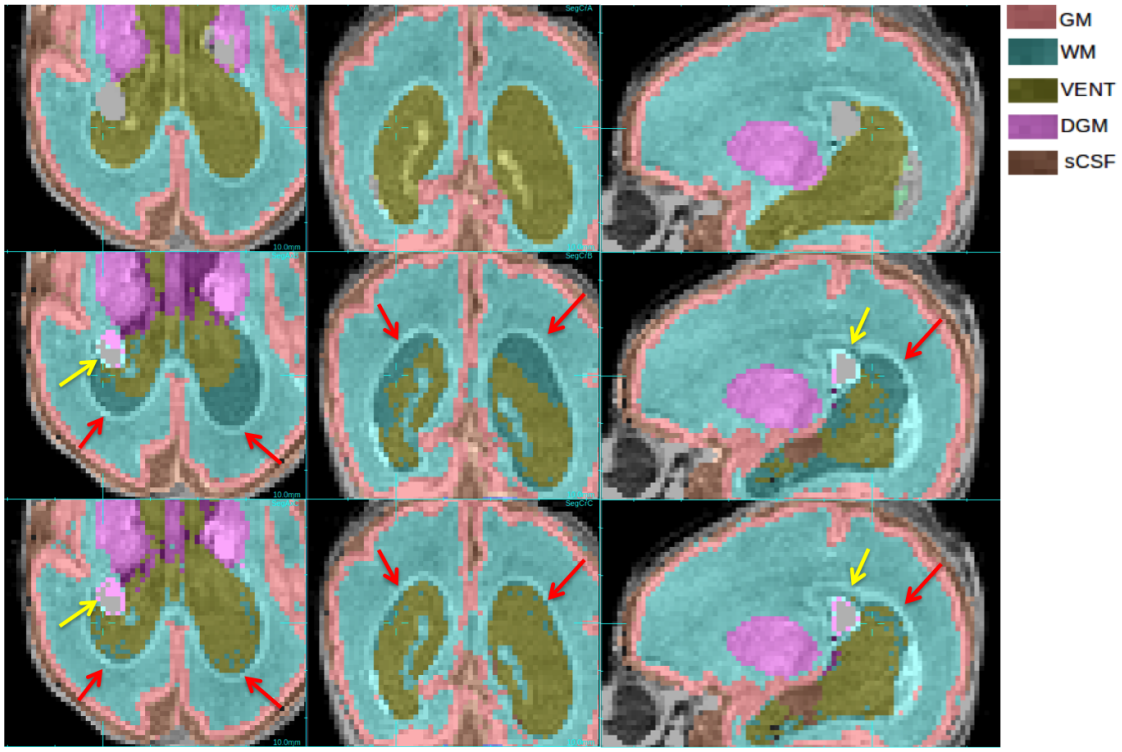

Figure 3.14: Labeling of abnormal anatomy not represented in the atlas. Comparison of manual labeling (top row), conventional atlas-based (middle row) and PBAEM-WF (bottom row) automated labeling in a case of severe ventriculomegaly and Grade 2 IVH. Red arrows: PBAEM-WF approach produced a more valid and accurate labeling of the enlarged VENT; Yellow arrows: IVH was partially labeled as BG.

mated tissue labeling, as seen in the middle row of Figure 3.14. In particular, the abnormally enlarged ventricular regions were not captured. The bottom row of Figure 3.14 shows the ability of our proposed PBAEM-WF approach to adapt to the unseen anatomies which greatly differed from the atlas. The abnormal ventricular regions were more accurately contoured and the blood clot was partially labeled as non-brain background (BG) based on its abnormal intensity at its location.

3.5 Discussion and Conclusion

In this chapter, I present a hybrid approach that augments the conventional model-based EM segmentation technique for labeling rapidly developing brain anatomy. The proposed method combines the best features in both atlas- and patch-based approaches in order to obtain more accurate automated labeling of brain tissues. We introduced the patch-based tissue probability estimates as a representation of any new subject's own anatomy within the atlas-based EM segmentation framework. In particular, we adapted the patch-based segmentation approach to use the parametric average estimate synthesized from the atlas as the patch-based dictionary images, which allows us to use atlases constructed with a relatively sparse distribution of manually segmented scans over the age range. A series of experiments with challenging examples of MRI from a range of clinical cases of premature birth showed a superior performance of the proposed PBAEM framework over either the conventional atlas-based segmentation approach or the local patch-based methods. Successful application of PBAEM to the abnormal cases also further highlighted its ability to adapt to the new unseen anatomies that are not captured in the atlas.

Chapter 4

PBAEM USING VARIABILITY-CONSTRAINED SEARCH (PBAEM-VCS) FOR HEALTHY PREMATURE NEONATAL BRAIN MRI TISSUE SEGMENTATION

4.1 Introduction

In the previous chapter, the atlas- and patch-based techniques were integrated in a way of creating a new combined tissue probability computed as the weighted average of atlas- and patch-based tissue probabilities. Using such method, segmentation accuracy of cortical regions was significantly improved compared to the conventional atlas-based EM approach. However, in PBAEM-WF, the computation was excessive. To build the spatio-temporal VLA models, EM segmentation needs to be run with both atlas- and patch-based tissue probabilities. In addition, a spatio-temporal PC model is required to account for the correlation between two sources of tissue probability estimates. As far as the patch search is concerned, for each subject, the search range is globally defined. Therefore, in challenging regions (e.g. cortical regions), the search range may not be big enough to find matching anatomy, while in structures with less variation (e.g. DGM and inside WM), most computation of patch similarity within the search range is wasted.

The spatio-temporal atlas can only account for the age-dependent variability across ages. However, scans of subjects at the same age are anatomically different as well, especially the depth and location of cortical foldings. Nonetheless, this level of variability at any specific age was not captured in the atlas. In PBAEM, the purpose of introducing in the patch-based local search is to correct for the segmentation error caused by this level of variability.

In this chapter, I present a novel and simplified approach to integrate the two techniques by using the atlas to locally define the patch-based search range. In PBAEM-VCS, the

anatomical variability at each certain age for all ages is modeled using a spatio-temporal and used as a guideline for determining how far should the patch-based local search for each location in the brain. In this way, the patch-based search could optimally compensate for the population variability at any certain age. In this approach, besides the spatio-temporal tissue atlas, only one additional spatio-temporal model of anatomical variability map is required which can be directly learnt from the atlas and anatomy without running EM segmentation pipeline. Besides, since the search range of locally leanrt, patches can be better matched in highly variable regions while computation in regions with less variation can be greatly reduced. Despite of straightforwardness and less computation, PBAEM-VCS is shown to improve the segmentation accuracy even more that PBAEM-WF.

4.2 Methodology

4.2.1 Overview

In this section, I first define the anatomical variability and explain the rationale behind modeling it. Then details of modeling and synthesizing age-specific variability map are given. Last, I explain how to use the locally learnt variability map to constrain the patch search range

4.2.2 Define Variability at Each Location in Brain

Here the purpose is to examine the anatomical variability that is not captured by the atlas. Recall that in the spatio-temporal atlas, we have a displacement field modeled across ages which measures the difference of voxel locations between each subject space and the average reference (i.e. atlas) space. The displacement field contains three values at each voxel which are the displacement in x, y and z direction respectively. The true displacement field of each subject is computed from group warp when constructing the atlas[53], and what atlas captures is reflected by the synthesized age-specific displacement field. Figure 4.1 shows the comparison between the true and estimate displacement field for one subject. Essentially,

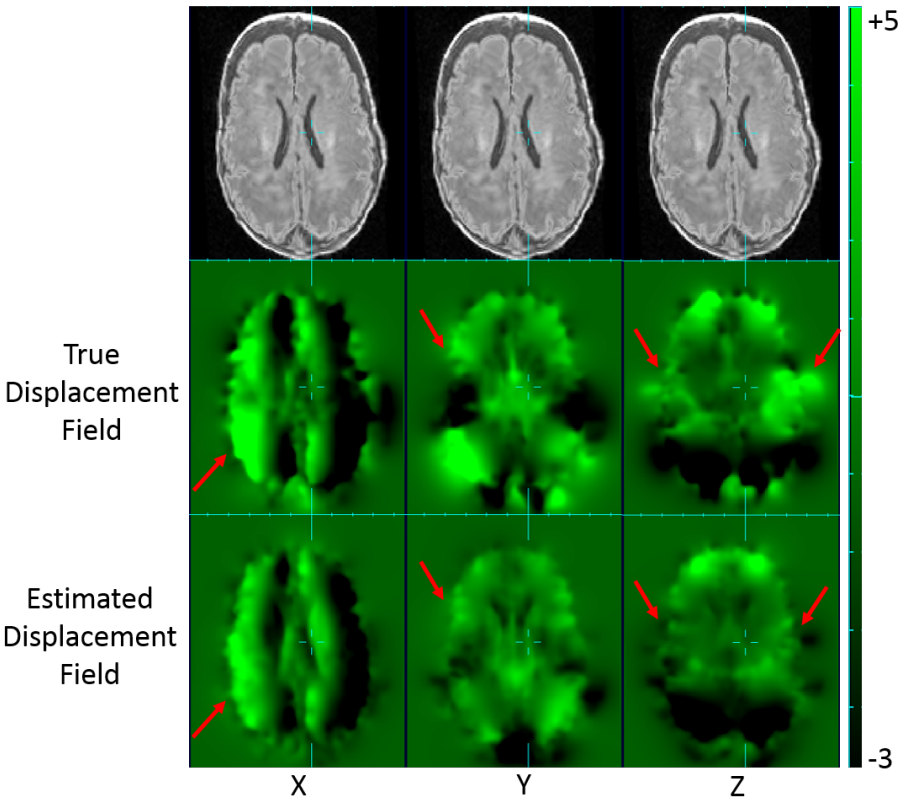

Figure 4.1: An example of differences between the true and estimated displacement fields of one subject (value in number of voxels). Red arrows point to regions with significant difference in cortical regions.

the variability that we want to measure is, for each subject, the difference between the true displacement field and the synthesized age-specific displacement field, in other words, the residual of the displacement field model. For convenience, I will refer to it as the residual displacement field (RDF).

4.2.3 Model RDFs Using Spatio-Temporal Models and Synthesize Age-Specific Variability Map

In order to obtain an estimate of variability at any age, a spatio-temporal model is constructed from RDFs using the training data. I followed the same approach of temporal

modeling the shape deformation as in [53]. Then for any given age of the scan to be segmented, an age-specific variability map can be synthesized. Three values at each voxel location indicate, on average, how far its matching point could be.

4.2.4 Using Variability Map to Define Patch-based Search Range

This locally learnt age-specific variability map can then be used to define the patch-based local search range. Essentially, for each voxel, the variability in x, y and z direction gives us a 3D ellipsoid centered at the voxel under investigation. This ellipsoid defines the search range for its center voxel. Therefore, a voxel at coordinate (X', Y', Z') is in the search range of voxel (X, Y, Z) if the following inequality is satisfied:

$$\frac{(X - X')^2}{Vx^2} + \frac{(Y - Y')^2}{Vy^2} + \frac{(Z - Z')^2}{Vz^2} \leq 1 \tag{4.1}$$

where Vx, Vy, Vz is the variability at voxel (X, Y, Z) in x, y and z direction respectively. In this way, in regions with less variation, such as DGM and inside WM, the variability-defined search range is restricted. In extremely cases, when variability is minimal, no search will be conducted and the patch-based tissue probability will be the same as the atlas-based tissue probability at that voxel, which is desired for regions with extremely low variation.

4.2.5 Patch-based Augmentation of EM using Variability Constrained Search (PBAEM-VCS) Framework Framework

Similarly to PBAEM-WF, in PBAEM-VCS, we found that best automated tissue labeling performance was achieved using patch-based tissue probability as the initialization prior and the combined tissue probability as the running prior in the EM algorithm. Therefore, in the initialization step, Eq. (3.2) should be replaced with:

$$P_{init}(k \mid y_i) = P_p(k \mid y_i), k \in [1, K] \tag{4.2}$$

In the iterations, Eq. (3.7) should be replaced with:

$$P_{run}^{(t)}(k) = P_w(k \mid y_i), k \in [1, K] \tag{4.3}$$

4.3 Experiments and Validation

To validate the proposed PBAEM-VCS, I utilized the same premature neonatal brain MR dataset as in the previous chapter. The spatio-temporal was chosen to be linear to avoid any potential over-fitting. To experimentally determine the optimal threshold of search range, $1.5 \sim 3.0$ times the variability were tested and results will be shown in the following section. Leave-one-out cross-validation was used in the entire pipeline for both spatio-temporal atlas and spatio-temporal variability map. That is, for each subject to be segmented, the age-specific displacement field was generated using a leave-one-out atlas, and the variability models were also constructed using difference maps of the other 31 scans. To quantify the performance of the automatic segmentation, the Dice Similarity Coefficient (DSC), Hausdorff distance (HD) [63] and mean distance (MD) of the tissue volume between manual and automatic tissue labellings were used. Detailed distance measurements of GM and WM divided by 8 lobes (Frontal Left/Right, Temporal Left/Right, Parietal Left/Right and Occipital Left/Right) were also computed. To challenge the validity of proposed methods without using computationally expensive leave-on-out cross-validation for PC and VLA maps, experiments on non-atlas scans which are not used in the atlas training dataset consisting of 32 scans) were conducted. We present the results in the following section.

4.4 Results

4.4.1 Patch Search Range Threshold Selection

Experiments were conducted using different search ranges of from 1.5 to 3.0 times of variability at each voxel. The average and standard deviation of DSC of 31 subjects are compared, as shown in Figure 4.2. We can see that the segmentation performance improves with larger

search range until an optimal threshold, and then decreases. This is because when smaller than the optimal, the search range is not large enough to cover most potential matching voxels. However, with a search range too big, the tissue probabilities of the best matches will be diluted by the non-matches. Therefore, the optimal search range exits. To balance the performances of all tissue classes, the threshold of twice the variability was set.

4.4.2 Age-Specific Variability Map

To visualize the age effect on variability in neonatal brain MRIs, synthesized age-specific variability maps at four different gestational ages are shown and compared in Figure 4.3. We can clearly see that the variability at cortical regions dramatically grows with age, especially at major cortical gyrifications at central sulcus (red arrow) and primary visual cortex (yellow arrow).

4.4.3 Patch-based Tissue Probability with Locally Constrained Search

Using the threshold of search range chosen above, patch-based tissue probability was computed for each voxel in the scan being studied. Figure 4.4 shows the comparison of GM tissue probability between the atlas-based, patch-based using a global search range as in PBAEM-WF and using a locally defined search range as in PBAEM-VCS. Compared to atlas-based tissue probability, the VCS patch-based tissue probability demonstrates higher accuracy at cortical GM (pointed by yellow arrows). Compared to patch-based tissue probability using globally set search range, VCS patch-based tissue probability shows superior accuracy in hippocampus (red arrow) and sharper tissue probability boundaries in GM. Such more definite tissue probability could avoid ambiguity in EM segmentation to achieve a more accurate automated tissue labeling.

Figure 4.5 shows the improvement to accurately estimate the tissue probability in regions with less variation, DGM. Since the variability is low in DGM, the variability constraint search range is also small. Thus, in such regions, the patch-based tissue probability is dominated by the atlas-based tissue probability of its closest neighbor and will not be further

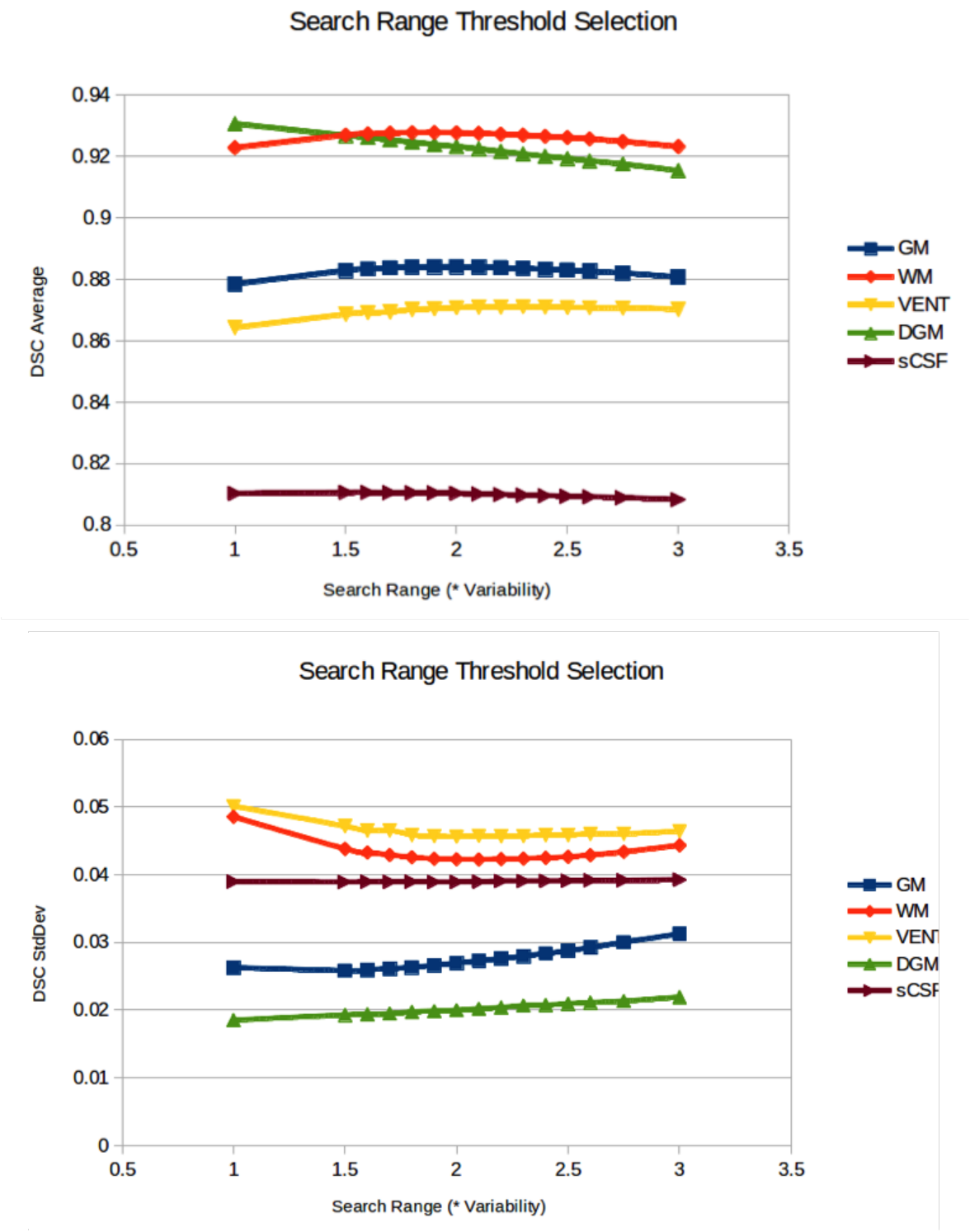

Figure 4.2: Average (left) and standard deviation (right) of DSC of 31 subjects using different search range threshold.

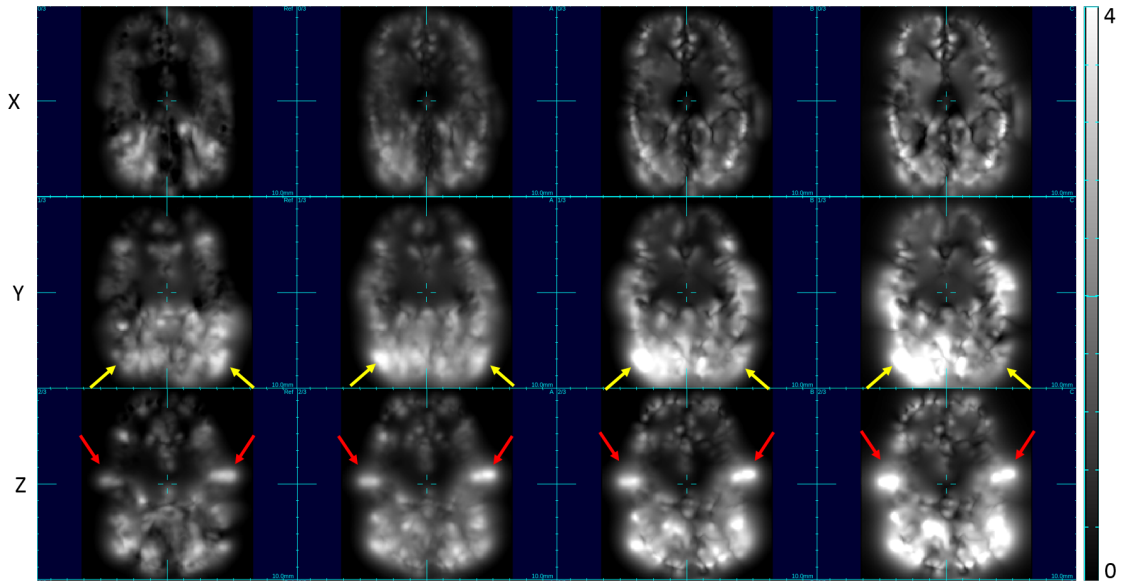

Figure 4.3: Age-specific variability maps in x(upper row), y(middle row) and z(lower row) direction at four different gestional ages (value in number of voxels). 1st column: 28.57 GW; 2nd column: 35.29 GW; 3rd column: 40.29 GW; 4th column: 45.29 GW. Red arrows: central sulcus; Yellow arrrows: primary visual cortex.

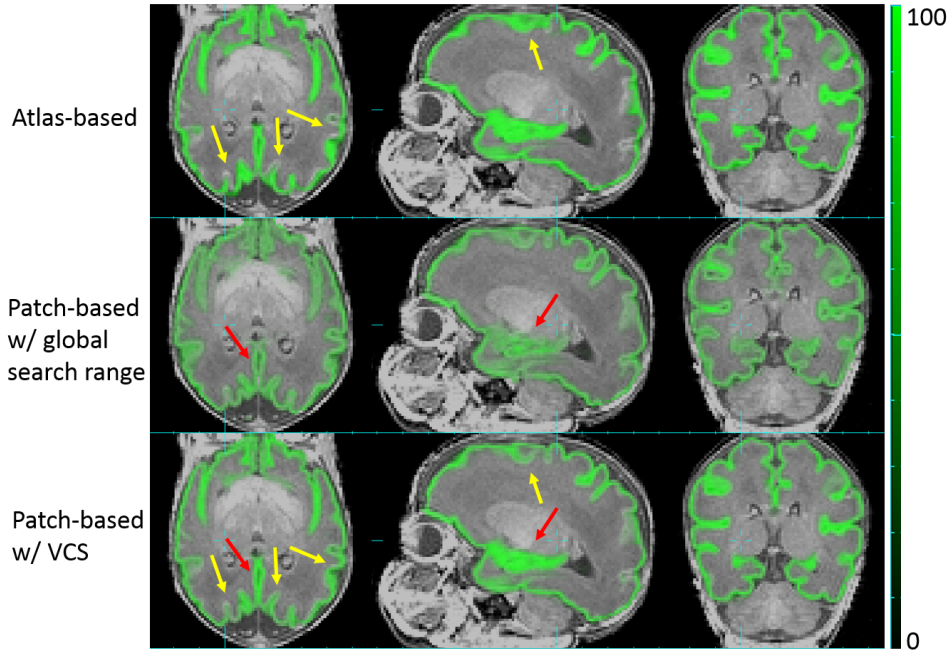

Figure 4.4: Comparison of GM tissue probabilities between atlas-based, patch-based with globally and locally set search range. Red arrows: VCS patch-based TP better than patch-based with global search range; Yellow arrows: VCS patch-based TP better than atlas-based TP.

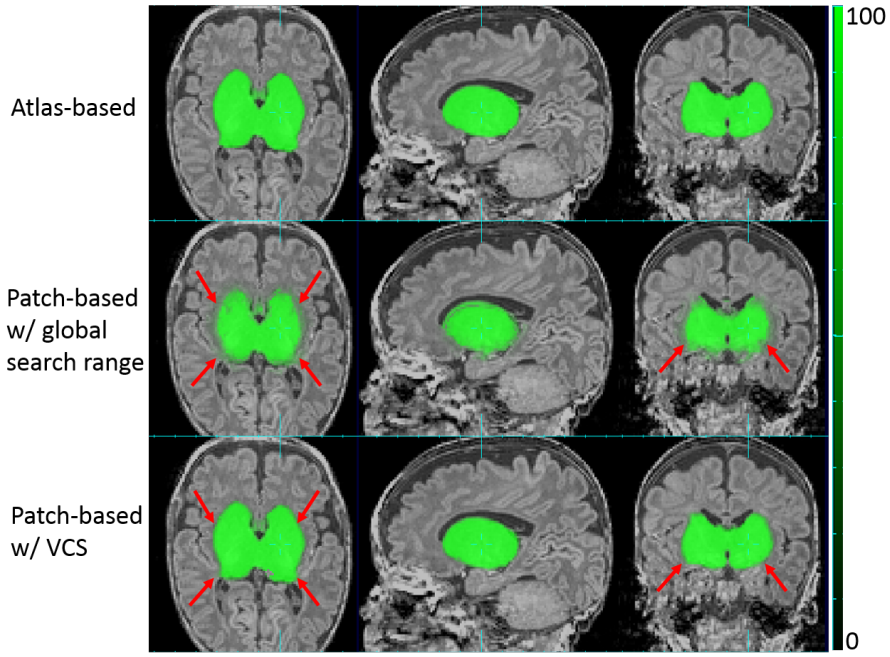

Figure 4.5: Comparison of DGM tissue probabilities between atlas-based, patch-based with globally and locally set search range. Red arrows: VCS patch-based TP better than patch-based with global search range.

diluted. This is one of the major advantages of locally constraint patch-based search so that no additional fusion with the atlas-based tissue probability is needed.

4.4.4 Comparison with PBAEM-WF

The final automated tissue segmentation were compared between PBAEM-WF and PBAEM-VCS. Table 4.1 compares the average DSC of 31 subjects using two methods. We can see that the segmentation accuracy is significantly ($p < 0.05$) improved using PBAEM-VCS, especially in cortical GM. Even though, in GM and VENT, the standard deviations are higher than PBAEM-WF, the differences are minimal. Table 4.2 shows the comparison of HDs between two PBAEM approaches. Improved HDs using PBAEM-VCS are observed in all tissue classes except for VENT. In fact, PBAEM-VCS achieves the best performances of

Table 4.1: Comparison of average DSCs of 31 subjects between using PBAEM-WF and PBAEM-VCS. Tissues with improvements in DSC are highlighted.* indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test..

Tissue Class	PBAEM-WF	PBAEM-VCS	Difference in Average	Difference in StdDev
GM	**0.8686 ± 0.0236**	**0.8840 ± 0.0269**	**+0.0154***	**+0.0033**
WM	**0.9184 ± 0.0452**	**0.9276 ± 0.0423**	**+0.0092***	**-0.0029**
VENT	**0.8673 ± 0.0438**	**0.8707 ± 0.0456**	**+0.0034**	**+0.0018**
sCSF	**0.8014 ± 0.0428**	**0.8103 ± 0.0389**	**+0.0089***	**-0.0039**
DGM	0.9260 ± 0.0192	0.9232 ± 0.0200	-0.0028	+0.0008

HD in GM, WM, DGM. The reason why PBAEM-VCS fails to perform as well is that the search range may be not large enough in cases of slightly enlarged ventricles.

Figure 4.6 gives an example of improvement in automatic labeling of GM when using PBAEM-VCS compared to PBAEM-WF. At corpus callosum (pointed by blue arrows), due to WM myelination, its intensity is similar to cortical GM in its neighborhood. If the search range is too big, it will be more likely to be labeled as cortical GM instead of WM as in PBAEM-WF. However, using PBAEM-VCS, we are able to learn that the variability in corpus callosum is low, and thus, patch-based search will be conducted more locally to avoid mislabeling.

Figure 4.7 provides a visualization of improvement of PBAEM-VCS compared to atlas-based segmentation and PBAEM-WF with different ages. Improvements are seen especially in older age range, where segmentation errors were relatively larger compared to younger age range using both atlas-based approach and PBAEM-WF. PBAEM-VCS helps to reduce inconsistency of segmentation accuracy across ages significantly. However, in WM, we still see decreasing DSC values with older ages. There are two main reasons. Firstly, since WM is adjacent to all other tissue classes, it collects all the segmentation errors in other tissue

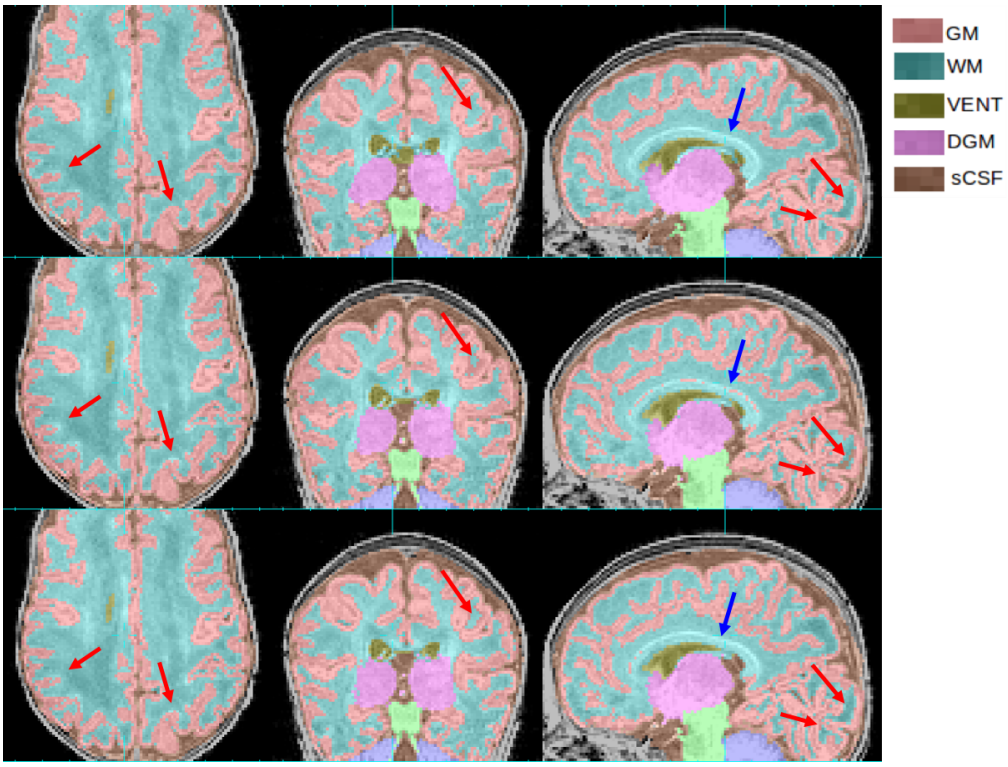

Figure 4.6: Improvements of GM and WM segmentation in a subject where the cortex is significantly more folded than the age-specific average template. Top row: manual segmentation; Middle row: PBAEM-WF (DSC: GM 0.8946, WM 0.9153); Bottom row: PBAEM-WF (DSC: GM 0.9171 with improvement of 0.0225, WM 0.9330 with improvement of 0.0177). Red arrows: Cortical GM-WM boundaries where PBAEM-VCS was proved to generate more accurate labeling than PBAEM-WF approach; Blue arrows: corpus callosum

Table 4.2: Comparison of average HDs of 31 subjects between using PBAEM-WF and proposed PBAEM-VCS approach. Improved items are bold. * indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test.

Tissue Class	PBAEM-WF	PBAEM-VCS	Difference in Average	Difference in StdDev
GM	6.7364 ± 1.0156	6.3480 ± 1.5212	**-0.3884**	+0.5056
WM	6.9771 ± 4.0929	6.4155 ± 4.3259	**-0.5616***	+0.2330
VENT	18.9204 ± 6.0749	20.4287 ± 6.3525	+1.5083	+0.2776
sCSF	10.1193 ± 1.6036	9.8709 ± 1.6759	**-0.2484**	+0.0723
DGM	4.7416 ± 1.3220	4.0894 ± 0.9585	**-0.6522***	**-0.3635**

classes. With older age, the scan is more difficult to segment, and thus more error in most tissue classes, WM would have a even lower DSC at older ages. Secondly, volume-wise, WM is the largest tissue class which we are labeling. Therefore, there are more random error on the boundary for a more folded brain at older age. This would cause a more dramatic decrease of DSC with age in WM than other tissue classes.

4.4.5 Application to Severe Abnormalities

I further tested PBAEM-VCS on the scan of a subject suffering from IVH. Figure 4.8 shows an example when the blood clot was mislabeled as BS by PBAEM-WF while was correctly labeled as BG by PBAEM-VCS. This is because, in PBAEM-VCS, I was able to learn a relatively smaller search range at the hemorrhage site (inside WM) than the globally defined search range in PBAEM-WF. Therefore, the patch-based search did not wrongfully search in the BS region to find a match, and instead, couldn't find a match in its close neighborhood. Thus, it gave a more accurate patch-based tissue probability estimate as BS for the blood clot. However, this also caused segmentation errors in the greatly enlarged ventricles due to too limited search range (yellow arrows). This issue will be further addressed in the next

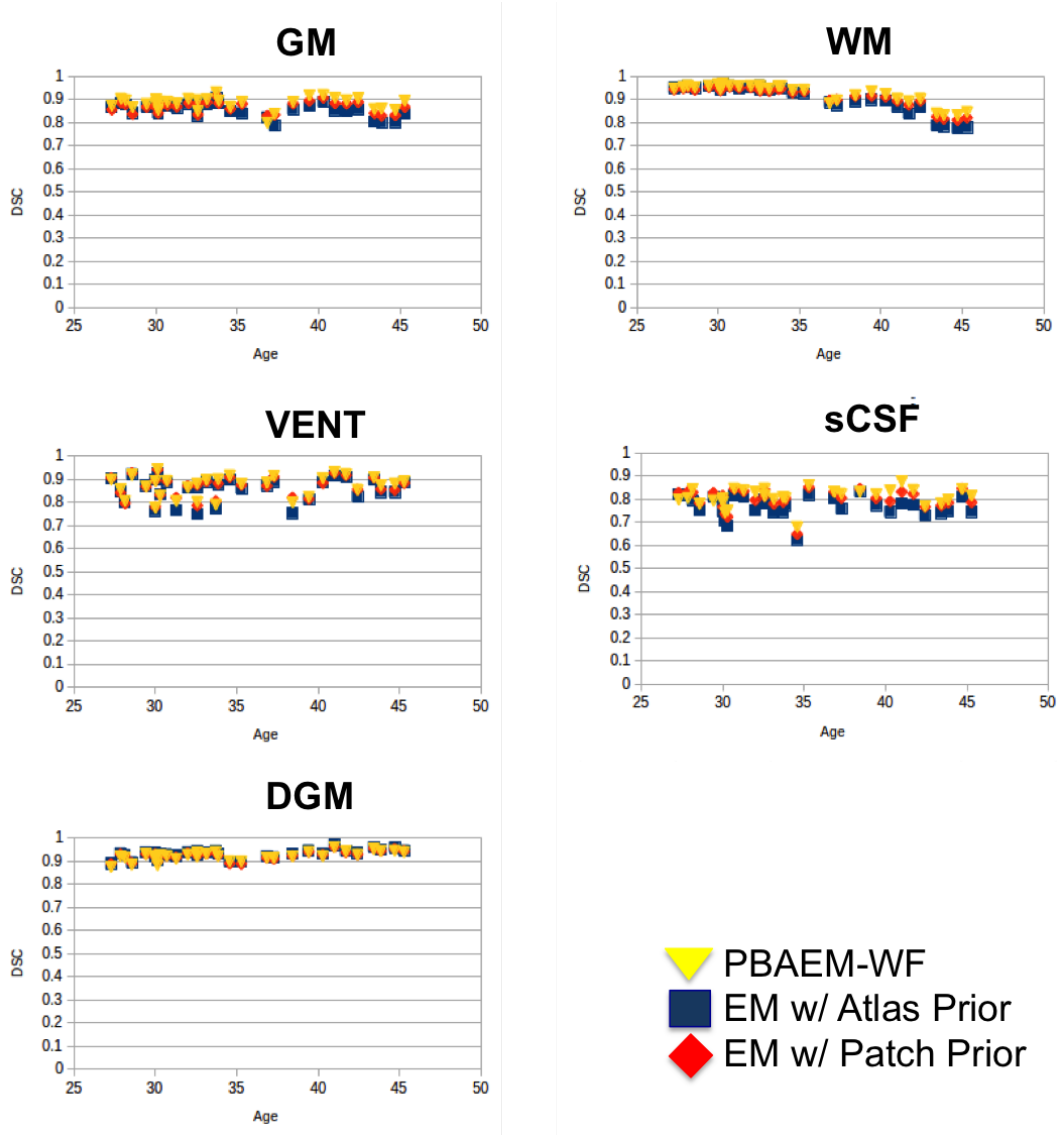

Figure 4.7: DSCs of five tissue classes of 31 individual scans plotted with age. Comparison is shown between segmentation performances of conventional atlas-based EM approach, PBAEM-WF and PBAEM-VCS.

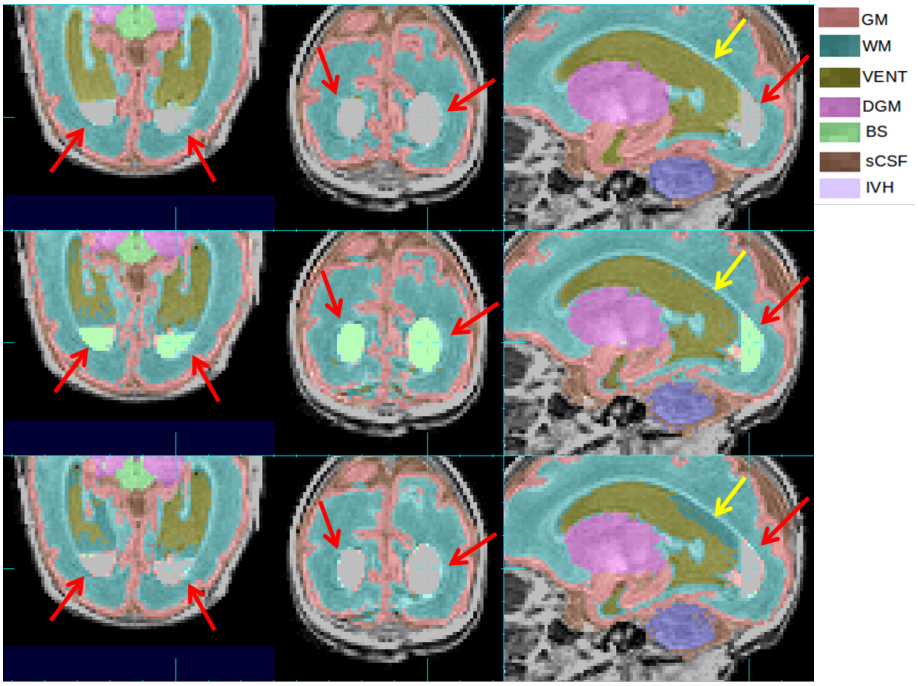

Figure 4.8: Comparison of PBAEM-WF and PBAEM-VCS on an abnormal scan with IVH. Red arrow: IVH mislabeled as BS by using PBAEM-WF while correctly labeled as BG by using PBAEM-VCS; Yellow arrows: mislabeled VENT as WM by PBAEM-VCS while correctly labeled by PBAEM-WF.

chapter.

4.4.6 Distance Evaluation

Based on comparison of DSC in Table 3.1 and Table 3.2, we see that PBAEM-VCS shows the most superior segmentation performance. Here we further computed and compared the average HD and MD (Table 4.3) between PBAEM-VCS and conventional atlas-based EM technique to show its superiority. We can see that in most cases, PBAEM-VCS provided better performance than the conventional age-specific atlas-based approach. The majority of the large errors we observe occurred outside brain tissue, e.g. around skull due to partial volume effect. This could be solved with a more accurate skull stripping or masking in the future.

Table 4.3: Comparison of average HDs and MDs of 31 subjects between PBAEM-VCS and atlas-based EM technique. Lowest HDs and MDs are bold. * indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test between the two methods.

Tissue Class	EM w/ Atlas-based Prior		PBAEM-VCS	
	MD	HD	MD	HD
GM	0.17 ± 0.04	**5.58 ± 1.11**	**0.13* ± 0.03**	5.79 ± 1.19
WM	0.12 ± 0.10	5.67 ± 2.15	**0.09* ± 0.06**	**5.64 ± 1.89**
VENT	0.35 ± 0.30	23.61 ± 6.24	**0.23* ± 0.15**	**19.78* ± 6.10**
sCSF	0.31 ± 0.09	**9.45** ± 1.90	**0.26* ± 0.06**	9.60 ± **1.88**
DGM	**0.08*** ± 0.03	5.13 ± 9.62	0.09 ± 0.03	**3.94 ± 0.89**

We also computed distance measurements of cortical GM by individual lobes (Table 4.4). The lobe definition is shown in Fig. 4.9. We see that both methods achieved consistent results across different lobes and PBAEM-VCS showed significant superiority in all lobes of the cortex. We also noted that segmentation accuracy in the occipital lobes was improved by using PBAEM-VCS over conventional age-specific atlas-based method.

4.5 Discussion and Conclusion

In this chapter, I describe an improved approach on PBAEM-WF to locally learn the search range to achieve higher segmentation accuracy while reducing unnecessary computation. Especially, significant improvement in DSC is seen in cortical regions, where are most challenging.

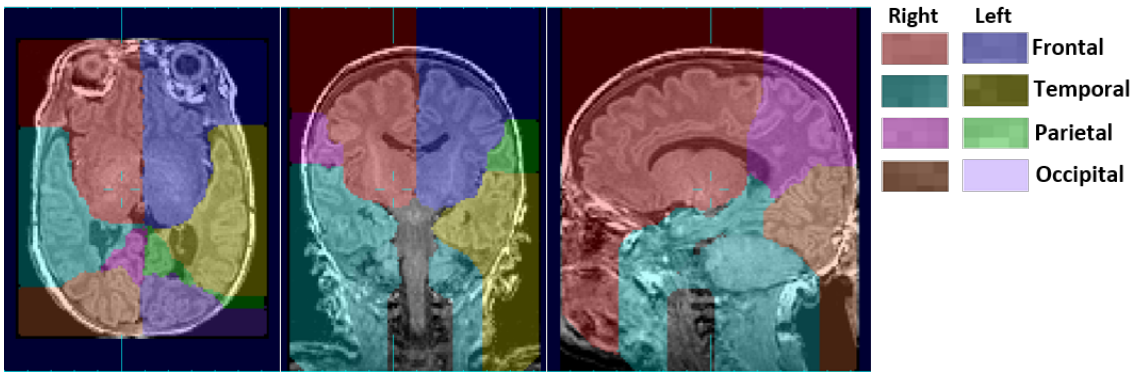

Figure 4.9: Labeling of 8 cerebral lobes.

Table 4.4: Comparison of average HDs and MDs of 8 lobes of cortical GM from 31 subjects between PBAEM-VCS and atlas-based EM technique. Lowest HDs and MDs are bold. * indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test between the two methods.

Lobe	EM w/ Atlas-based Prior		PBAEM-VCS	
	MD	HD	MD	HD
Frontal R	0.15 ± 0.04	**4.25 ± 1.34**	**0.12* ± 0.03**	4.38 ± 1.43
Frontal L	0.15 ± 0.04	4.43 ± **1.41**	**0.12* ± 0.03**	4.43 ± 1.47
Temporal R	0.17 ± 0.04	4.16 ± **0.91**	**0.14*** ± 0.04	**4.05** ± 0.99
Temporal L	0.17 ± 0.05	4.00 ± **0.90**	**0.14* ± 0.04**	**3.91** ± 1.04
Parietal R	0.17 ± 0.07	**4.40* ± 1.10**	**0.13* ± 0.06**	4.90 ± 1.20
Parietal L	0.15 ± 0.05	**3.95 ± 1.21**	**0.11* ± 0.03**	4.24 ± 1.28
Occipital R	0.22 ± 0.07	4.20 ± **1.15**	**0.16* ± 0.06**	**4.07** ± 1.31
Occipital L	0.19 ± 0.06	3.93 ± **1.19**	**0.13* ± 0.04**	**3.70** ± 1.35

Table 4.5: Comparison of average HDs and MDs of 8 lobes of WM from 31 subjects between PBAEM-VCS and atlas-based EM technique. Lowest HDs and MDs are bold. * indicates statistical significance with $p < 0.05$ from pair-wise two-tailed t-test between the two methods.

Lobe	EM w/ Atlas-based Prior		PBAEM-VCS	
	MD	HD	MD	HD
Frontal R	0.10 ± 0.07	4.28 ± 1.78	**0.08* ± 0.05**	**4.25 ± 1.58**
Frontal L	0.09 ± 0.06	4.10 ± 1.06	**0.08* ± 0.04**	**3.95 ± 0.94**
Temporal R	0.12 ± 0.09	4.15 ± 1.69	**0.10* ± 0.06**	**4.05 ± 1.42**
Temporal L	0.12 ± 0.09	4.01 ± 1.22	**0.10* ± 0.06**	**3.87 ± 1.13**
Parietal R	0.13 ± 0.12	**3.46 ± 1.32**	**0.09* ± 0.07**	3.52 ± 1.58
Parietal L	0.12 ± 0.12	3.68 ± **1.20**	**0.09* ± 0.07**	**3.53** ± 1.50
Occipital R	0.18 ± 0.17	**3.95* ± 1.34**	**0.13* ± 0.10**	4.33 ± 1.56
Occipital L	0.19 ± 0.22	**4.11** ± 2.28	**0.13* ± 0.13**	4.16 ± **1.90**

Chapter 5

MULTI-CENTER CROSS-SCANNER CROSS-PROTOCOL VALIDATION OF PBAEM-VCS USED FOR HEALTHY PREMATURE NEONATAL BRAIN MRI TISSUE SEGMENTATION

5.1 Introduction

Large-scale MRI clinical studies often require image analysis on datasets acquired by different sites using different scanners and imaging protocols[129]. Additionally, multi-center structural MRI studies can greatly increase statistical power and single-center studies[117]. Because of these, many multi-center studies on children[93] and adults[135, 67] have been carried out. However, when dealing with multi-center data, especially acquired using different scanners running different imaging protocols, across-center differences may cause difficulties in image processing and inconsistency in algorithm performances. Ideally, MR acquisition protocols should be optimized in advance by scanning calibration subjects before starting multi-center data acquisition[135, 27] to allow reliable combination of MRI data acquired at different sites in a multi-site MR study[118]. However, this is not always feasible. In some cases, due to scanner or facility restrictions, it may not be able to adopt the optimal protocol. Besides, in our case, it is both unethical and impossible to acquire brain MRIs of the same premature neonate within a short time to minimize the age effect. Last but not least, the decision to analyze multi-center data collectively is usually made after all the data has been collected using site-specific protocols[27].

Nonetheless, in such large-scale multi-site MRI-based brain studies, automatic delineation of brain tissues in the MR scan remains a crucial step and allows for further volumetric analysis due to the large amount of data that needs to be processed[61]. This can be even

more challenging than single-site MR segmentation due to potential large-scale cross-platform variations. Especially in studies that involve MRI-based quantitative measurements such as tissue volumes, cortical thickness etc., the susceptibility of the measurements to the MR scanner and protocol properties needs to be taken into consideration. However, previous studies have shown that the choice of a segmentation algorithm had the largest impact on variability of the results[27]. Therefore, the cross-vendor cross-site cross-protocol robustness of automated tissue labeling methods is greatly desired.

This is also extremely important for application of automated image analysis in clinical settings. Different hospitals and different imaging sessions may use different machines and MR acquisition protocols. It is impossible to build all the possible atlases for automated analysis of all types of MR images. Therefore, the adaptation from the atlas constructed using MR images acquired by using one protocol to the ones using other protocols is of great importance in clinical application.

5.1.1 Facility, Vendor, Imaging Protocol's Effect on Quantitative MRI Study

Factors such as pulse sequence, image artifacts, magnetic field strength and scanner hardware and software may affect compatibility in multi-center studies. Investigators should have knowledge of how different MR acquisition parameters could influence the image analysis procedures. Ideally, the image analysis pipeline can be optimized based on the performances on the whole multi-site datasets.

5.1.1.1 Facility & Vendor Differences

As discussed above, different sites usually use different MRI equipment due to facility constraints. Even though, in some cases, different manufacturers of MRI equipment adopt different name and implementation of similar pulse sequence (for example, SIEMENS uses the name FLASH while GE uses the name SPGR), vendor-specific differences in pulse sequence characteristics, coil sensitivity profiles and localization methods may also exist. In addition, each site may have different types of personnel performing the scans and overseeing

other aspects of the study. The professional skills and standards of "a good scan" may differ greatly due to training in different facilities. Thus, patient/subject preparation and setup etc. could also be different across sites. Therefore, the facility and vendor differences are the major contributor to the potential incompatibility between multi-center datasets. Numerous studies have studied the impact of different facility and scanner in multi-site studies using phantom[39], post-mortem human brain phantom[39, 35] and human adult study[126]. Filippi et al. suggested that the inter-scanner agreement of observer measured lesion load from multiple scans of the same subject acquired using different MRI scanners was lower than the intra-observer agreements of the same scan. Higher inter-scanner agreement was achieved using same magnetic field strength of 1.5T [39]. However, to the best of our knowledge, no such study on premature neonates has been reported.

Clark et al. reported, based on a calibration study, that the pulse sequence had the second largest impact on variability of results following the choice of a segmentation algorithm as the first [27]. Patwardhan et al. studied the effect of image orientation and acquisition parameters on volumetric measurements of pediatric MR Data using same scanner (1.5T GE Signa) with same pulse sequence (3D-SPGR)[93]. Differences were observed in regions of temporal and occipital GM and may resulted from PV and susceptibility-induced geometric distortions. It was recommended by their results that multi-center volumetric studies should control slice selection and image resolution and use thin slice profile to minimize PV averaging[93]. Last not but least, the emphasis of the scans may also contribute to the inconsistency in the resulting acquisition. For example, if doctors suspect lesions in a certain brain structure, then images emphasizing on that particular structure would be considered optimal.

5.1.1.2 Methods for Quantitative Analysis of Multi-center MRI Data

Generally speaking, there are three ways to analyze the quantitative measurements (e.g. tissue volumes, cortical thickness etc.) from multi-center MRI data[118]. First, each dataset can be analyzed individually by computing the quantitative measurements followed by statis-

tical meta-analysis[31]. Second, all datasets are analyzed according to one single established procedure at a central site to extract quantitative measurement and carry out statistical analysis. In these first two methods, the site-related variable could be further introduced to detect and calibrate systematic differences between the measurements[118]. Third and ideally, the processing pipeline can be optimized based on the comparability of the measurements across all contributing sites[118]. Of course, it would be most ideal if the optimization includes MR acquisition parameters for each site. Schnack et al. reported different approaches to test compatibility and optimize processing pipeline for different morphometric measurements. For volumtric measurements, adult volunteers were rescanned using different MR scanners and interclass correlation coefficient was computed to assess the cross-dataset variations[118]. For VBM and cortical thickness meaurements, calibration subjects were also used to learn between-site reliability maps which can be used for future analysis. However, to the best of our knowledge, no such study has been reported for population of premature neonates due to unavailability of calibration subjects.

5.2 Methodology

In this section, the population statistics and different MR acquisition parameters used in each dataset are listed. The spatio-temporal atlas used to analyse both dataset was constructed using SIEMENS data. 30 manually labeled MR images with age at scan ranging from 27.29 to 46.43 GW (34.63 $\pm$ 5.55 GW) from SIEMENS data were used to build the spatio-temporal model of tissue growth as well as VLA and PC model. Then PBAEM-WF was used to obtain the automated tissue labeling of 8 scans with similar ages from either dataset. Note that since the 8 testing scans from SIEMENS dataset are included in the atlas, leave-one-out cross-validation was adopted during automated segmentation to ensure fair comparison. The testing data population statistics is shown in Table 5.1. Different MRI machine vendor and imaging protocols are listed in Table 5.2. To minimize the cross-dataset differences, similar neonatal head coil, pulse sequence (FLASH by Siemens and SPGR by GE) and thin slice profile were adopted for acquisition. FLASH by Siemens is short for Fast Low Angle Shot

Table 5.1: Population statistics of the testing data from SIEMENS and GE dataset.

Usage	**Training**	**Testing**	
Dataset	SIEMENS data	SIEMENS data	GE data
Number of Scans	30	8*	8
Age Range	27.29 $\sim$ 46.43 GW	30.30 $\sim$ 36.86 GW	30.30 $\sim$ 36.70 GW
Age Mean $\pm$ Std	34.63 $\pm$ 5.55 GW	32.95 $\pm$ 2.37 GW	33.05 $\pm$ 2.45 GW

sequence.

The DSC were computed between the manual and automated labeling of all testing images and further compared between two dataset. In this study, due to the fact that, in GE dataset, only manual labelings of GM, WM, VENT and DGM are currently available, only these 4 tissue classes were studied.

5.3 Results

This cross-center cross-vendor cross-protocol validation of PBAEM automated segmentation technique has shown promising results. The segmentation performances of similar ages were comparable. Table 5.3 shows the average DSC achieved when segmenting the SIEMENS and GE datasets. The average values are comparable, however, we see that the standard deviations of GE data are larger than SIEMENS data. This is probably due to the fact that the atlas constructed using SIEMENS data can capture the features and variability in SIEMENS data better than the GE data. However, this hypothesis still needs to be further studied. Besides, the largest difference in segmentation performance exit in GM, which is in accordance to the results reported by Clark et al [27]. To visualize the segmentation performance change with age in two datasets, the DSCs of each scan in both dataset are plotted and compared in Figure 5.1.

Figure 5.2 gives an example of comparison between automated segmentation of a scan of

Table 5.2: MR acquisition parameters of the SIEMENS and GE dataset.

Parameters	**SIEMENS data**	**GE data**
Scanner	SIEMENS Avanto	GE Signa
Facility	UBC (BC Children's Hospital)	UCSF
Magnetic Field	1.5T	1.5T
Voxel Dimension	$1.04 \times 1.04 \times 1mm^3$	$1 \times 1 \times 1mm^3$
TE	9.2 ms	6 ms
TR	36 ms	35 ms
Sequence	FLASH	SPGR
Slice Selection	Coronal	Sagittal
Pixel Bandwidth	80	122.109
Flip Angle	30	35
% **Phase FOV**	$45 \sim 55$	75

Table 5.3: Average DSC of the automated tissue segmentations of SIEMENS and GE dataset.

Dataset	GM	WM	VENT	DGM
SIEMENS Data	0.8487 ± 0.0194	0.9300 ± 0.0249	0.8435 ± 0.0453	0.9047 ± 0.0241
GE Data	0.8056 ± 0.0363	0.8962 ± 0.0362	0.8439 ± 0.0649	0.8832 ± 0.0464

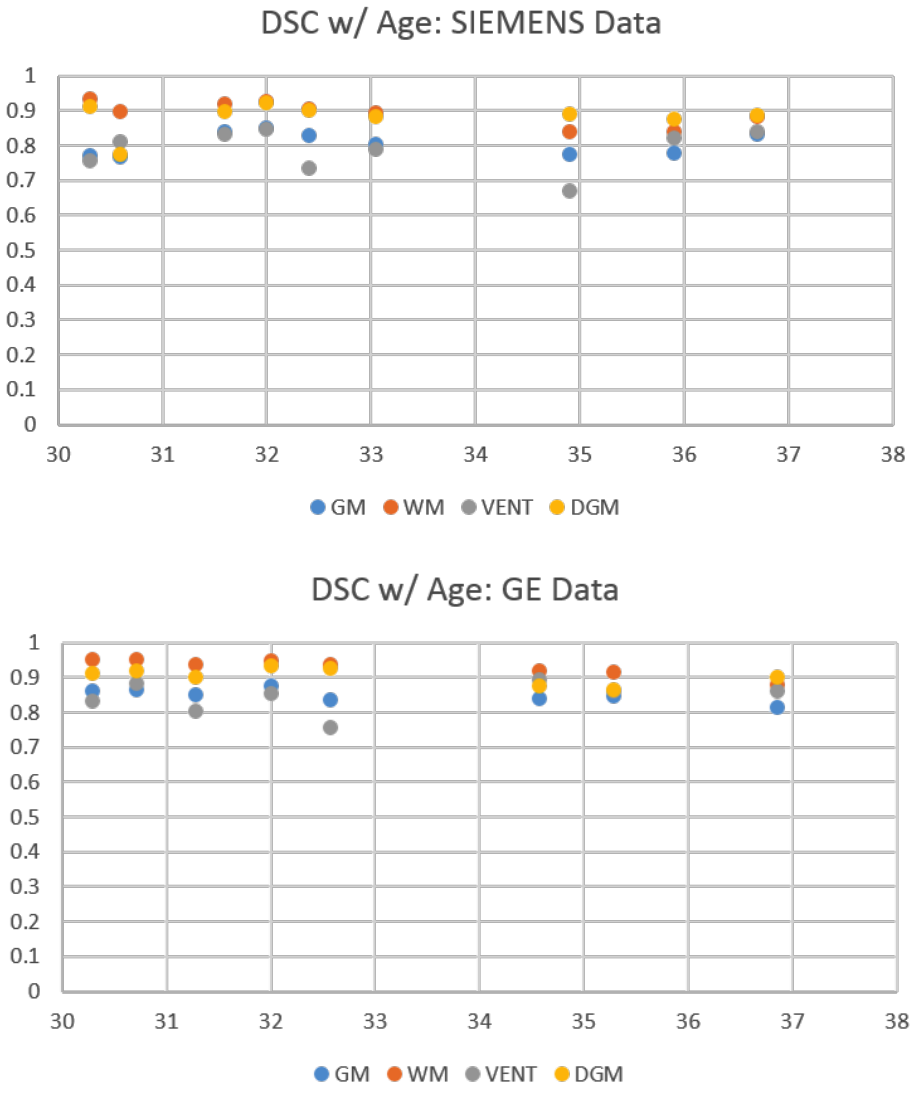

Figure 5.1: DSCs of 8 scans plotted with age in SIEMENS and GE datasets.

a 32-GW neonate in SIEMENS and GE data. The segmentations obtained were comparable both visually and DSC-wise.

These results showed that the previously describe PBAEM approach is capable to obtain comparable automated tissue segmentation results when applied on data acquired using different scanners and imaging protocols. Therefore, it is promising that, in studies involving multi-center cross-vendor cross-protocol large datasets, PBAEM techniques can be used to produce automated tissue labeling for any future analysis.

5.4 Discussion and Conclusion

The study in this chapter builds a bridge from the development automated tissue segmentation approach to applying it to study the early human brain development using multicenter data, which will be discussed in the next chapter. Only with such validation, are we confident to use the previously developed automatic image segmentation pipeline to process multi-center MRI data and extract quantitative measurements, such as tissue volume, cortical thickness and cortical surface area etc in the application section of this work.

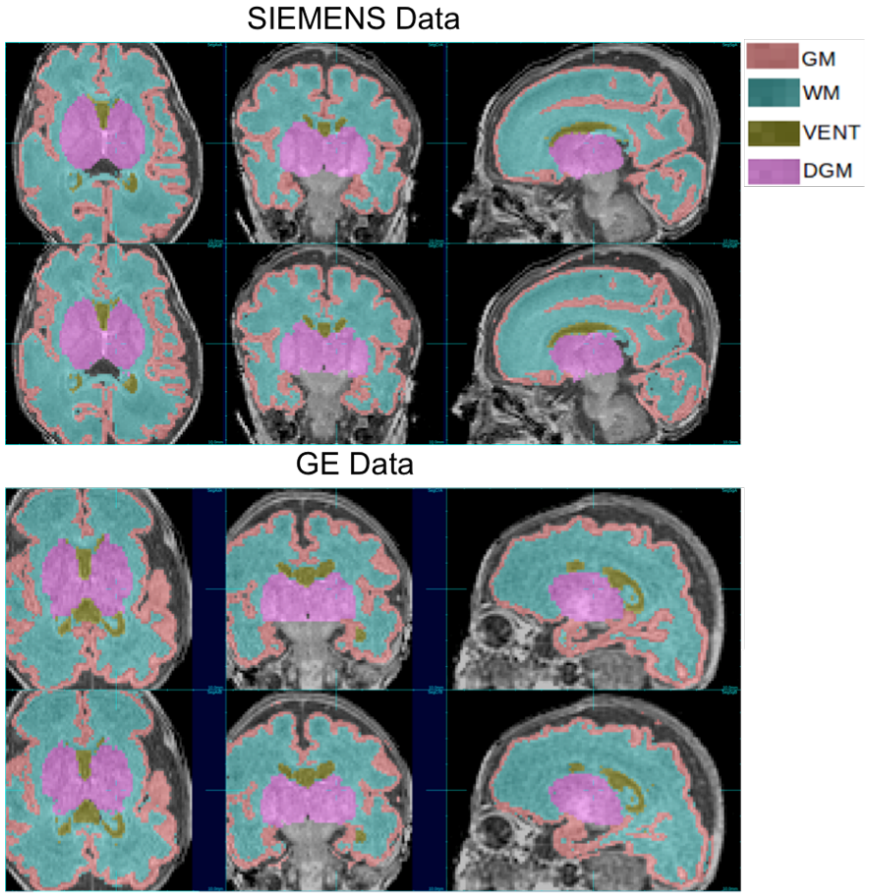

Figure 5.2: Manual (top row) and automated (bottom row) tissue labeling of a scan of a 32 GW neonate in SIEMENS (top) and GE (bottom) data. DSC: (SIEMENS) GM 0.8761, WM 0.9474, VENT 0.8533, DGM 0.9338; (GE) GM 0.8525, WM 0.9305, VENT 0.9036, DGM 0.9242.

Chapter 6

COMBINED SPATIAL AND NON-SPATIAL DICTIONARY SEARCH FOR FULL BRAIN AND INTRA-VENTRICULAR HEMORRHAGE (IVH) SEGMENTATION IN PREMATURE NEONATAL BRAIN MRI

6.1 Introduction

In neonates born prematurely, focal pathology is one of the abnormalities that can be visualized by medical imaging techniques. Magnetic Resonance Imaging (MRI) has been an essential tool to diagnose and monitor the condition, which can be greatly improved by an accurate volumetric and morphological analysis of the neonatal brain anatomy[62], especially in cases of high-risk preterm newborns[149]. Intra-ventricular hemorrhage (IVH) is one type of focal lesion and can often be accompanied by an enlargement of cerebral ventricles (VENT), termed Ventriculomegaly (VM)[139]. In these cases, automated tissue delineation can be badly impacted by the combination of large ventricles and multiple regions of IVH. The challenge of the task is two-fold. Firstly, the presence of IVH and VM makes the accurate non-rigid mapping of a normal atlas or even an IVH subject to a new IVH scan challenging because of the changes in topology (creation/deletion of regions) required to map between the anatomy with and without regions of IVH. Secondly, it is difficult to approach the problem by building an exhaustive dictionary that collects all possible shape, size and location of IVH and the enlarged VENT.

Due to the differences in tissue contrasts in premature neonatal brain imaging, specialized atlases and methods to use and validate them have been developed [65]. Cheng et al. proposed a stochastic process based approach for white matter injury detection in premature neonates [24]. Qiu et al. developed a multi-phase geodesic level-sets method that specifi-

cally targeted post-hemorrhagic ventricle dilation [101]. However, neither method labeled the normal tissue structures in the image. Wang et al. developed a patch-driven level set approach for normal term-birth neonatal T1-weighted MRIs [142]. Liu et al. [80] proposed to integrate a local patch-based search into a spatio-temporal atlas-based method to more accurately delineate the detailed structures in pre-natal scans of varying ages. In related work, Roy et al. presented a subject-specific sparse non-local dictionary learning approach for adult brain lesion segmentation [110]. Tong et al. proposed an approach to construct a discriminative dictionary to describe the hippocampus and used a sparse coding learning method to segment the hippocampus [132]. However, to the best of our knowledge, there has been little previous work in developing automated whole-brain segmentation methods for premature neonatal MRI scans with IVH and severe VM. We previously attempted aimed to automatically detect IVH voxels as outliers to the Gaussian Mixture Model [79]. Preliminary study showed reasonable segmentation accuracy with subject-level fine-tuned parameters but failed to generalize an universal parameter set.

In this work, we propose to utilize a specially designed dictionary, which consists of a spatial and a non-spatial component to account for both healthy and abnormal structures. The spatial dictionary encodes normal variation in anatomy, while we use a non-spatial dictionary to capture the shape and occurrence of IVH voxels with respect to their commonly neighboring tissues. An Elastic Net algorithm is used to ensure the sparsity of the dictionary learning in both the dictionaries. The two dictionaries are collectively used to estimate a probability of normal and abnormal tissues for each voxel, which is then used to initiate an Expectation-Maximization based tissue labeling of the image data[138, 10, 80]. We examine the behavior of this approach using a dataset of premature neonatal MRI scans with severe IVH and VM. Results indicate improvements in the segmentation accuracy compared to the conventional spatial-only dictionary learning segmentation methods. This approach provides the first automatic whole-brain segmentation framework for severe IVH and VM in premature neonatal brain MRIs.

6.2 Methods

6.2.1 Preliminaries

The problem being addressed is to assign an initial tissue probability map to a new unseen scan. Let I be the new image under investigation, R_h be the number of sets of lesion-free labelled MR template images $I^r(r = 1, ..., R_h)$ with labels $L^r(r = 1, ..., R_h)$, and R_l be the number of sets of labelled images with IVH and VM. At voxel location i of testing image I, its intensity patch of its $p \times p \times p$ neighboring voxels is represented as a column vector Y_i, and its corresponding dictionary is denoted as D_i with size d. The sparse dictionary search task is to determine the sparse coefficients β multiplied by which the dictionary can represent the image patch Y_i under investigation. We estimate β by solving a minimization of the non-negative Elastic-Net problem [153]:

$$\min_{\beta,\beta\geq 0} \frac{1}{2} \parallel Y_i - D_i\beta \parallel_2^2 + \lambda_1 \parallel \beta \parallel_1 + \frac{\lambda_2}{2} \parallel \beta \parallel_2^2 \tag{6.1}$$

The L_1-norm regularization ensures sparsity of β, and the L_2-norm regularization encourages similar dictionary patches to have similar coefficients. Conventionally in brain tissue segmentation, a spatial dictionary $\overline{D_i^{sp}}$ is constructed [142] to capture locally specific information. In this work, we consider parts of the anatomy for which we do not have enough training data to capture the full range of possible locations of pathology. We propose to use an additional non-spatial dictionary $\overline{D^n}$ which is combined with the spatial dictionary, such that $\overline{D_i} = \{\overline{D_i^{sp}}, \overline{D^n}\}$. This non-spatial component is used to augment the assignment of tissue labels where abnormalities are known to occur. In the following section, we focus on the construction of our proposed combined dictionary which includes spatial samples to match the normal tissue structures such as GM and WM, and non-spatially encoded samples of abnormal structures, i.e. IVH and VM.

6.2.2 Dictionary Construction

6.2.2.1 Non-Spatial Dictionary

The aim of this dictionary is to learn the appearance of focal pathologies and their occurrence with surrounding normal tissues, but to encode them without spatial constraints of where they may occur. This then can be used in regions where we assume the pathology can occur. In the problem considered here, the non-spatial component of the dictionary is constructed from ventricular regions with IVH and severe VM from lesion templates, i.e. $I^r, L^r, r = 1, ..., R_l$. For each voxel j within this region, we extract its $p \times p \times p$ intensity patch Y_j in the form of a column vector with unit L_2 norm, arranged to form a dictionary matrix $\widetilde{D^n}$. Due to the volume of severely enlarged ventricles, the number of columns (denoted as C) of this matrix can be large ($C \sim 10^4$) with many similar columns. To reduce computation time, we remove the duplicate dictionary samples while keeping the unique ones, by thresholding the similarity measurement between samples. We define the correlation between the u-th column and v-th column of $\widetilde{D^n}$ as $corr(\widetilde{D^n}(u), \widetilde{D^n}(v))$. Then we consider the v-th column of $\widetilde{D^n}$ as a duplicate of the u-th column and remove it if

$$\max_{c,c\in[1:C]} |corr(\widetilde{D^n}(v), \widetilde{D^n}(c)) - corr(\widetilde{D^n}(u), \widetilde{D^n}(c))| < a \tag{6.2}$$

where a is a chosen threshold. The use of correlation mimics the patch matching criteria in the LARS sparse dictionary search process [37]. After removing the duplicate dictionary samples, we obtain a succinct non-spatial dictionary $\overline{D^n}$, which is independent of voxel location x.

6.2.2.2 Spatial Dictionary

Using a conventional approach, the spatial component of the dictionary is constructed using lesion-free templates with similar gestational ages, i.e. $I^r, L^r, r = 1, ..., R_h$. For voxel i, we build its spatial dictionary as follows. Let $\mathcal{N}_i^r$ denote the $n \times n \times n$ neighborhood of voxel i in r-th ($r = 1, ..., R_h$) template image. For each voxel $j \in \mathcal{N}_i^r$, we extract its intensity patch

from I^r, normalize it to have a unit L_2 norm and then rewrite it into a p^3-sized column vector Y_j. By arranging $n^3 \times R_h$ column vectors, we obtain the spatial dictionary matrix $\overline{D_i^{sp}}$ for each voxel i.

6.2.2.3 Combined Dictionary

For each voxel the correspondent dictionary $\overline{D_i}$ is the combination of the spatial and non-spatial component: $\overline{D_i} = \{\overline{D_i^{sp}}, \overline{D^n}\}$. To further simplify computation, we conduct a pre-screening of the mean intensity of the dictionary patches. We exclude the dictionary patch at v-th column of $\overline{D_i}$ if

$$\frac{|avg(\overline{D_i}(v)) - avg(Y_i)|}{avg(Y_i)} \geq b \tag{6.3}$$

where $avg()$ computes the mean intensity of the patch intensities before unit L_2-normalization, b is a chosen threshold. Another benefit of the mean pre-screening is to remove the confusion caused by the dictionary sample patches with similar intensity pattern but very different absolute intensity level. For example, an uniform patch inside VENT should not be matched to the uniform patches inside WM with their similar pattern but different absolute intensity. After this, we have the final dictionary D_i for voxel i.

6.2.3 Implementation Details

6.2.3.1 Pre-processing

To construct the dictionary, we first linearly align all training images I and globally standardize the intensity scaling factor. For the non-spatial component, we extract voxels label as IVH or VENT (Fig. 6.1 (1-a)), and smoothly dilate the region to include the outer boundaries (Fig. 6.1 (1-b)). An example non-spatial dictionary $\overline{D^n}$ obtained after the duplicate removal process is shown in Fig. 6.1 (1-c). A conventional atlas-based automated segmentation is used to provide outer cerebral boundary.

6.2.3.2 Sparse Dictionary Search using LARS

For the sparse dictionary search, we use the combined dictionary for regions inside the cerebral boundary where the pathology can occur, and spatial-only dictionary for the other regions to save computation. The Elastic-Net problem (Eq. 6.1) is a convex optimization problem and, in our implementation, β is solved by the LARS algorithm with non-negative constraints [37]. Each element of β represents the similarity between Y_i and the corresponding dictionary sample $Y_j, j \in N_i^r$. For LARS, the similarity is based on correlation for matching the pattern in the two patches. Under the assumption that similar patches share the same tissue label, we can compute the estimate the tissue probability $P(k|Y_i)$ of the voxel i belonging to tissue class k, from β as follows:

$$P(k|Y_i) = \frac{\sum_{i=1}^{d} \beta_i L_i}{\sum_{i=1}^{d} \beta_i} \tag{6.4}$$

6.2.3.3 Post-Dictionary EM Segmentation

The patch-based dictionary-learnt tissue probability estimate is used to initiate an EM-based tissue labelling framework. The EM algorithm clusters the voxels with similar intensities into same tissue classes given the prior tissue estimates. The final automated tissue labeling is obtained. Our full segmentation driven has in total 8 labels. In the following section, we focus only on the 5 cerebral tissue structures that contain lesions: GM, WM, VENT, DGM and IVH.

6.3 Experimental Results

6.3.1 Dataset and Validation

Our data consists of a total of 12 T1-weighted MR scans of premature neonatal brains with manual tracing into GM, WM, VENT, DGM, CBL, BS, sCSF and IVH, 4 of which have IVH and severe VM. To test the approach we used 2 age matched normal scans to construct a spatial dictionary and 3 out of 4 of the IVH scans to construct the non-spatial dictionary

Test Case	Test Scan	Spatial Dict. (2 scans)		Non-Spatial Dict. (3 scans)		
1	28.7 GW	28.6 GW	29.4 GW	29.4 GW	32.7 GW	33.6 GW
2	29.4 GW	28.6 GW	29.4 GW	28.7 GW	32.7 GW	33.6 GW
3	32.7 GW	32.6 GW	33.1 GW	28.7 GW	29.4 GW	33.6 GW
4	33.6 GW	33.1GW	33.7 GW	28.7 GW	29.4 GW	32.7 GW

Table 6.1: Gestational ages of testing datasets and the corresponding dictionary data.

leaving one to be automatically segmented. This was repeated for each of the 4 IVH cases and Dice Similarity Coefficients (DSC) calculated against the corresponding manual tracing. The experimental data is summarized in Table. 6.1. The Dice Similarity Coefficient (DSC) between the manual and automatic tissue labelings of the lesion scans were calculated to quantitatively evaluate the proposed segmentation method.

6.3.2 Parameter Selection

Optimal parameter values were determined by leave-one-out cross-validation on all 4 IVH scans. Values for L_1-regularization coefficient $\lambda_1 = 0.1, 0.2, 0.3, 0.4, 0.5$, patch size $p = 3, 5$ and neighborhood size $n = 3, 5, 7$ were compared using DSC (Fig. 6.2), finding optimal values: $\lambda_1 = 0.2, p = 5, n = 7$. We also tested on a smaller scale and chose L_2-regularization coefficient $\lambda_2 = 0.01$ as in [142], dictionary thresholds $a = 0.04$ and $b = 0.2$. The impact of duplicate dictionary sample removal using $a = 0.04$ is shown in Fig. 6.1(2)(3).

6.3.3 Results

To show the contribution from the spatial and non-spatial components of the dictionary, we compare the number of positive coefficients in the spatial and non-spatial part of β for each voxel in Fig. 6.4, confirming that for normal tissue the primary contribution is from the spatial dictionary, while locations with abnormal ventricles or IVH are determined by the non-spatial dictionary.

Scan	Before EM Algo					After EM Algo				
	IVH	VENT	GM	WM	DGM	IVH	VENT	GM	WM	DGM
1	0.5232	0.7802	0.8051	0.8937	0.8595	0.6522	0.9524	0.8849	0.9520	0.8951
2	0.7409	0.8630	0.7525	0.8756	0.8991	0.8813	0.9386	0.8885	0.9377	0.9313
3	0.7649	0.8583	0.7822	0.9071	0.8811	0.8746	0.9144	0.8660	0.9434	0.8982
4	0.7432	0.8526	0.8014	0.9319	0.9011	0.8436	0.9230	0.8724	0.9564	0.9219
Avg	0.6931	0.8385	0.7853	0.9021	0.8852	0.8129	0.9321	0.8780	0.9474	0.9116

Table 6.2: Comparison of individual and average DSC of 5 main tissue classes obtained before using EM clustering (left section) and after EM clustering (right section).

To show the accuracy of the sparse coded tissue probability and the effect of the EM algorithm, we compared the before-EM and after-EM DSC as shown in Table. 6.3. We can clearly see that dictionary labelling provides an accurate initial tissue label estimate. EM then adapts this labeling further and allows modeling of subtle residual bias field to improve the final labels.

Fig. 6.5 summarizes the key results with red arrows indicating where the combined dictionary improved performance. In particular scan #1 in Fig. 6.5 illustrates a case where the IVH location was not present in the training data leading to a failure in the spatial-only dictionary approach, but a correct labelling when also using the non-spatial dictionary.

We further compare the average DSC of 5 main cerebral tissue classes between using the combination of spatial and non-spatial dictionary and using only spatial dictionary in Table. 6.3. Significantly higher segmentation accuracy in IVH was obtained by the proposed sparse combined dictionary search method compared to using the spatial-only dictionary, while comparable accuracy was obtained for the other tissue classes.

6.4 Discussion and Conclusion

This paper describes a novel hybrid technique to address segmentation of highly variable focal abnormalities that is motivated by the study of abnormally developing premature neonatal

Scan	Spatial + Non-Spatial Dictionary					Spatial-only Dictionary				
	IVH	VENT	GM	WM	DGM	IVH	VENT	GM	WM	DGM
1	0.6522	0.9524	0.8849	0.9520	0.8951	0.4833	0.9608	0.8872	0.9599	0.9405
2	0.8813	0.9386	0.8885	0.9377	0.9313	0.9234	0.9509	0.9047	0.9554	0.9429
3	0.8746	0.9144	0.8660	0.9434	0.8982	0.8373	0.9209	0.8672	0.9435	0.9064
4	0.8436	0.9230	0.8724	0.9564	0.9219	0.8329	0.9247	0.8815	0.9589	0.9262
Avg	0.8129	0.9321	0.8780	0.9474	0.9116	0.7692	0.9393	0.8852	0.9544	0.9290

Table 6.3: Comparison of individual and average DSC of 5 main tissue classes obtained by using the proposed combined dictionary (left section) and spatial-only dictionary (right section).

brain anatomy. The proposed method seeks to label brain anatomy with a tissue probability using a collective sparse search of a combined spatial and a non-spatial dictionary to provide a more accurate estimate of the tissue labels, for both focal lesions and surrounding normal tissues. The spatial component represents the normal anatomical variations and the non-spatial component encodes the variable appearance of IVH and VM. Experimental analysis of the results of EM segmentation driven by this prior, compared against manually delineated premature neonatal brain MRIs indicated improved performance.

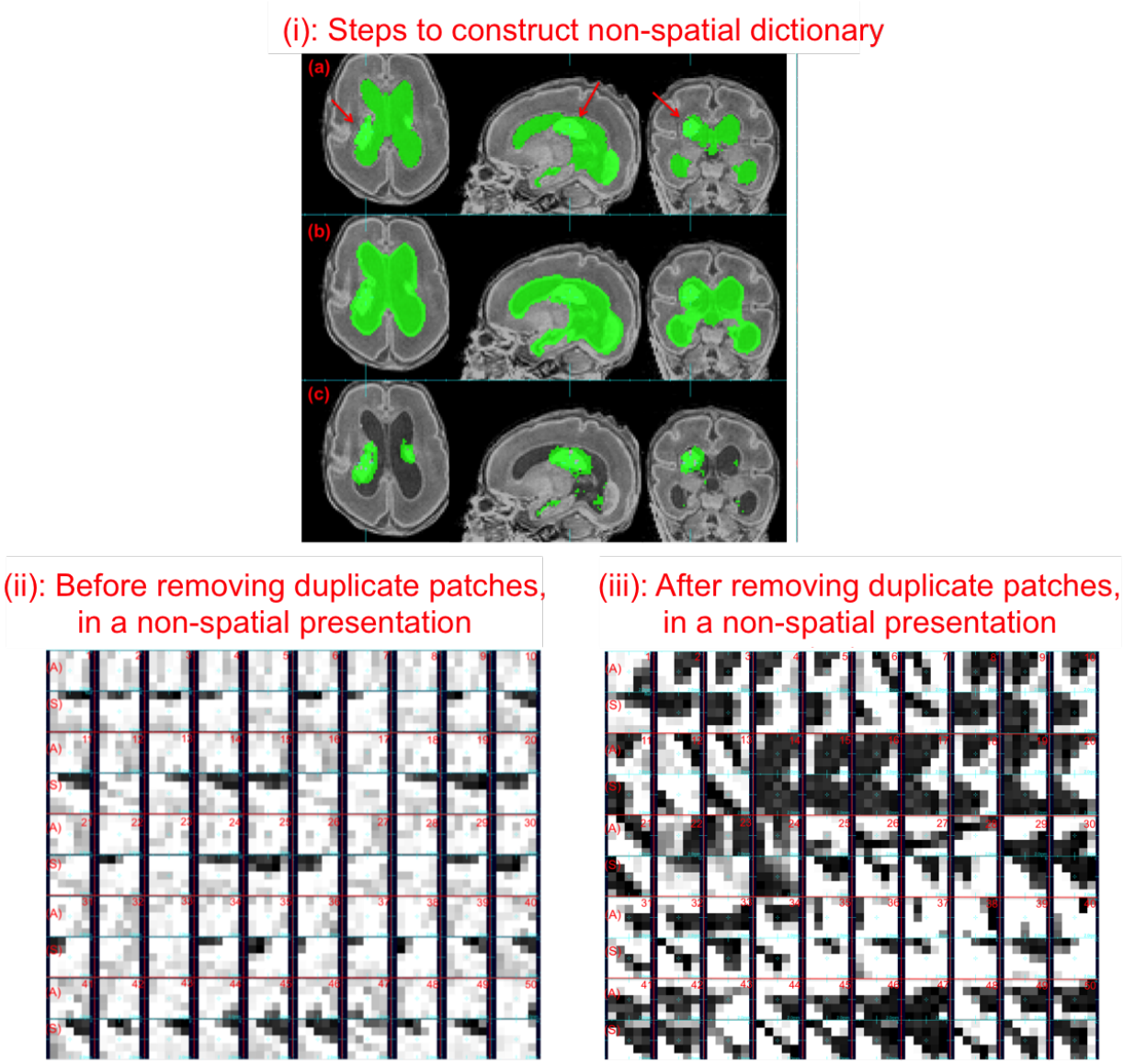

Figure 6.1: Non-spatial dictionary construction. (1): Example showing the construction of non-spatial dictionary in the form of a mask (green) overlaying the subject MRI. (a) IVH and VENT mask extracted from manual labeling; (b) dilated mask that includes duplicate voxels with similar intensity profile; (c) remaining voxels after removing duplicates. Red arrow: IVH. (2)(3): 50 randomly selected sample patches (shown in axial and sagittal view) in the non-spatial dictionary before (2) and after (3) removing the duplicate patches. It is clearly shown that, before removal (2), more patches share the same intensity profile and will hence contribute same information to the non-spatial dictionary while unnecessarily increasing the computation time. After removal duplicates (3), we obtain more structural diversity given the same number of dictionary samples.

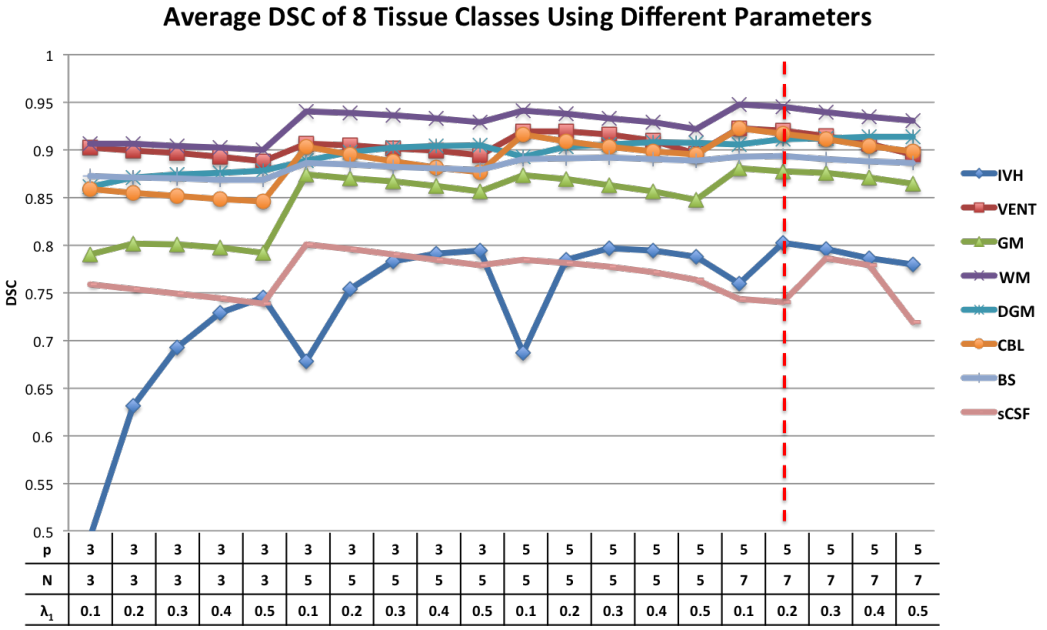

Figure 6.2: Average DSC of 8 tissue classes (IVH, GM, WM, VENT, DGM, BS, CBL and sCSF) with respect to the different combinations of parameters λ_1, p and N.

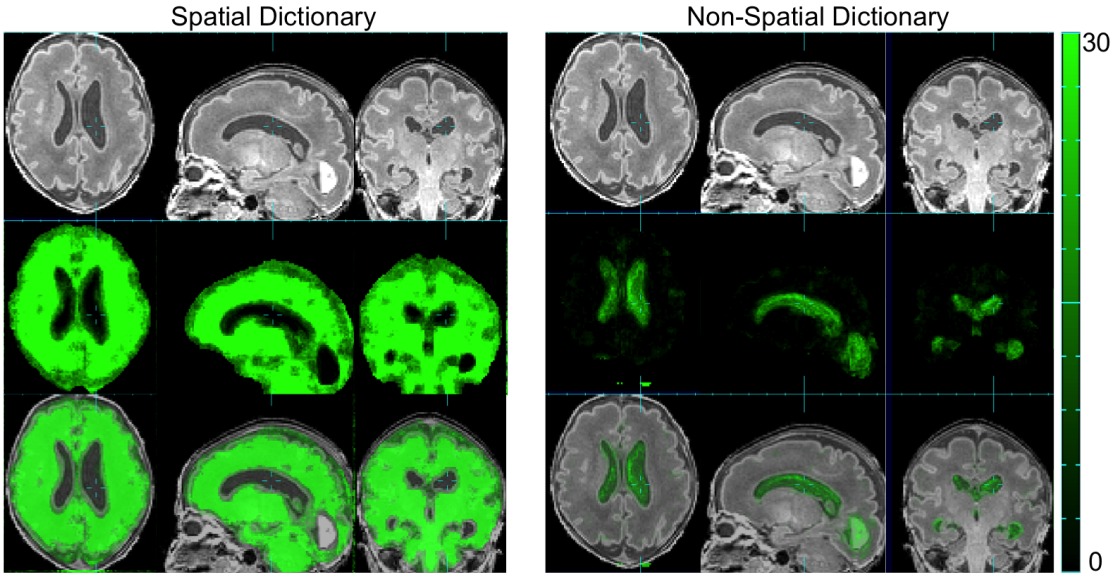

Figure 6.3: Number of positive similarity coefficients from spatial (left) and non-spatial (right) dictionaries for each voxel. The green level indicates the number of positive coefficients. The greener a voxel is, the more matches were found in the dictionary for that voxel.

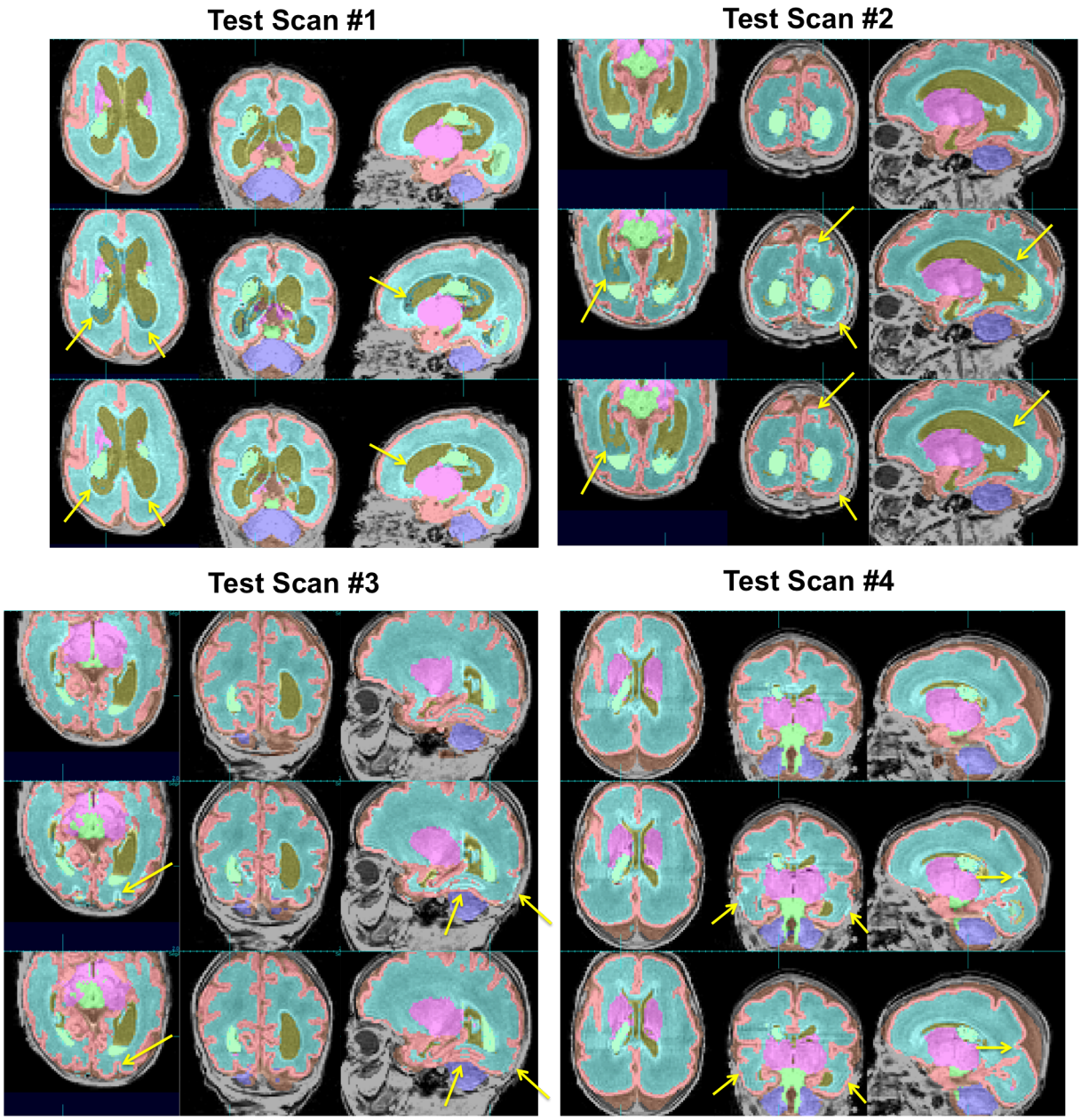

Figure 6.4: Comparison of manual (top row), before EM (middle row) and after EM (bottom row) tissue segmentation of all 4 testing scans using combined dictionary. Yellow arrow: the normal tissue region that is correctly labeled after EM clustering while mislabeled before.

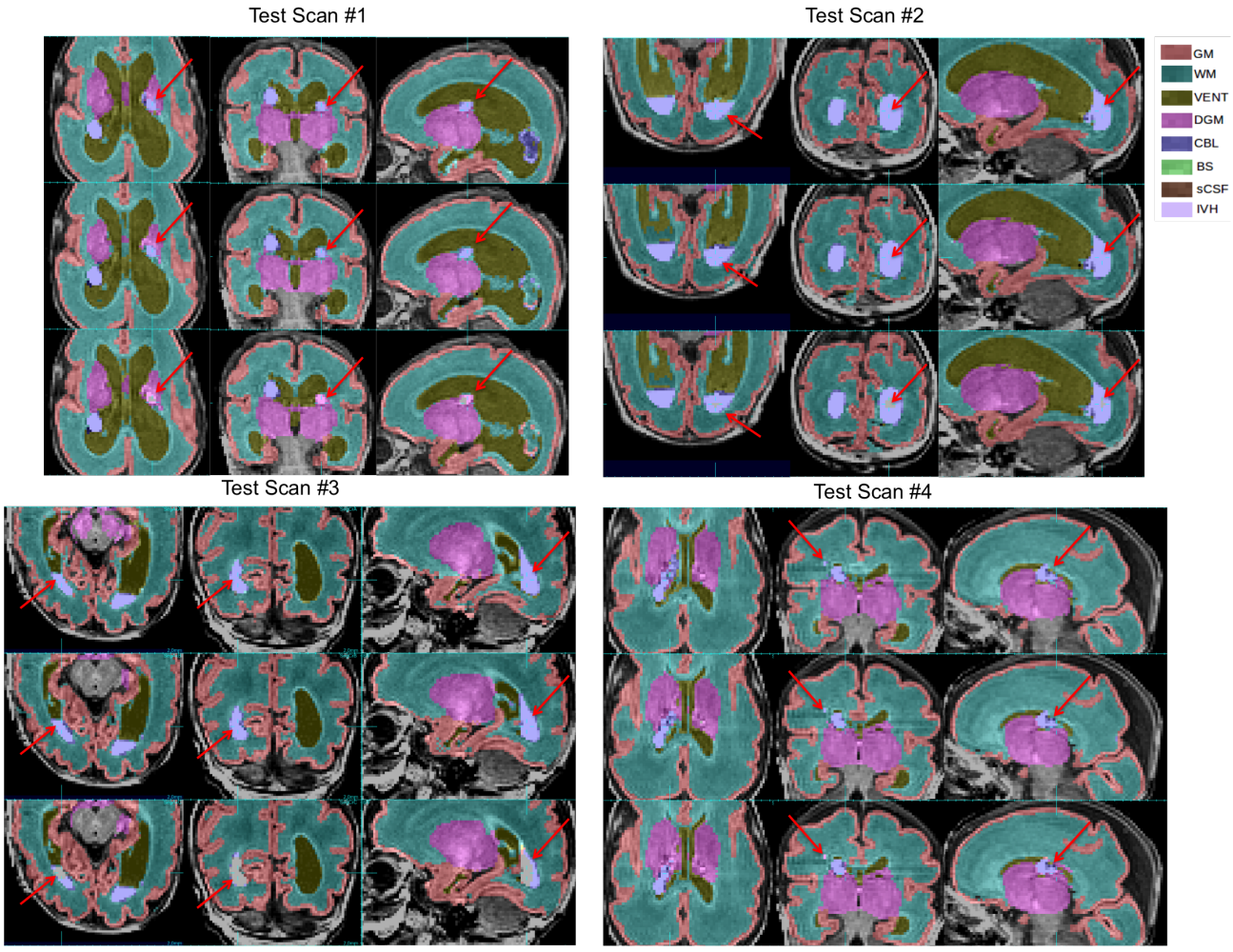

Figure 6.5: Comparison of manual (top row), automatic using spatial and non-spatial dictionary (middle row) and automatic using only spatial dictionary (bottom row) tissue segmentation of all 4 testing scans. Red arrow: the IVH region that is correctly labeled using the proposed combined dictionary while mislabeled using spatial-only dictionary.

Chapter 7

APPLICATIONS OF METHODOLOGY: EARLY HUMAN BRAIN DEVELOPMENT AND NEURO-DEVELOPMENTAL OUTCOMES

In this chapter, I present the collaborative work that has been done to study the early brain development and neuro-developmental outcomes based on quantitative MRI studies with the help of the previously described automated tissue segmentation algorithms. Results from several collaborative studies from both our group and collaborating groups are presented. Note that most of these studies used volumetric measurements that I extracted using the PBAEM or sparse dictionary search automated segmentation, and other collaborators completed the subsequent analysis and interpretation of results as notated in the subsection titles. I only have *partial* contributions to the work presented in this section.

7.1 Background

In the third trimester of pregnancy, the brain rapidly grows. Abrupt and premature termination of pregnancy during preterm birth would result in delayed development of human brains. Understanding the impact of preterm birth on neuro-development is key to early detection and prediction of brain function deficits in premature infants.

One region that grows particularly fast in the third trimester is the cerebral cortex (see Figure 7.1). Previous studies have shown that, in preterm births, the growth of cortical surface area slows down relative to total brain volume growth. Such reduced growth correlates with lower PMA and lower global abilities at 2 and 6 years of age, particularly with respect to complex cognitive functions but not motor functions[103, 68].

The cerebellum also undergoes proliferative growth in the third trimester, with a 5-fold

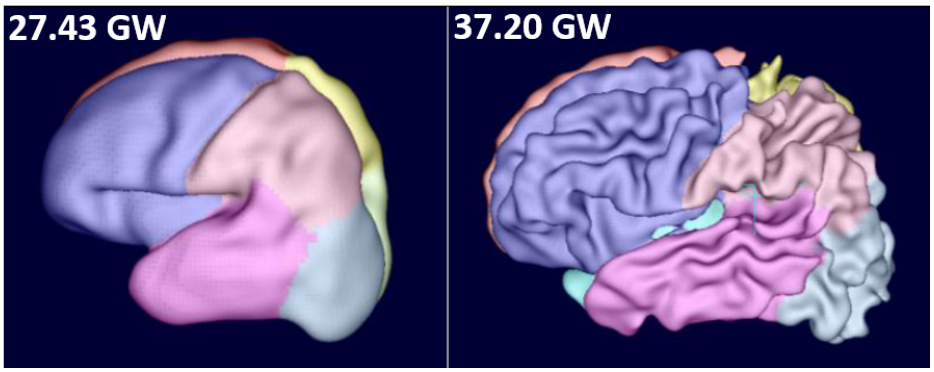

Figure 7.1: Comparison of cortical smoothness between subjects at 27.43 GW (left) and 37.2 GW (right) [74].

increase in volume from 24 to 40 GW. Therefore, it is particularly vulnerable to preterm birth[141].

Many anatomical abnormalities are prevalent in premature neonates, for example, IVH and WMI, which usually result in delayed brain maturation, even far from the lesion site, as well as other adverse outcomes[89, 73]. Tam et al. reported that decreased cerebellar volume was associated with IVH[130]. Chau et al. reported that abnormal brain maturation involving WMI studied serially using DTI and MRSI was associated with adverse cognitive, language and motor outcomes at 18 months corrected age[23]. In a long-term follow-up study, Futagi et al. reported that approximately 50% of children who had IVH as infants, suffered from neuro-developmental illness, including cerebral palsy (22.4%), mental retardation (10.2%) and borderline intelligence (11.3%)[42]. Furthermore, different grades of IVH resutled in statistically different outcomes[42]. Therefore, understanding how different grades of lesions can affect the course of neuro-development is also critical to predict outcomes at a later age.

In our study, Bayley Scales of Infant and Toddler Development - Third Edition (BSID-III) [3] was administered for each subject at 18 months and used as a quantitative neuro-development measurement. BSID-III examined three distinct functions: cognitive, language

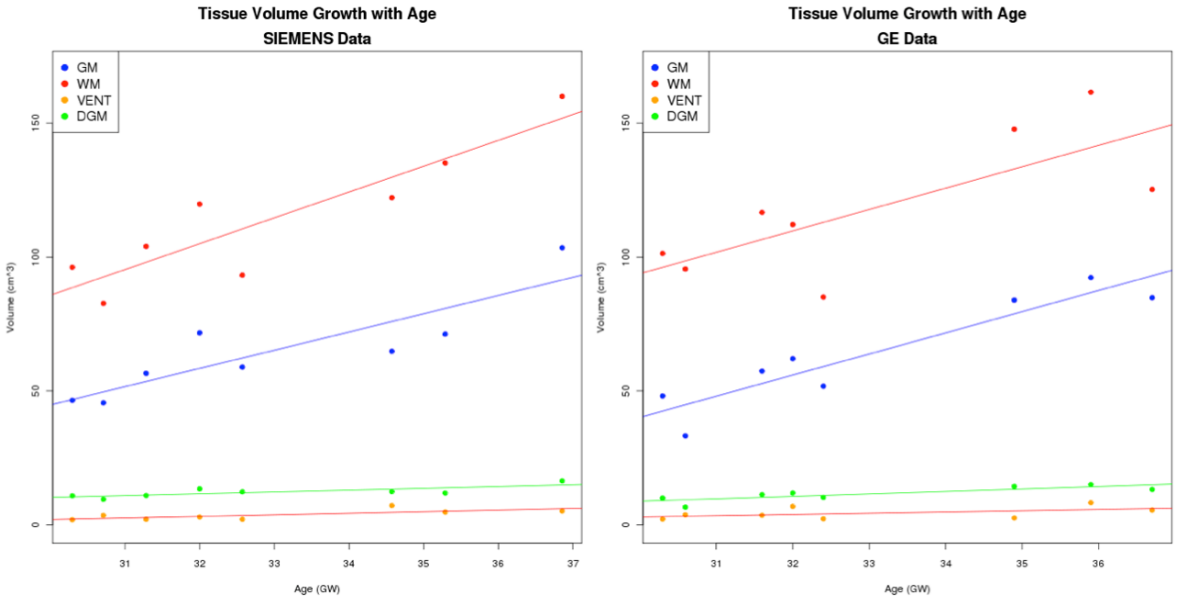

Figure 7.2: Tissue volume growth with age of the SIEMENS (left) and GE (right) dataset.

and motor.

7.2 Brain Tissue Growth

Based on the results obtained from the validation of PBAEM automated segmentation approach using multi-center data, the tissue volume growth trajectories were plotted and compared between the SIEMENS and GE datasets (see Figure 7.2). Linear regression models were used to model the volumetric growth over age of GM, WM, VENT and DGM respectively. The resulting linear trends are comparable between the two datasets. The results, on one hand, further validate the comparable segmentation performances on two datasets; on the other hand, they serve as a preliminary study of cerebral tissue volume growth between 30 to 37 GW.

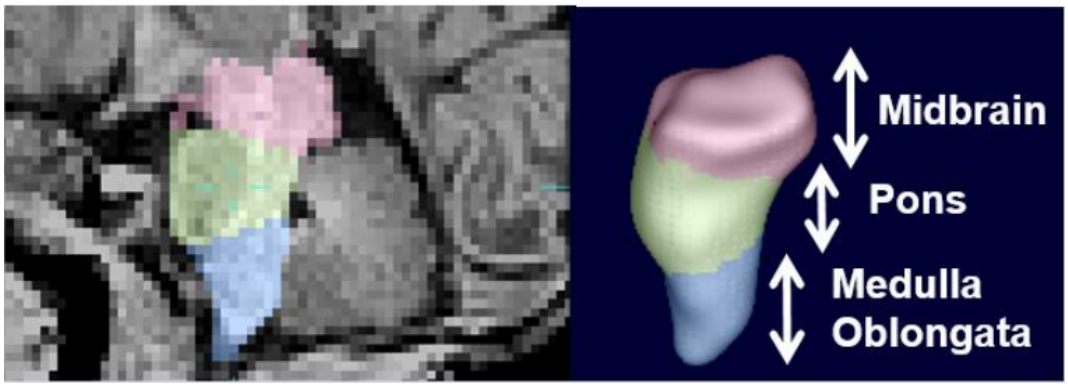

Figure 7.3: Three structure parcellations of brainstem. Figure adapted by author from [58].

7.3 Neuro-Developmental Outcomes [Co-investigator: Averi Kitsch, BS; Lisa Harrylock, BS; Nicole Riley, BS]

7.3.1 Population & Quantitative Measurement Extraction

To maximize the statistical power, a total of 269 premature neonatal T1w brain MR scans were studied. 269 scans came from 152 premature neonates, 117 of which were scanned twice at birth age and term-equivalent age. 269 scans consisted of scans of 122 males and 145 females. All MR scans were acquired using the same scanner and protocol (see Table. 5.2 SIEMENS dataset). The proportion of infants that scored one SD below the mean (85) were: cognitive 24.0%, language 38.6% and motor 38.5%.

Automated tissue segmentations were obtained using the PBAEM-WF approach with a 28-scan spatio-temporal atlas. Additionally, the cerebrum was parcellated into 8 lobes (left and right frontal, temporal, parietal and occipital lobe) using registration-based label propagation (Figure 4.9). Three brainstem structures (midbrain, pons and medulla oblongata) were also automatically parcellated (Figure 7.3) in the parcellation pipeline to allow for analysis of regional volumetric data. Scans with poor tissue segmentations were excluded. The volumes of tissue classes and parcellations were extracted.

7.3.2 Brainstem Volume

The brainstem plays an important role in motor and sensory functions and thus is important for studying early brain development. In this initial attempt, we looked at the correlation between volumes of three anatomical structures, i.e., midbrain, pons and medulla oblongata, as well as neuro-developmental outcomes at 18 months in order to study their predictive power[58]. Linear regression models were used to analyze the correlation between structure volumes and neuro-development. The multivariate regression model co-varied for gestational age at scan and supratentorial volume:

$$\begin{aligned} TissueVolume = a \times ScanAge + b \times BirthAge + c \times Gender \\ + d \times SupratentorialVolume + e \times BSID - III + f \end{aligned} \tag{7.1}$$

The regression results are listed in Figure 7.4. The total brainstem volume was found to be significantly ($p < 0.05$) correlated with decreased Bayley scales at birth but not term-equivalent age, while no statistical significant correlation was found for total supratentorial. This shows that brainstem volume is predictive of neuro-developmental outcome at a later age. As for the sub-structure in the brainstem, the pons was found to be significantly ($p < 0.05$) associated with motor function at term age, while the medulla oblongata was found to be significantly associated with all three neuro-cognitive functions at both birth and term age. No significant association was found for the midbrain. The results indicate that the brainstem is a promising candidate to serve as a neuro-cognitive predictor, while analysis for more detailed sub-structures within the brainstem is not required except for the meduall oblongdata.

7.3.3 Regional Tissue Volume

The predictive power of supratentorial and lobe-wise GM and WM volumes were also investigated preliminarily as a collaborative study within our research group[74]. Linear regression models were used to analyze the correlation between tissue volumes and neuro-development to identify potential biomarkers for brain health. The multivariate regression model co-varied

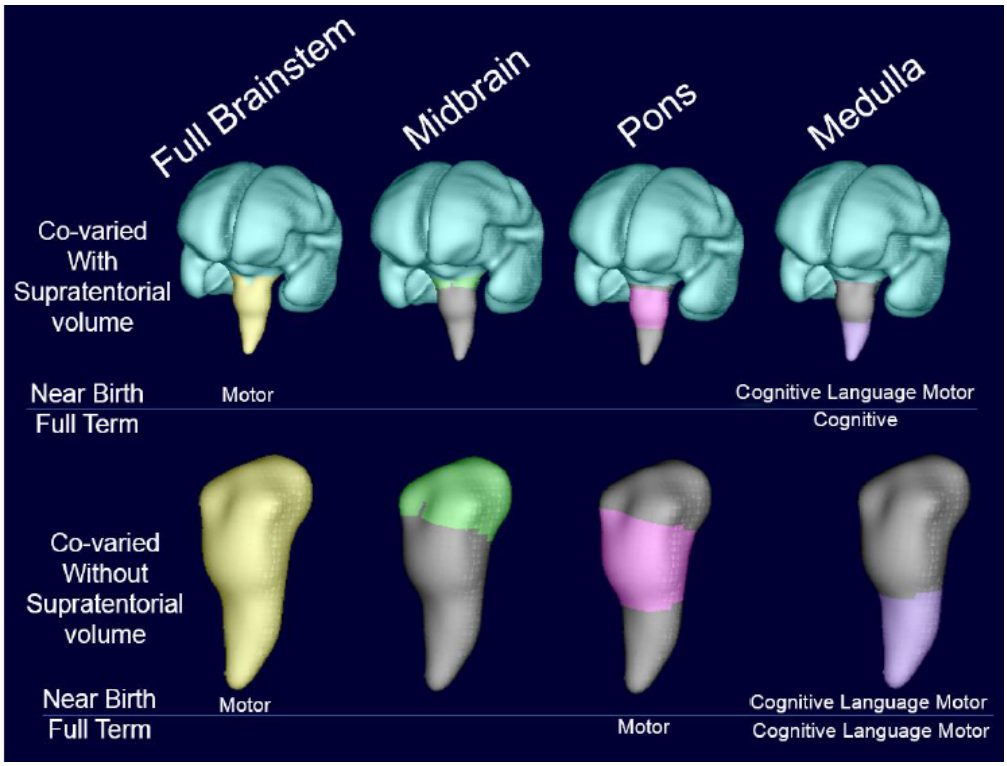

Figure 7.4: Results from multivariate analysis of brainstem volumes as a function of age at scan, age at birth and supratentorial volume (upper row only). Significant models ($p < 0.05$) are listed underneath the structure. Figure adapted by author from [58]

.

for gestational age at birth, gender, age at scan and specific BSID-III scores:

$$TissueVolume = a \times ScanAge + b \times BirthAge + c \times Gender + d \times BSID - III + e \tag{7.2}$$

The analysis found that the extent of prematurity is strongly correlated to decreased neuro-cognitive abilities. The regions with statistical significance ($p < 0.05$) are listed in Figure 7.5. As we can see, the regional volumetric measurements of GM and WM provide better prediction of neuro-cognitive delay at 18 months than total tissue volumes. The parietal and occipital lobe volumes correlate with motor and language neuro-cognitive scales, while the temporal lobe is more associated with cognitive functions, and the frontal lobe is only associated with motor abilities. The results from this preliminary study show the promise in quantitative measurements from more anatomically specific brain structure providing more accurate outcome prediction of neuro-development after preterm birth.

7.3.4 Occipital Lobe Subdivision Volume

In this work, we specifically investigated one of the 4 lobes, the occipital lobe, where the GM volume was found to be associated with both language and motor skills at 18 months in the work above. It has also been established that the occipital lobe is associated with visual outcomes [131]. In particular, inferior occipital lobe volumes have been reported to be correlated with visual measures (such as rate of saccade) at 2 years old [122]. In the work by our group, Riley et al. [106] studied the relationship between the focal anatomy of the occipital lobe and neuro-developmental outcomes at 18 months in premature neonates.

A sub-parcellation protocol was first defined for the occipital lobe as shown in Figure 7.6. The calcarine sulcus was used to define the boundary between the cuneus region and the lingual region. Then the automatic tissue segmentation and lobe parcellation pipeline was applied with an extension to include new occipital lobe sub-divisions. The DSC of the 32 scans that were used to construct the spatio-temporal atlas was used to quantitatively

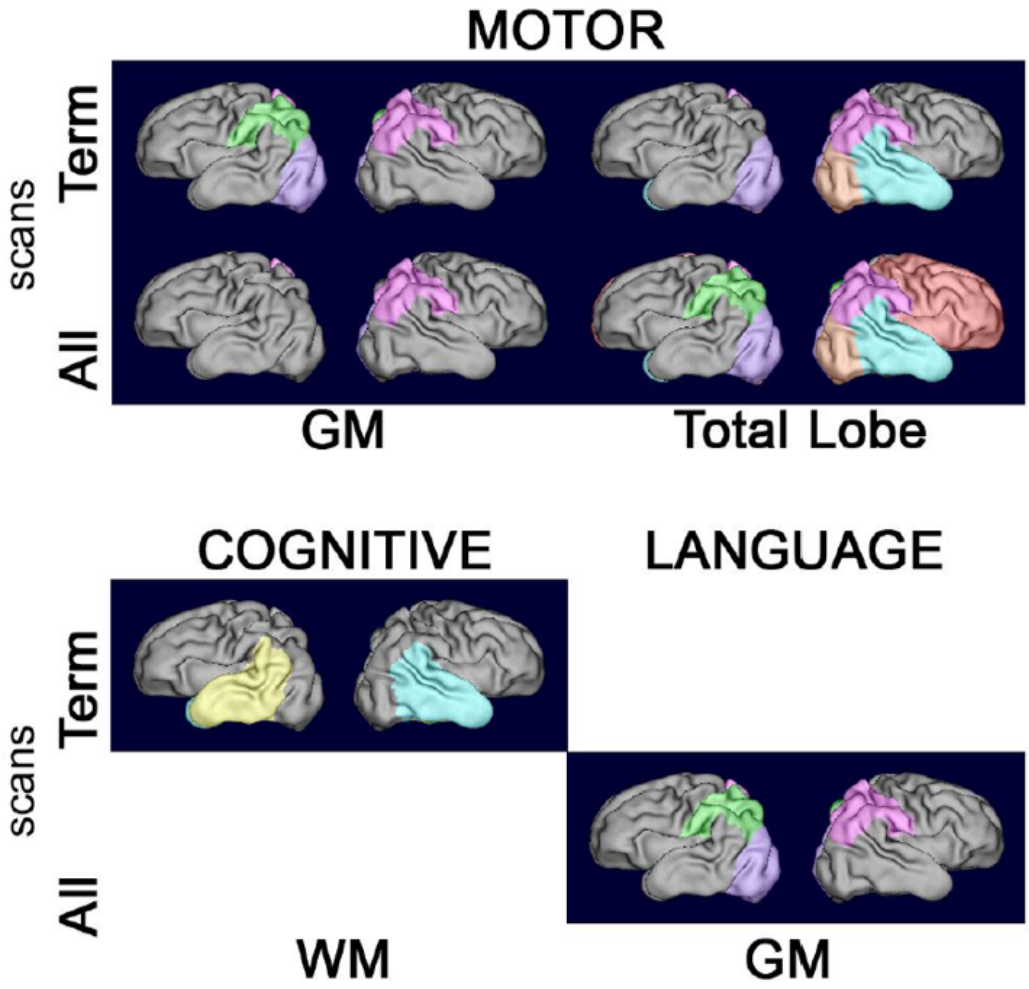

Figure 7.5: Lobe parcellation volumes, by hemisphere, of GM, WM and LOBE (combined GM and WM) that predict certain neurocognitive abilities at 18 months. Term: only term-equivalent age scans were used; All: both scans at birth age and term-equivalent age were used. Figure adapted by author from [74].

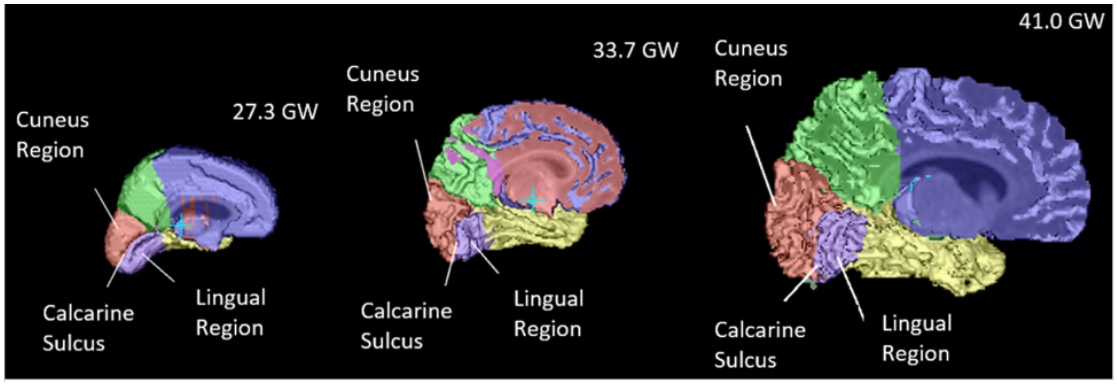

Figure 7.6: Occipital lobe manual sub-division protocol on the left cerebral cortex at different ages. The calcarine sulcus divides the occipital lobe into two parts: the cuneus region and the lingual region. [106].

demonstrate the validity of the automatic segmentation and parcellation pipeline. Volumetric measurements were then extracted from the entire dataset of 170 scans.

Regression analysis was conducted similarly to previous work. The multivariate regression model co-varied for gestational age at birth, gender, age at scan, head size and specific BSID-III scores:

$$\begin{aligned} CorticalGMVolumeofOccipitalLobeRegion = a \times ScanAge + b \times BirthAge + c \times Gender \\ + d \times HeadSize \\ + e \times BSID - III + f \end{aligned} \tag{7.3}$$

The analysis found that cortical GM volume in the lingual region is associated with motor and cognitive functions with statistical significance $p < 0.05$, while both lingual and total cortical GM volumes are associated with language function with statistical significance $p < 0.05$ at 18 months. The results suggested the potential predictive power of focal anatomy or lobe subdivision for prediction of later neuro-developmental outcomes.

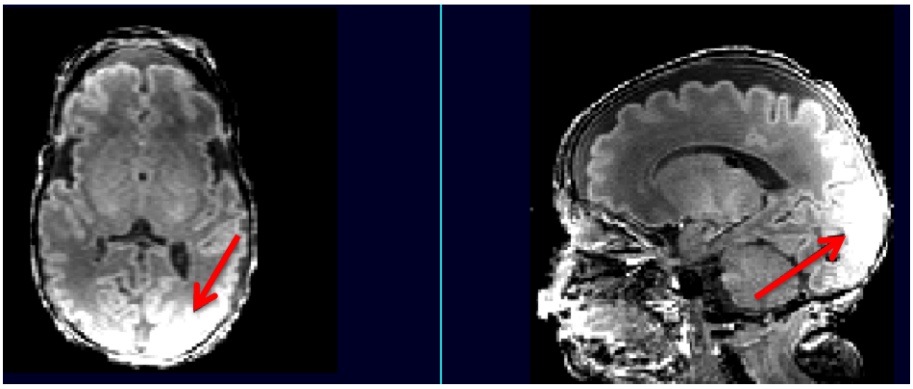

Figure 7.7: An example of severely bias-corrupted premature neonatal brain T1w brain MRI acquired on a Siemens Verio system using a 15 channel knee coil from PENUT [115] study [150]. Red arrow: hyper-intense WM that may be mis-labeled as GM without bias correction

7.4 Refined pipeline for Clinical Bias-Corrupted MRIs Acquired without Dedicated Neonatal Coils [Co-investigator: Zachary Hill, BS]

To reiterate, the bias field is the intensity variation across the field of view which can cause one tissue class to have highly varied intensity levels. Even though bias correction is incorporated into the EM segmentation pipeline, it can only handle bias correction to an extent. In the previous pipeline, the bias correction is carried out after estimating the initial tissue probability, so we assume that the accuracy of the tissue probability estimate accuracy is not affected by the bias field dramatically. However, when an MRI contains extreme bias corruption as in Figure 7.7, there is limitations to the accuracy of the initial tissue probability. For example, the hyper-intensive patch in WM might be mis-categorized as GM. Therefore, a step of bias correction before the patch-based search is required to ensure accurate tissue probability estimates can be found.

This is particularly an issue for large-scale multi-center studies when image qualities are mixed between centers, especially when certain centers use older scanners with coils that are not specifically designed for neonatal brain imaging. As a matter of fact, even at modern hospitals, dedicated neonatal imaging techniques are not always available. Such imaging can

result in severe bias field and fundamental differences in contrast which can result in failures in the previously described segmentation pipeline. The mutli-center MRI scans from the PENUT study[115] suffer from this issue.

Hill et al. [150] proposed a hybrid MRI tissue segmentation pipeline which integrated PBAEM with an additional bias correction step. To begin, an initial rough template-based bias correction was carried out. We first non-rigidly align the age-specific MR template with the scan to label. Then the relative distortion between the scan and AS template was estimated by dividing the 5th-order polynomial bias field models fitted to both images. This coarse bias estimate was then used to rescale the intensities in the scan to label. To further refine this estimate, a second round of the same process was carried out using the coarsely bias-corrected MRI from the first round. This refined bias-corrected MRI was used as the MRI to segment in the previously described PBAEM pipeline.

The proposed hybrid pipeline was evaluated on 4 highly bias-corrupted T1-weighted coronal brain MRI images (TR=50ms, TE=9.2ms, 3D FLASH, voxel size $1 \times 1 \times 1 mm^3$) from the PENUT study[115]. The scans were acquired with a 3T Siemens Verio system (B17 software) using a 15 channel TX/RX adult knee coil to provide improved signal-to-noise ratio in the premature neonates. DSC and HD between the manual segmentation and automatic segmentation were calculated. Comparison showed significant improvement from the proposed hybrid method on top of PBAEM (see Figure 7.8).

In summary, a refined pipeline was proposed based on PBAEM to achieve accurate segmentation on severely bias-corrupted MRI scans. This is particularly useful in large-scale multi-center studies when some centers are not equipped with dedicated neonatal head coil. The results show significant improvement in segmentation accuracy when compared to conventional PBAEM. The proposed method will be used in the group analysis of the PENUT study as introduced in the following section.

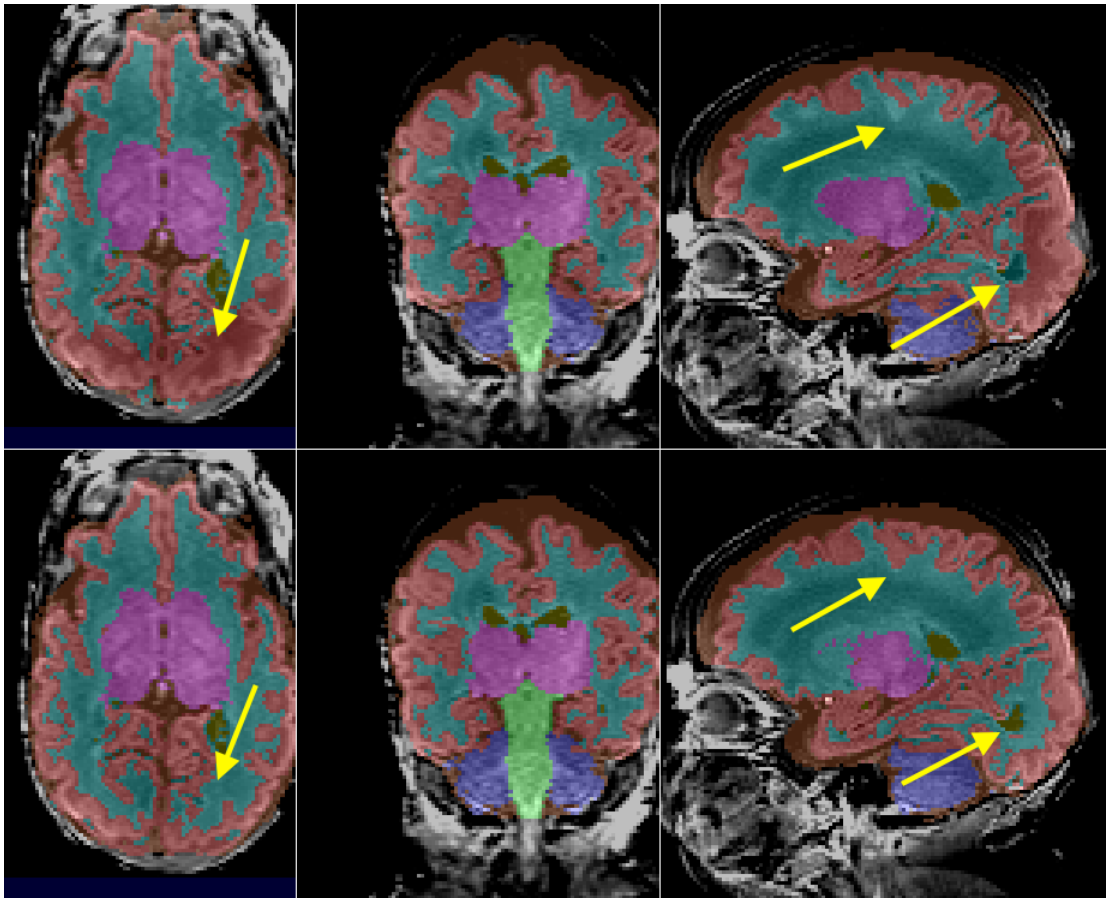

Figure 7.8: Comparison between tissue segmentations using PBAEM (top row) and hybrid pipeline (bottom row) from PENUT [115] study [150]. Yellow arrows: areas mislabeled by PBAEM while correctly labeled by the proposed hybrid pipeline.

7.5 Clinical Study: Effect of Morphine Exposure on Early Brain Development [Co-investigator: Jill Zwicker, PhD]

This study is to examine the relationship between morphine exposure and cerebellar growth from early postnatal period to term-equivalent age in premature neonates[154]. Given the role of the cerebellum in motor and cognitive functions[75], we also hypothesized that morphine exposure could cause motor and cognitive deficits at a later age. Note that this study was primarily carried out by our collaborators, and my contribution was to obtain the auto-

mated tissue labeling including cerebellum and to extract the volumetric measurements for further statistical analysis.

The study sample was the same as the previous subsection. However, more subjects were excluded when clinical evidence of major malformation (e.g. congenital syndrome, large parenchymal hemorrhagic infarction) was shown. A group of 136 very preterm neonates (birth age: $24 \sim 32$ GW) with detailed morphine exposure recorded were included in the study.

Generalized least squares models were used to examine the relationship between log-transformed morphine exposure and cerebellar volume, co-varying for age at birth, age at scan, total brain volume, early brain severity, number of invasive procedures and additional confounders. Linear regression models were adopted to examine the relationship of morphine exposure and age-adjusted brain volumes to motor and cognitive outcomes.

Results showed a significant association between morphine exposure and smaller cerebellar volume ($p < 0.001$), which persisted after adjusting for confounders ($p = 0.03$). Neonates, regardless of age at birth, show smaller cerebellar volumes with increased morphine exposure (Figure 7.9). However, no significant correlation was found for cerebral volumes. As far as neuro-cognitive functions, it was found that morphine exposure was significantly associated with poorer total motor scores at 18 months ($p < 0.001$), with a less robust relationship with cognitive scales ($p = 0.006$). The relationship of morphine with motor and cognitive outcomes were attenuated after adjusting for cerebellar and cerebral volumes, suggesting that the relationship between higher morphine exposure with poorer neurodevelopmental outcomes is mediated, at least in part, by slower brain growth.

7.6 Clinical Study: Effect of High-dose Epo on Preterm Neonatal Neurodevelopment [Co-investigator: Juul Sandra, PhD; Debosmita Biswas, MS]

The Preterm Epo Neuroprotection (PENUT) Trial is a NINDS funded trial to asses whether early high dose Epo will improve survival without neuro-developmental impairment in infants born between 24 to 28 GW, i.e. the extremely premature neonates [115]. PENUT is

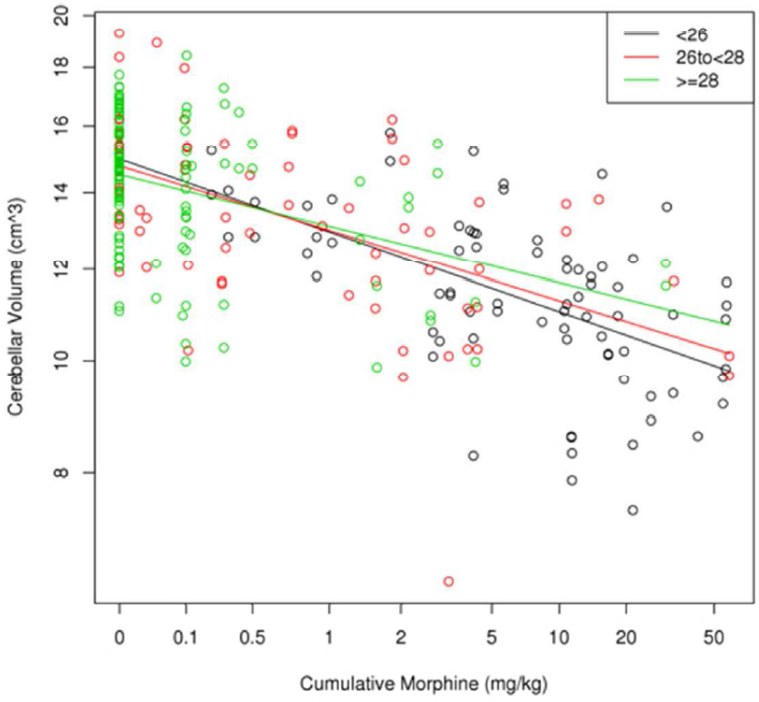

Figure 7.9: Cerebellar volumes by morphine exposure, stratified by age at birth. Figure reused by author from [154].

Site	#Scans Acquired	#Scans Processed	#Scans Used in Analysis
Orlando (FHO)	55	22	20
Texas (MCH)	48	25	10
Gainsville (UFL)	16	15	13
Minnesota (UMN)	16	11	3
Seattle (UW)	31	26	23
Baltimore (JHU)	7	6	2
NewYork (MFC)	12	6	1

Table 7.1: Number of scans acquired, processed and used in the analysis from each site in the PENUT study.

a multi-site study and the dataset contains MRI scans from 7 sites across the US. All brain MRIs were acquired at a term-equivalent age of 36 GW.

To automatically process the whole dataset, we further adapted the hybrid method by Zach et al [150] to replace PBAEM with the sparse dictionary learning methodology to further improve the accuracy of the estimated tissue probability for the PENUT study. The sparse dictionary learning approach was preferred over PBAEM because of two reasons. 1) PENUT scans were all acquired at the same age of 36 GW, therefore no spatio-temporal model was needed. Using individual MRIs as templates instead of an age-match average provides more detailed and less blurry dictionary patches. 2) Due to the multi-site nature of the PENUT study, the image quality varies across sites. Therefore, selecting template images from several sites instead of one single site to construct the dictionary can improve the chance of finding a correct match from the sparse patch-based search.

Due to the nature of clinical scans, some scans fail to meet the basic image quality requirement to be processed. These scans mainly suffer from severe motion artifacts or extremely undistinguishable contrast, even to the human eye. These scans were excluded from

the study. After completing the automatic segmentation pipeline, image quality analysis and automatic segmentation quality analysis of all processed scans, those that did not meet the image quality or segmentation accuracy were excluded from the analysis as well. The detailed information of scan counts can be found in Table 7.1. Volumetric measurements were extracted from automatic segmentation. Given the drug placebo group assignment, a two-tailed non-paired student t-test was conducted to analyze the difference between groups.

We observed a statistically significant difference between the two groups (drug and placebo) in WM-GM volume ratio, WM-(GM+DGM) volume ratio, GM-total volume ratio and (GM+DGM)-total volume ratio. Thus, we preliminarily concluded that the GM/WM volume ratio could be different depending on exposure to high-dose Epo. With that being said, the PENUT study is far from completed. Due to the nature of its double blinded clinical trial, group assignment of drug vs placebo is unkown at the current time of writing this thesis, and will remain unknown until all data is acquired. Future directions include deeper investigation of the excluded scans to avoid bias in the exclusion process, adoption of other morphological measurements such as cortical surface area.

Chapter 8

CONCLUSION AND DISCUSSION

This chapter summarizes the contribution and impact of the automatic MRI processing methodology developed in this dissertation and the results from the experiments. The overall outcomes of the studies are described, the distinguished contributions compared to previous literature are summarized, the limitations are discussed and future directions are motivated.

8.1 Outcomes and Contributions

This work developed an automatic MRI image segmentation methodology to achieve automated labeling of the normal and abnormal tissue structures in the premature neonatal brain by incorporating a patch-based method.

Methodology-wise, the patch-based techniques developed in this work can motivate more application of such techniques in image processing and analysis algorithms. Essentially, the main purpose of patch-based techniques is to correct the mismatching from one-to-one voxel location correspondence. The mismatching can happen in highly variable structures. Such structures may include both normal (e.g. cortical gyrifications) and abnormal (e.g. VM, focal lesions such as IVH etc.) structures. When we know that it is very likely to find a true match within a voxel's immediate neighborhood, we can search more effectively and efficiently by avoiding unnecessary computation using an estimate of the range inside which the potential match is in. The variability-constrained search developed in PBAEM-VCS is a novel approach to locally set the search range for any patch-based technique by learning the variability at each location. In Chapter 4, this estimated range is learnt from the variability of each location at any specific age and proved to improve the effectiveness and efficiency of the patch-based local search. Nonetheless, this locally set search range can

be further adapted based on the sources of the mismatching. Images that have focal lesions with unique intensity profiles, no matter which type of lesions, can be automatic segmented using the dictionary learning method from Chapter 6. The only requirement is to obtain manually delineated dictionary examples that can fully represent the lesion intensity profile. Therefore, this approach can be further adapted to automatically label other MRIs with lesions.

In terms of applications of our method, these automated tissue segmentation results on premature neonatal scans covering the full range of clinical ages provide a basis for a variety of morphometric analyses of brain growth. In particular, they enable the study of local anatomical variations using deformation based or surface based morphometry [102] that have so far not been able to be applied to such large scale studies of the brain after premature birth. In particular such techniques for local anatomical mapping in premature neonates may reveal markers that are specific to given functional deficits later in life, which can help direct early cognitive interventions soon after birth. A clinical application can also be explored based on automatic labeling of focal lesions present in a MRI scan. The automatic labeling can be used for visual hints for radiologist while providing rapid quantitative morphological estimates. Adopted in large scale studies, such as the multi-site large-scale study of post-hemorrhagic hydrocephalus, this method allows consistent lesion segmentation in a timely manner, enabling a study that would not have been possible before with only a manual labeling approach.

In summary, multiple automatic segmentation algorithms and pipelines were developed and implemented, including PBAEM-WF and PBAEM-VCS for normal tissue segmentation and a discriminative dictionary learning technique for focal lesion labeling. The developed algorithms and pipelines were applied to multiple original research studies, and could be further adapted and applied to more future studies as well as utilized in clinical settings.

8.2 Limitations and Future Work

The work presented in Chapter 3 and 4 focuses on the automatic image segmentation of normal tissue structures in healthy premature neonatal brain MRI. Although improved segmentation accuracy has been achieved and presented, there are some limitations in the developed PBAEM methodology, which motivate more future work. I discuss the limitations and propose the future steps as follows.

- **Utilization of other age-related information.** One of the limitations of our method is that we have not used all the age-related information that is available. This could be addressed by incorporating additional information into the atlas model. In the current work, the birth age of the atlas subjects ranges from 24 to 32 GW. Previous studies have shown that neonates born very pretermly (VPT) (less than 30 GW) [146, 90] and with low body weights (LBW) [90] may suffer from more severe neuro-developmental abnormalities compared to infants born less prematurely. Thus, it would be interesting to investigate if incorporating such information into the model could allow us to further adjust the local patch-based search for the very premature neonates.

- **Refinement of the spatio-temporal atlas.** The ages-at-scan of the atlas subjects are not evenly distributed within the whole age range from 27.29 to 46.43 GW, with particularly less density between 39 to 46 GW. Although the algorithm still produced anatomically valid automated tissue segmentation for that age range, we believe that the labeling accuracy in GM and WM could be greatly improved if more scans between 39 and 46 GW were included in the atlas. This is especially true because at this age, the cortical GM becomes more folded and thus makes the automated segmentation at the GM-WM boundary very challenging. However, we are currently limited in terms of good quality image with manual tracing.

- **Incorporation of tissue estimates from other sources.** Future work would also

include improving the methods of combining tissue probabilities from different sources. The proposed approach has introduced priors and constraints from sources other than the atlas itself into the conventional atlas-based EM automatic segmentation framework. A potential direction would be to incorporate priors acquired from other geometric models, such as the laminar structure model of the developing human brain [52] for further improvement of the segmentation accuracy at earlier gestational ages and in fetuses.

- **More detailed characterization of errors in PBAEM-VCS.** Future steps to continue this work may include more detailed characterizations of other types of errors that may affect the voxel location correspondence in the processing pipeline. For example, in registration between the age-specific MRI and the new subject MRI, errors could be introduced. The locally defined search range could be more accurate once these errors are taken into consideration.

- **Adaptation to fetal brain MRI analysis.** Another future direction is to adapt the methodology to fetal brain MRI analysis. This provides a valuable tool to compare in-utero and ex-utero brain development by comparing fetal and premature neonatal brains at the same gestational age.

Chapter 5 presented an application of the PBAEM methodology on a multi-site study. Although consistent performance across different sites, scanners and image protocols was proved, there are multiple future steps can be taken to extend the scale of the current study. Future steps to extend the current work includes investigation of compatibility or reproducibility of the multi-center data, optimization of the processing pipeline using multi-center data and extended validation on both normal and abnormal scans. In particular, in my opinion, optimizing the pipeline using the entire multi-center dataset will result in a great boost of analysis quality in such studies. Currently, the optimal parameters were chosen based on the performance from only one subset of the multi-center data, leading to

a slightly unbalanced performance.

Chapters 6 presented a patch-based dictionary learning approach for automatic image segmentation of normal tissue structure and focal lesions (with IVH as an example). Although we achieved, for the first time, whole brain tissue and lesion segmentation of premature neonatal brain MRI with high accuracy, there are some limitations and future work worth discussing as follows.

- **Construction of the discriminative dictionary.** In this work, we adopted a measure, correlation, to eliminate duplicate dictionary elements and construct the non-spatial dictionary, which is consistent with later sparse coding. The threshold is crucial for the descriptive power of the dictionary. The threshold might also be different for different dataset. In this work, the threshold was chosen using a leave-one-out cross-validation. One of the future directions is to better construct a discriminative dictionary. Tong et al. [132] proposed a method to construct a discriminative dictionary using the k-singular vector decomposition (K-SVD) algorithm. K-SVD was proposed by Aharon et al. [2] to design overcomplete dictionaries for sparse representation. This method uses an iterative approach to determine the optimal discriminative dictionary that best represents the data. However, due to the iterative nature of the method, it might be computationally intensive. Therefore, this requires further investigation.

- **Adaptation to other age ranges and extensive validation experiments.** Due to limited high-quality MRI scans with IVH and manual delineation, we were only able to conduct experiments on 4 subjects that fell into two narrow age ranges. Theoretically, the proposed method should work for any age range if and only if there is age-matched training examples in the dictionary. The matched age is particularly important in premature neonatal brain studies because of the changing contrast with age. Therefore, more extensive validation experiments, especially in other age brackets, should conducted when appropriate manually delineated reference data becomes

available.

- **Adaptation to other local lesions.** The method proposed theoretically should work for all types of highly variable focal lesions as long as there are a small set of annotated lesion images by trained professionals, for example, Periventricular leukomalacia (PVL). The method should also be able to correctly label extremely enlarged ventricles caused by bleeding in post-hemorrhagic hydrocephalus. Moreover, it would be interesting to explore the ability of the methodology to distinguish IVH from intraparenchymal blood clot. However, we currently lack good quality professionally traced data for these cases. Therefore, future work is motivated in these cases when manually delineated reference data becomes available.

Looking at the whole quantitative analysis pipeline, one major missing component is the automated detection of lesions. If we are given a healthy premature neonatal brain MRI, we can obtain its accurate tissue labeling by using methods from Chapter 3 and 4; if we are given a MRI scan with IVH, then we are confident that the combined dictionary search of both spatial and non-spatial dictionary can generate an accurate labeling of both normal tissue structures and IVH. However, there is a lack of automatic sorting of the randomly given MRI scans into the two categories: normal vs abnormal. Of course, we can always use a combined dictionary for both normal and abnormal scans which most likely will generate an accurate segmentation, but it has the potential to be a waste of computation. Therefore, future steps can be taken to design an algorithm to first automatically detect if there are any known lesions present in the scan and then assign the appropriate segmentation method to process the image.

Deep learning techniques have also attracted a lot of research interest in general imaging and computer vision during the time of this work. Deep learning refers to learning the computational models, often via neural networks, which consist of multiple processing layers to learn representations of data using multiple abstraction levels [76]. Moeskops et al [91] reported the satisfactory performance of convolutional neural networks (CNNs) on 5 different

datasets of subjects from pre-term infants to older adults. Zhang et al [151] presented a method to use deep conventional neural networks to segment multi-modality infant brain images at the iso-intense stage and showed significant improvement on this challenging task. In many more examples, deep learning has shown its power in achieving results in traditional biomedical image analysis tasks. Thus, future directions can include further developing new deep learning methodologies to achieve a more accurate and faster automatic quantitative analysis pipeline for premature neonatal brain studies.

8.3 Conclusions

An automatic quantitative analysis pipeline was developed for premature neonatal brain MRI images, with the focus on developing new methodology to achieve more accurate automatic tissue segmentation. The first two methods focused on brain scans without severe pathology but included large inter-subject variation, and demonstrated higher segmentation accuracy and consistent performance on multi-center cross-protocol scans. The third method focused on incorporating automatic labeling of focal lesions such as IVH into full-brain segmentation. This method achieved for the first time accurate full-brain segmentation with IVH in premature neonatal brain MRI. The field of fetal and neonatal brain imaging has advanced significantly over the past few decades, making more quantitative studies of early brain development have been made possible. In particular, larger datasets are being acquired, and more manually-traced datasets has become available. This work presented an automatic quantitative analysis pipeline that will enable such studies to be conducted in a consistent and timely fashion, contribute to the understanding of human brain development and enable early detection of pathologies.

BIBLIOGRAPHY

[1] Faiza Admiraal-Behloul, DMJ Van Den Heuvel, Hans Olofsen, Matthias JP van Osch, Jeroen van der Grond, MA Van Buchem, and JHC Reiber. Fully automatic segmentation of white matter hyperintensities in mr images of the elderly. *Neuroimage*, 28(3):607–617, 2005.

[2] Michal Aharon, Michael Elad, and Alfred Bruckstein. *rmk*-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[3] Craig A Albers and Adam J Grieve. Review of bayley scales of infant and toddler development–. 2007.

[4] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal, and Daniel Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[5] Petronella Anbeek, Ivana Išgum, Britt JM van Kooij, Christian P Mol, Karina J Kersbergen, Floris Groenendaal, Max A Viergever, Linda S de Vries, and Manon JNL Benders. Automatic segmentation of eight tissue classes in neonatal brain mri. *PLoS One*, 8(12):e81895, 2013.

[6] Petronella Anbeek, Koen L Vincken, Matthias JP van Osch, Robertus HC Bisschops, and Jeroen van der Grond. Automatic segmentation of different-sized white matter lesions by voxel probability estimation. *Medical image analysis*, 8(3):205–215, 2004.

[7] Petronella Anbeek, Koen L Vincken, Matthias JP van Osch, Robertus HC Bisschops, and Jeroen van der Grond. Probabilistic segmentation of white matter lesions in mr imaging. *NeuroImage*, 21(3):1037–1044, 2004.

[8] Xabier Artaechevarria, Arrate Munoz-Barrutia, and Carlos Ortiz-de Solórzano. Combination strategies in multi-atlas image segmentation: Application to brain mr data. *IEEE transactions on medical imaging*, 28(8):1266–1277, 2009.

[9] John Ashburner and Karl J. Friston. Unified segmentation. *NeuroImage*, 26(3):839 – 851, 2005.

[10] John Ashburner and Karl J Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.

[11] Andrew J Asman and Bennett A Landman. Non-local statistical label fusion for multi-atlas segmentation. *Medical image analysis*, 17(2):194–208, 2013.

[12] Wenjia Bai, Wenzhe Shi, Declan P O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, and Daniel Rueckert. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac mr images. *Medical Imaging, IEEE Transactions on*, 32(7):1302–1315, 2013.

[13] Malcolm R Battin, Elia F Maalouf, Serena J Counsell, Amy H Herlihy, Mary A Rutherford, Denis Azzopardi, and A David Edwards. Magnetic resonance imaging of the brain in very preterm infants: visualization of the germinal matrix, early myelination, and cortical folding. *Pediatrics*, 101(6):957–962, 1998.

[14] Elizabeth Berry and Andrew J Bulpitt. *Fundamentals of MRI: an interactive learning approach*. CRC Press, 2008.

[15] Andrew CG Breeze, Peta Alexander, Edile M Murdoch, Hannah H Missfelder-Lobos, Gerald A Hackett, and Christoph C Lees. Obstetric and neonatal outcomes in severe fetal ventriculomegaly. *Prenatal diagnosis*, 27(2):124–129, 2007.

[16] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[17] Graeme M Bydder, RE Steiner, IR Young, AS Hall, DJ Thomas, J Marshall, CA Pallis, and NJ Legg. Clinical NMR imaging of the brain: 140 cases. *American Journal of Roentgenology*, 139(2):215–236, 1982.

[18] Mariano Cabezas, Arnau Oliver, Xavier Llad, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158 – e177, 2011.

[19] M Jorge Cardoso, Andrew Melbourne, Giles S Kendall, Marc Modat, Nicola J Robertson, Neil Marlow, and Sebastien Ourselin. AdaPT: an adaptive preterm segmentation algorithm for neonatal brain mri. *NeuroImage*, 65:97–108, 2013.

[20] Jimmy D Cardoza, Ruth B Goldstein, and Roy A Filly. Exclusion of fetal ventriculomegaly with a single measurement: the width of the lateral ventricular atrium. *Radiology*, 169(3):711–714, 1988.

[21] Alan Carrington and Andrew D McLachlan. Introduction to magnetic resonance: with applications to chemistry and chemical physics. 1967.

[22] Vann Chau, Kenneth J Poskitt, Deborah E McFadden, Tim Bowen-Roberts, Anne Synnes, Rollin Brant, Michael A Sargent, Wendy Soulikias, and Steven P Miller. Effect of chorioamnionitis on brain development and injury in premature newborns. *Annals of neurology*, 66(2):155–164, 2009.

[23] Vann Chau, Anne Synnes, Ruth E Grunau, Kenneth J Poskitt, Rollin Brant, and Steven P Miller. Abnormal brain maturation in preterm neonates associated with adverse developmental outcomes. *Neurology*, 81(24):2082–2089, 2013.

[24] Irene Cheng, Steven P Miller, Emma G Duerden, Kaiyu Sun, Vann Chau, Elysia Adams, Kenneth J Poskitt, Helen M Branson, and Anup Basu. Stochastic process for white matter injury detection in preterm neonates. *NeuroImage: Clinical*, 7:622–630, 2015.

[25] Je G Chi, Elizabeth C Dooling, and Floyd H Gilles. Gyral development of the human brain. *Annals of neurology*, 1(1):86–93, 1977.

[26] Myong-sun Choe, Silvia Ortiz-Mantilla, Nikos Makris, Matt Gregas, Janine Bacic, Daniel Haehn, David Kennedy, Rudolph Pienaar, Verne S. Caviness, April A. Benasich, and P. Ellen Grant. Regional infant brain development: An mri-based morphometric analysis in 3 to 13 month olds. *Cerebral Cortex*, 23(9):2100–2117, 2013.

[27] Kristi A Clark, Roger P Woods, David A Rottenberg, Arthur W Toga, and John C Mazziotta. Impact of acquisition protocols and processing streams on tissue segmentation of t1 weighted mr images. *NeuroImage*, 29(1):185–202, 2006.

[28] Mayfield Clinic. Mayfield brain & spine, https://www.mayfieldclinic.com/pe-anatbrain.htm, 2016.

[29] Chris A Cocosco, Alex P Zijdenbos, and Alan C Evans. A fully automatic and robust brain mri tissue classification method. *Medical Image Analysis*, 7(4):513–527, 2003.

[30] Pierrick Coupé, José V Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and D Louis Collins. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954, 2011.

[31] Jorrit F de Kieviet, Jan P Piek, Cornelieke S Aarnoudse-Moens, and Jaap Oosterlaan. Motor development in very preterm and very low-birth-weight children from birth to adolescence: a meta-analysis. *Jama*, 302(20):2235–2242, 2009.

[32] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

[33] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[34] John Dobbing and Jean Sands. Quantitative growth and development of human brain. *Archives of disease in childhood*, 48(10):757–767, 1973.

[35] Amgad Droby, Carsten Lukas, Anne Schänzer, Isabella Spiwoks-Becker, Antonio Giorgio, Ralf Gold, Nicola De Stefano, Harald Kugel, Michael Deppe, Heinz Wiendl, et al. A human post-mortem brain model for the standardization of multi-centre mri studies. *NeuroImage*, 2015.

[36] Leigh E Dyet, Nigel Kennea, Serena J Counsell, Elia F Maalouf, Morenike Ajayi-Obe, Philip J Duggan, Michael Harrison, Joanna M Allsop, Joseph Hajnal, Amy H Herlihy, et al. Natural history of brain lesions in extremely preterm infants studied with serial magnetic resonance imaging from birth and neurodevelopmental assessment. *Pediatrics*, 118(2):536–548, 2006.

[37] Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[38] Simon F. Eskildsen, Pierrick Coupe, Vladimir Fonov, Jose V. Manjon, Kelvin K. Leung, Nicolas Guizard, Shafik N. Wassef, Lasse Riis Ostergaard, and D. Louis Collins. BEaST: Brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362 – 2373, 2012.

[39] M Filippi, JH Van Waesberghe, MA Horsfield, S Bressi, C Gasperini, TA Yousry, ML Gawne-Cain, SP Morrissey, MA Rocca, F Barkhof, et al. Interscanner variation in brain mri lesion load measurements in ms: implications for clinical trials. *Neurology*, 49(2):371–377, 1997.

[40] Lynn M Fletcher, John B Barsotti, and Joseph P Hornak. A multispectral analysis of brain tissues. *Magnetic Resonance in Medicine*, 29(5):623–630, 1993.

[41] Margaret A Foster. Magnetic resonance in medicine and biology. 1984.

[42] Yasuyuki Futagi, Yasuhisa Toribe, Kana Ogawa, and Yasuhiro Suzuki. Neurodevelopmental outcome in children with intraventricular hemorrhage. *Pediatric neurology*, 34(3):219–224, 2006.

[43] P Gaglioti, D Danelon, S Bontempo, M Mombro, S Cardaropoli, and T Todros. Fetal cerebral ventriculomegaly: outcome in 176 cases. *Ultrasound in obstetrics & gynecology*, 25(4):372–377, 2005.

[44] Jay N Giedd. Structural magnetic resonance imaging of the adolescent brain. *Annals of the New York Academy of Sciences*, 1021(1):77–85, 2004.

[45] Jay N Giedd, Jonathan Blumenthal, Neal O Jeffries, F Xavier Castellanos, Hong Liu, Alex Zijdenbos, Tomáš Paus, Alan C Evans, and Judith L Rapoport. Brain development during childhood and adolescence: a longitudinal mri study. *Nature neuroscience*, 2(10):861–863, 1999.

[46] Orit A. Glenn. Normal development of the fetal brain by {MRI}. *Seminars in Perinatology*, 33(4):208 – 219, 2009. Advances in Fetal Neuroimaging.

[47] Catriona D Good, Ingrid S Johnsrude, John Ashburner, Richard NA Henson, KJ Friston, and Richard SJ Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. In *Biomedical Imaging, 2002. 5th IEEE EMBS International Summer School on*, pages 16–pp. IEEE, 2002.

[48] Ioannis S Gousias, Alexander Hammers, Serena J Counsell, Latha Srinivasan, Mary A Rutherford, Rolf A Heckemann, Jo V Hajnal, Daniel Rueckert, and A David Edwards. Magnetic resonance imaging of the newborn brain: automatic segmentation of brain images into 50 anatomical regions. *PloS one*, 8(4):e59990, 2013.

[49] Hayit Greenspan, Amit Ruf, and Jacob Goldberger. Constrained gaussian mixture model framework for automatic segmentation of mr brain images. *Medical Imaging, IEEE Transactions on*, 25(9):1233–1245, 2006.

[50] Hákon Gudbjartsson and Samuel Patz. The rician distribution of noisy mri data. *Magnetic resonance in medicine*, 34(6):910–914, 1995.

[51] Debra A Gusnard and Marcus E Raichle. Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews Neuroscience*, 2(10):685–694, 2001.

[52] Piotr A Habas, Kio Kim, Dharshan Chandramohan, Francois Rousseau, Orit A Glenn, and Colin Studholme. Statistical model of laminar structure for atlas-based segmentation of the fetal brain from in utero MR images. In P. W. Pluim and B. M. Dawant, editors, *Medical Imaging: Image Processing*, volume 7259 of *Proc. SPIE*, pages 725917–8, 2009.

[53] Piotr A Habas, Kio Kim, James M Corbett-Detig, Francois Rousseau, Orit A Glenn, A James Barkovich, and Colin Studholme. A spatiotemporal atlas of mr intensity, tissue probability and shape of the fetal brain with application to segmentation. *Neuroimage*, 53(2):460–470, 2010.

[54] Piotr A Habas, Kio Kim, Francois Rousseau, Orit A Glenn, A James Barkovich, and Colin Studholme. Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses. *Human brain mapping*, 31(9):1348–1358, 2010.

[55] G Hambleton and JS Wigglesworth. Origin of intraventricular haemorrhage in the preterm infant. *Archives of Disease in Childhood*, 51(9):651–659, 1976.

[56] B.E. Hamilton, J.A. Martin, M.J.K Osterman, and S.C. Curtin. Births: Preliminary data for 2013. *National vital statistics reports*, 63(2):1 – 19, 2014.

[57] Xiao Han and Bruce Fischl. Atlas renormalization for improved brain mr image segmentation across scanner platforms. *Medical Imaging, IEEE Transactions on*, 26(4):479–486, 2007.

[58] Lisa Harrylock, Averi Kitsch, Mengyuan Liu, Steven Miller, Vann Chau, Ken Poskitt, and Colin Studholme. Brainstem volume after premature birth and developmenttal outcome at 18 months. Presented at Organization for Human Brain Mapping (OHBM) 2014 Annual Meeting, Hamburg, Germany, 2014.

[59] Rolf A. Heckemann, Joseph V. Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115 – 126, 2006.

[60] Alfred L Horowitz. *MRI physics for physicians*. Springer Science & Business Media, 2012.

[61] Petra S. Huppi, Simon Warfield, Ron Kikinis, Patrick D. Barnes, Gary P. Zientara, Ferenc A. Jolesz, Miles K. Tsuji, and Joseph J. Volpe. Quantitative magnetic resonance imaging of brain development in premature and mature newborns. *Annals of Neurology*, 43(2):224–235, 1998.

[62] Petra S Hüppi, Simon Warfield, Ron Kikinis, Patrick D Barnes, Gary P Zientara, Ferenc A Jolesz, Miles K Tsuji, and Joseph J Volpe. Quantitative magnetic resonance imaging of brain development in premature and mature newborns. *Annals of neurology*, 43(2):224–235, 1998.

[63] Daniel P. Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.

[64] Terrie E Inder, Scott J Wells, Nina B Mogridge, Carole Spencer, and Joseph J Volpe. Defining the nature of the cerebral abnormalities in the premature infant: a qualitative magnetic resonance imaging study. *The Journal of pediatrics*, 143(2):171–179, 2003.

[65] Ivana Išgum, Manon JNL Benders, Brian Avants, M Jorge Cardoso, Serena J Counsell, Elda Fischi Gomez, Laura Gui, Petra S Hűppi, Karina J Kersbergen, Antonios Makropoulos, et al. Evaluation of automatic neonatal brain segmentation algorithms: the neobrains12 challenge. *Medical image analysis*, 20(1):135–151, 2015.

[66] MA Johnson, JM Pennock, GM Bydder, RE Steiner, DJ Thomas, R Hayward, DR Bryant, JA Payne, MI Levene, and A Whitelaw. Clinical nmr imaging of the brain in children: normal and neurologic disease. *American Journal of Roentgenology*, 141(5):1005–1018, 1983.

[67] Jorge Jovicich, Moira Marizzoni, Roser Sala-Llonch, Beatriz Bosch, David Bartrés-Faz, Jennifer Arnold, Jens Benninghoff, Jens Wiltfang, Luca Roccatagliata, Flavio Nobili, et al. Brain morphometry reproducibility in multi-center 3t mri studies: A comparison of cross-sectional and longitudinal segmentations. *NeuroImage*, 83:472–484, 2013.

[68] Olga Kapellou, Serena J. Counsell, Nigel Kennea, Leigh Dyet, Nadeem Saeed, Jaroslav Stark, Elia Maalouf, Philip Duggan, Morenike Ajayi-Obe, Jo Hajnal, Joanna M Allsop, James Boardman, Mary A. Rutherford, Frances Cowan, and A. David Edwards. Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. *PLoS Med*, 3(8):e265, 2006.

[69] Vladimir Katkovnik, Alessandro Foi, Karen Egiazarian, and Jaakko Astola. From local kernel to nonlocal multiple-model image denoising. *International journal of computer vision*, 86(1):1–32, 2010.

[70] Edmond N Kelly, Victoria M Allen, Gareth Seaward, Rory Windrim, and Greg Ryan. Mild ventriculomegaly in the fetus, natural history, associated findings and outcome of isolated mild ventriculomegaly: a literature review. *Prenatal diagnosis*, 21(8):697–700, 2001.

[71] O Khwaja and Joseph J Volpe. Pathogenesis of cerebral white matter injury of prematurity. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 93(2):F153–F161, 2008.

[72] Hiroyuki Kidokoro, Peter J Anderson, Lex W Doyle, Lianne J Woodward, Jeffrey J Neil, and Terrie E Inder. Brain injury and altered brain growth in preterm infants: predictors and prognosis. *Pediatrics*, pages peds–2013, 2014.

[73] Hiroyuki Kidokoro, Peter J. Anderson, Lex W. Doyle, Lianne J. Woodward, Jeffrey J. Neil, and Terrie E. Inder. Brain injury and altered brain growth in preterm infants: Predictors and prognosis. *Pediatrics*, 134(2):e444–e453, 2014.

[74] Averi Kitsch, Mengyuan Liu, Lisa Harrylock, Steven Miller, Vann Chau, Ken Poskitt, and Colin Studholme. Regional neonatal brain tissue volumes after premature birth and neurocognitive delay at 18 months. Presented at Organization for Human Brain Mapping (OHBM) 2014 Annual Meeting, Hamburg, Germany, 2014.

[75] Leonard F Koziol, Deborah Budding, Nancy Andreasen, Stefano DArrigo, Sara Bulgheroni, Hiroshi Imamizu, Masao Ito, Mario Manto, Cherie Marvel, Krystal Parker, et al. Consensus paper: the cerebellum's role in movement and cognition. *The Cerebellum*, 13(1):151–177, 2014.

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

[77] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging*, 18(10):897–908, 1999.

[78] Rhoshel K Lenroot and Jay N Giedd. Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neuroscience & Biobehavioral Reviews*, 30(6):718–729, 2006.

[79] Mengyuan Liu, Lisa Harrylock, Averi Kitsch, Dennis Shaw, Steven Miller, Vann Chau, Kenn Poskitt, Francois Rousseau, and Colin Studholme. Automated outlier-based intraventricular hemorrhage segmentation (obihs) in premature neonatal {MRI}. In *Proc. of Organization of Human Brain Mapping Annual Meeting 2014*, 2014.

[80] Mengyuan Liu, Averi Kitsch, Steven Miller, Vann Chau, Kenneth Poskitt, Francois Rousseau, Dennis Shaw, and Colin Studholme. Patch-based augmentation of expectation–maximization for brain MRI tissue segmentation at arbitrary age after premature birth. *NeuroImage*, 127:387–408, 2016.

[81] Mengyuan Liu, Julia Scott, Steven Miller, Vann Chau, Ken Poskitt, and Colin Studholme. An evaluation of age-specific atlas-based MRI brain segmentation in premature neonates. Proc. BMES, 2013.

[82] Mengyuan Liu, Sharmishtaa Seshamani, Lisa Harrylock, Averi Kitsch, Steven Miller, Van Chau, Kenneth Poskitt, Francois Rousseau, and Colin Studholme. Spatially adapted augmentation of age-specific atlas-based segmentation using patch-based priors. In *SPIE Medical Imaging*, pages 90341H–90341H. International Society for Optics and Photonics, 2014.

[83] Antonios Makropoulos, Ioannis S Gousias, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V Hajnal, A David Edwards, Serena J Counsell, and Daniel Rueckert. Automatic whole brain MRI segmentation of the developing neonatal brain. 2014.

[84] Elaine Nicpon Marieb and Katja Hoehn. *Human anatomy & physiology*. Pearson Education, 2007.

[85] Neil Marlow. Neurocognitive outcome after very preterm birth. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 89(3):F224–F228, 2004.

[86] E Martin, R Kikinis, M Zuerrer, Ch Boesch, J Briner, G Kewitz, and P Kaelin. Developmental stages of human brain: an mr study. *Journal of computer assisted tomography*, 12(6):917–922, 1988.

[87] Junko Matsuzawa, Mi Matsui, Tohru Konishi, Kyo Noguchi, Ruben C. Gur, Warren Bilker, and Toshio Miyawaki. Age-related volumetric changes of brain gray and white matter in healthy infants and children. *Cerebral Cortex*, 11(4):335–342, 2001.

[88] CB McArdle, CJ Richardson, DA Nicholas, M Mirfakhraee, CK Hayden, and EG Amparo. Developmental features of the neonatal brain: Mr imaging. part i. gray-white matter differentiation and myelination. *Radiology*, 162(1):223–229, 1987.

[89] Steven P Miller, Donna M Ferriero, Carol Leonard, Robert Piecuch, David V Glidden, J Colin Partridge, Marta Perez, Pratik Mukherjee, Daniel B Vigneron, and A James Barkovich. Early brain injury in premature newborns detected with magnetic resonance imaging is associated with adverse early neurodevelopmental outcome. *The Journal of pediatrics*, 147(5):609–616, 2005.

[90] Majid Mirmiran, Patrick D. Barnes, Kathy Keller, Janet C. Constantinou, Barry E. Fleisher, Susan R. Hintz, and Ronald L. Ariagno. Neonatal brain magnetic resonance imaging before discharge is better than serial cranial ultrasound in predicting cerebral palsy in very low birth weight preterm infants. *Pediatrics*, 114(4):992–998, 2004.

[91] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S de Vries, Manon JNL Benders, and Ivana Išgum. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5):1252–1261, 2016.

[92] Parashkev Nachev, Elizabeth Coulthard, H Rolf Jäger, Christopher Kennard, and Masud Husain. Enantiomorphic normalization of focally lesioned brains. *NeuroImage*, 39(3):1215–1226, 2008.

[93] Anil J Patwardhan, Stephan Eliez, Ilana S Warsofsky, Gary H Glover, Chris D White, Jay N Giedd, Bradley S Peterson, Donald C Rojas, and Allan L Reiss. Effects of image orientation on the comparability of pediatric brain volumes using three-dimensional mr data. *Journal of computer assisted tomography*, 25(3):452–457, 2001.

[94] Bradley S Peterson, Betty Vohr, Lawrence H Staib, Christopher J Cannistraci, Aaron Dolberg, Karen C Schneider, Karol H Katz, Michael Westerveld, Sara Sparrow, Adam W Anderson, et al. Regional brain volume abnormalities and long-term cognitive outcome in preterm infants. *Jama*, 284(15):1939–1947, 2000.

[95] Dzung L Pham and Jerry L Prince. An adaptive fuzzy segmentation algorithm for three-dimensional magnetic resonance images. In *Information Processing in Medical Imaging*, pages 140–153. Springer, 1999.

[96] Kilian M Pohl, John Fisher, Sylvain Bouix, Martha Shenton, Robert W McCarley, W Eric L Grimson, Ron Kikinis, and William M Wells. Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11(5):465–477, 2007.

[97] Marcel Prastawa, Elizabeth Bullitt, Sean Ho, and Guido Gerig. A brain tumor segmentation framework based on outlier detection. *Medical image analysis*, 8(3):275–283, 2004.

[98] Marcel Prastawa, Elizabeth Bullitt, Nathan Moon, Koen Van Leemput, and Guido Gerig. Automatic brain tumor segmentation by subject specific modification of atlas priors 1. *Academic radiology*, 10(12):1341–1348, 2003.

[99] Marcel Prastawa, John Gilmore, Weili Lin, and Guido Gerig. Automatic segmentation of neonatal brain mri. In Christian Barillot, DavidR. Haynor, and Pierre Hellier, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2004*, volume 3216 of *Lecture Notes in Computer Science*, pages 10–17. Springer Berlin Heidelberg, 2004.

[100] Marcel Prastawa, John H Gilmore, Weili Lin, and Guido Gerig. Automatic segmentation of mr images of the developing newborn brain. *Medical image analysis*, 9(5):457–466, 2005.

[101] Wu Qiu, Jing Yuan, Martin Rajchl, Jessica Kishimoto, Yimin Chen, Sandrine de Ribaupierre, Bernard Chiu, and Aaron Fenster. 3D MR ventricle segmentation in pre-term infants with post-hemorrhagic ventricle dilatation (phvd) using multi-phase geodesic level-sets. *NeuroImage*, 118:13–25, 2015.

[102] Vidya Rajagopalan, Julia A Scott, Mengyuan Liu, Kenneth Poskitt, Vann Chau, Steven Miller, and Colin Studholme. Complementary cortical gray and white matter developmental patterns in healthy, preterm neonates. *Human brain mapping*, 38(9):4322–4336, 2017.

[103] R Rathbone, SJ Counsell, O Kapellou, L Dyet, N Kennea, J Hajnal, JM Allsop, F Cowan, and AD Edwards. Perinatal cortical growth and childhood neurocognitive abilities. *Neurology*, 77(16):1510–1517, 2011.

[104] Hashemi Ray et al. Mri. *The Basics 2ED*, 2004.

[105] Art Riddle, Justin Dean, Joshua R Buser, Xi Gong, Jennifer Maire, Kevin Chen, Tahir Ahmad, Victor Cai, Thuan Nguyen, Christopher D Kroenke, et al. Histopathological correlates of magnetic resonance imaging–defined chronic perinatal white matter injury. *Annals of neurology*, 70(3):493–507, 2011.

[106] Nicole Riley, Mengyuan Liu, Raley Rewa, Steven Miller, Vann Chau, Ken Poskitt, and Colin Studholme. Occipital lobe subdivision volumes in relationship to neurological delays in premature neonates. Presented at Organization for Human Brain Mapping (OHBM) 2016 Annual Meeting, Geneva, Switzerland, 2016.

[107] Claudia E Rodriguez-Carranza, Pratik Mukherjee, Daniel Vigneron, J Barkovich, and Colin Studholme. A framework for in vivo quantification of regional brain folding in premature neonates. *Neuroimage*, 41(2):462–478, 2008.

[108] Mounira Rouaïnia, Mohamed Salah Medjram, and Noureddine Doghmane. Brain mri segmentation and lesions detection by em algorithm. *Proc. World Acad. Sci. Eng. Technol*, 17:301–304, 2006.

[109] Francois Rousseau, Piotr Habas, and Colin Studholme. A supervised patch-based approach for human brain labeling. *IEEE Trans. Med. Imaging*, 30:1852–1862, 2011.

[110] Snehashis Roy, Aaron Carass, Jerry L Prince, and Dzung L Pham. Subject specific sparse dictionary learning for atlas based brain MRI segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 248–255. Springer, 2014.

[111] Su Ruan, Stéphane Lebonvallet, Abderrahim Merabet, and J Constans. Tumor segmentation from a multispectral mri images by using support vector machine classification. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 1236–1239. IEEE, 2007.

[112] Mary A Rutherford. *MRI of the Neonatal Brain*. Gulf Professional Publishing, 2002.

[113] Mert R Sabuncu, BT Thomas Yeo, Koen Van Leemput, Bruce Fischl, and Polina Golland. A generative model for image segmentation based on label fusion. *Medical Imaging, IEEE Transactions on*, 29(10):1714–1729, 2010.

[114] LJ Salomon, JP Bernard, and Y Ville. Reference ranges for fetal ventricular width: a non-normal approach. *Ultrasound in obstetrics & gynecology*, 30(1):61–66, 2007.

[115] Patrick Heagerty Sandra Juul. Preterm epo neuroprotection (penut) trial, 2013.

[116] Ana Sanjuán, Cathy J Price, Laura Mancini, Goulven Josse, Alice Grogan, Adam K Yamamoto, Sharon Geva, Alex P Leff, Tarek A Yousry, and Mohamed L Seghier. Automated identification of brain tumors from single mr images based on segmentation with refined patient-specific priors. *Frontiers in neuroscience*, 7, 2013.

[117] Hugo G Schnack, Neeltje EM van Haren, Rachel M Brouwer, G Caroline M van Baal, Marco Picchioni, Matthias Weisbrod, Heinrich Sauer, Tyrone D Cannon, Matti Huttunen, Claude Lepage, et al. Mapping reliability in multicenter mri: Voxel-based morphometry and cortical thickness. *Human brain mapping*, 31(12):1967–1982, 2010.

[118] Hugo G Schnack, Neeltje EM van Haren, Hulshoff Pol, E Hilleke, Marco Picchioni, Matthias Weisbrod, Heinrich Sauer, Tyrone Cannon, Matti Huttunen, Robin Murray, et al. Reliability of brain volumes from multicenter mri acquisition: a calibration study. *Human brain mapping*, 22(4):312–320, 2004.

[119] Arthur Schweiger and Gunnar Jeschke. *Principles of pulse electron paramagnetic resonance*. Oxford University Press on Demand, 2001.

[120] Julia A. Scott, Piotr A. Habas, Kio Kim, Vidya Rajagopalan, Kia S. Hamzelou, James M. Corbett-Detig, A. James Barkovich, Orit A. Glenn, and Colin Studholme. Growth trajectories of the human fetal brain tissues estimated from 3D reconstructed in utero MRI. *International Journal of Developmental Neuroscience*, 29(5):529 – 536, 2011.

[121] Mohamed L Seghier, Anil Ramlackhansingh, Jenny Crinion, Alexander P Leff, and Cathy J Price. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage*, 41(4):1253–1266, 2008.

[122] Divyen K Shah, Celeste Guinane, Philipp August, Nicola C Austin, Lianne J Woodward, Deanne K Thompson, Simon K Warfield, Richard Clemett, and Terrie E Inder. Reduced occipital regional volumes at term predict impaired visual function in early childhood in very low birth weight infants. *Investigative ophthalmology & visual science*, 47(8):3366–3373, 2006.

[123] Navid Shiee, Pierre-Louis Bazin, Jennifer L Cuzzocreo, Ari Blitz, and Dzung L Pham. Segmentation of brain images using adaptive atlases with application to ventriculomegaly. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 1–12. Springer, 2011.

[124] Zhuang Song, Suyash P Awate, Daniel J Licht, and James C Gee. Clinical neonatal brain mri segmentation using adaptive nonparametric data models and intensity-based markov priors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 883–890. Springer, 2007.

[125] Elizabeth R Sowell, Doris A Trauner, Anthony Gamst, and Terry L Jernigan. Development of cortical and subcortical brain structures in childhood and adolescence: a structural mri study. *Developmental Medicine & Child Neurology*, 44(1):4–16, 2002.

[126] Cynthia M Stonnington, Geoffrey Tan, Stefan Klöppel, Carlton Chu, Bogdan Draganski, Clifford R Jack, Kewei Chen, John Ashburner, and Richard SJ Frackowiak. Interpreting scan data acquired from multiple scanners: a study with alzheimer's disease. *Neuroimage*, 39(3):1180–1185, 2008.

[127] Colin Studholme. Mapping fetal brain development in utero using magnetic resonance imaging: The big bang of brain mapping. *Annual Review of Biomedical Engineering*, 13:345–368, 2011. cited By (since 1996)19.

[128] Colin Studholme and Franois Rousseau. Quantifying and modelling tissue maturation in the living human fetal brain. *International Journal of Developmental Neuroscience*, 32(0):3 – 10, 2014. Special Issue: Imaging Developing Brain.

[129] Martin A Styner, H Cecil Charles, Jin Park, and Guido Gerig. Multisite validation of image analysis methods: assessing intra-and intersite variability. In *Medical Imaging 2002*, pages 278–286. International Society for Optics and Photonics, 2002.

[130] Emily WY Tam, Steven P Miller, Colin Studholme, Vann Chau, David Glidden, Kenneth J Poskitt, Donna M Ferriero, and A James Barkovich. Differential effects of intraventricular hemorrhage and white matter injury on preterm cerebellar growth. *The Journal of pediatrics*, 158(3):366–371, 2011.

[131] Emily WY Tam, Elysa Widjaja, Susan I Blaser, Daune L MacGregor, Prakash Satodia, and Aideen M Moore. Occipital lobe injury and cortical visual outcomes after neonatal hypoglycemia. *Pediatrics*, 122(3):507–512, 2008.

[132] Tong Tong, Robin Wolz, Pierrick Coupé, Joseph V Hajnal, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Segmentation of mr images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. *NeuroImage*, 76:11–23, 2013.

[133] LC Tzarouchi, LG Astrakas, V Xydis, A Zikou, P Kosta, A Drougia, S Andronikou, and MI Argyropoulou. Age-related grey matter changes in preterm infants: an mri study. *Neuroimage*, 47(4):1148–1153, 2009.

[134] MS Van der Knaap, G van Wezel-Meijler, Peter G Barth, Frederik Barkhof, Herman J Adèr, and Jaap Valk. Normal gyration and sulcation in preterm and term neonates: appearance on mr images. *Radiology*, 200(2):389–396, 1996.

[135] Neeltje EM van Haren, Wiepke Cahn, Hilleke E Hulshoff Pol, Hugo G Schnack, Esther Caspers, Adriaan Lemstra, Margriet M Sitskoorn, Durk Wiersma, Rob J van den Bosch, Peter M Dingemans, et al. Brain volumes as predictor of outcome in recent-onset schizophrenia: a multi-center mri study. *Schizophrenia research*, 64(1):41–52, 2003.

[136] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan Colchester, and Paul Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *Medical Imaging, IEEE Transactions on*, 20(8):677–688, 2001.

[137] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan Colchester, and Paul Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *Medical Imaging, IEEE Transactions on*, 20(8):677–688, 2001.

[138] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE transactions on medical imaging*, 18(10):897–908, 1999.

[139] Joseph J Volpe. Intraventricular hemorrhage in the premature infantcurrent concepts. part i. *Annals of neurology*, 25(1):3–11, 1989.

[140] Joseph J Volpe. Cerebral white matter injury of the premature infantmore common than you think. *Pediatrics*, 112(1):176–180, 2003.

[141] Joseph J Volpe. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. *The Lancet Neurology*, 8(1):110 – 124, 2009.

[142] Li Wang, Feng Shi, Gang Li, Yaozong Gao, Weili Lin, John H Gilmore, and Dinggang Shen. Segmentation of neonatal brain mr images using patch-driven level sets. *NeuroImage*, 84:141–158, 2014.

[143] Simon K Warfield, Kelly H Zou, and William M Wells. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903–921, 2004.

[144] Williams M Wells III, W Eric L Grimson, Ron Kikinis, and Ferenc A Jolesz. Adaptive segmentation of mri data. *Medical Imaging, IEEE Transactions on*, 15(4):429–442, 1996.

[145] Nicholas S. Wood, Neil Marlow, Kate Costeloe, Alan T. Gibson, and Andrew R. Wilkinson. Neurologic and developmental disability after extremely preterm birth. *New England Journal of Medicine*, 343(6):378–384, 2000. PMID: 10933736.

[146] Lianne J. Woodward, Peter J. Anderson, Nicola C. Austin, Kelly Howard, and Terrie E. Inder. Neonatal mri to predict neurodevelopmental outcomes in preterm infants. *New England Journal of Medicine*, 355(7):685–694, 2006. PMID: 16914704.

[147] Guorong Wu, Qian Wang, Daoqiang Zhang, Feiping Nie, Heng Huang, and Dinggang Shen. A generative probability model of joint label fusion for multi-atlas based brain segmentation. *Medical image analysis*, 18(6):881–890, 2014.

[148] Michelle X.H. Yan and Joel S. Karp. An adaptive bayesian approach to three-dimensional mr brain segmentation. In *Information processing in medical imaging*, volume 3, page 201. Springer, 1995.

[149] Xintian Yu, Yanjie Zhang, Robert E Lasky, Sushmita Datta, Nehal A Parikh, and Ponnada A Narayana. Comprehensive brain MRI segmentation in high risk preterm newborns. *PLoS One*, 5(11):e13874, 2010.

[150] Sandra Juul Zachary Hill, Mengyuan Liu and Colin Studholme. A hybrid premature neonatal segmentation pipeline for clinical brain imaging acquired without dedicated neonatal coils. Presented at International Society for Magnetic Resonance in Medicine (ISMRM) 2016 Annual Meeting, Singapore, 2016.

[151] Wenlu Zhang, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, and Dinggang Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224, 2015.

[152] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on*, 20(1):45–57, 2001.

[153] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[154] Jill G Zwicker, Steven P Miller, Ruth E Grunau, Vann Chau, Rollin Brant, Colin Studholme, Mengyuan Liu, Anne Synnes, Kenneth J Poskitt, Mikaela L Stiver, and Emily Tam. Smaller cerebellar growth and poorer neurodevelopmental outcomes in very preterm infants exposed to neonatal morphine. *The Journal of pediatrics*, 172:81–87, 2016.

Appendix A

LIST OF PUBLICATIONS & CONFERENCE TALKS

Peer-Reviewed Journal Article

[1] Daphne Kamino, Colin Studholme, **Mengyuan Liu**, Vann Chau, Steven Miller, Anne Synnes, Elizabeth Rogers et al. "Postnatal polyunsaturated fatty acids associated with larger preterm brain tissue volumes and better outcomes." Pediatric research (2017).

[2] Vidya Rajagopalan, Julia Scott, **Mengyuan Liu**, Kenneth Poskitt, Vann Chau, Steven Miller, Colin Studholme. Complementary cortical gray and white matter developmental patterns in healthy, preterm neonates. Human Brain Mapping, Volume 38(9), September 2017, Pages 4322-4336.

[3] Jill Zwicker, Steven Miller, Ruth Grunau, Vann Chau, Rollin Brant, Colin Studholme, **Mengyuan Liu**, Anne Synnes, Kenneth Poskitt, Milaela Stiver, and Emily Tam. Smaller Cerebellar Growth and Poorer Neurodevelopmental Outcomes in Very Preterm Infants Exposed to Neonatal Morphine. The Journal of Pediatrics, Volume 172, May 2016, Pages 81-87.e2.

[4] **M**engyuan Liu, Averi Kitsch, Steven Miller, Vann Chau, Kenneth Poskitt, Francois Rousseau, Dennis Shaw and Colin Studholme. Patch-based augmentation of Expectation-Maximization for brain MRI tissue segmentation at arbitrary age after premature birth. NeuroImage, Volume 127, 15 February 2016, Pages 387-408.

Peer-Reviewed Conference Proceedings

[1] **Mengyuan Liu**, Steven Miller, Vann Chau, and Colin Studholme. Combining Spatial and Non-spatial Dictionary Learning for Automated Labeling of Intra-ventricular Hemor-

rhage in Neonatal Brain MRI. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 789-797. Springer, Cham, 2017.
[2] **Mengyuan Liu**, Sharmishtaa Seshamani, Lisa Harrylock, Averi Kitsch, Steven Miller, Vann Chau, Kenneth Poskitt, Francois Rouseau, and Colin Studholme. Spatially adapted augmentation of age-specific atlas-based segmentation using patch-based priors. In SPIE Medical Imaging, pages 9034H-90341H. International Society of Optics and Photonics, 2014.

Conference Abstracts

[1] Nicole Riley, **Mengyuan Liu**, Raley Rewa, Steven Miller, Vann Chau, Ken Poskitt, Ruth Grunau, Anne Synnes, Dennis Shaw and Colin Studholme. Occipital lobe subdivision volumes in relationship to neurological delays in premature neonates. OHBM 2016 Annual Meeting, Geneva, Switzerland, 2016.
[2] Zachary Hill, **Mengyuan Liu**, Sandra Juul and Colin Studholme. A hybrid premature neonatal segmentation pipeline for clinical brain imaging acquired without dedicated neonatal coils. ISMRM 2016 Annual Meeting, Singapore, 2016.
[3] **Mengyuan Liu**, Averi Kitsch, Emily Tam, Daphne Kamino, Steven Miller, Vann Chau, J A Barkovich, Duan Xu, and Colin Studholme. Cross-vendor, cross-site cross-protocol evaluation of premature neonatal MR brain tissue segmentation. OHBM 2015 Annual Meeting, Honolulu, Hawaii, USA, 2015.
[4] Averi Kitsch, Sinchai Tsao, **Mengyuan Liu**, Steven Miller, Vann Chau, Ruth Grunau, Anne Synnes, and Colin Studholme. The progression of regional cortical folding in preterm brains without MRI visible injury. OHBM 2015 Annual Meeting, Honolulu, Hawaii, USA, 2015.
[5] Sinchai Tsao, Averi Kitsch, **Mengyuan Liu**, Sharmishtaa Seshamani, Steven Miller, Vann Chau, Kennith Poskitt, and Colin Studholme. Cortical surface-based mapping of maturation-related MR tissue intensity changes in preterm neonates. OHBM 2015 Annual Meeting, Honolulu, Hawaii, USA, 2015.

[6] **Mengyuan Liu**, Lisa Harrylock, Averi Kitsch, Dennis Shaw, Steven Miller, Vann Chau, Kennith Poskitt, Francois Rousseau and Colin Studholme. Automated outlier-based Intraventricular Hemorrhage segmetnation (OBIHS) in premature neonatal MRI. OHBM 2014 Annual Meeting, Hamburg, Germany, 2014.

[7] Averi Kitsch, **Mengyuan Liu**, Sharmishtaa Seshamani, Lisa Harrylock, Steven Miller, Vann Chau, Kennith Poskitt, and Colin Studholme. Regional neonatal brain tissue volumes after premature birth and neurocognitive delay at 18 months. OHBM 2014 Annual Meeting, Hamburg, Germany, 2014.

[8] Lisa Harrylock, Averi Kitsch, **Mengyuan Liu**, Sharmishtaa Seshamani, Steven Miller, Vann Chau, Kennith Poskitt, Anne Synnes, and Colin Studholme. Brainstem volume after premature birth and developmental outcome at 18 months. OHBM 2014 Annual Meeting, Hamburg, Germany, 2014.

[9] **Mengyuan Liu**, Julia Scott, Vann Chau, Kenneth Poskitt, Steven Miller and Colin Studholme. An evaluation of age-specific atlas-based MRI brain segmentation in premature neonates. BMES 2013 Annual Meeting, Seattle, Washington, USA, 2013.

VITA

Mengyuan Liu was born in Dalian, Liaoning Province, China in 1989. She attended Dalian Maritime University Affiliated Elementary and Junior High School and Yuming Senior high school, one of the top schools in Liaoning Province. During her time there, she found love for science, biology and received the First Prize in the Olympic Biology Competition in China. This honor gave her the free pass to Zhejiang University, one of the top universities in China. She also enjoyed sports and sports, especially piano playing.

Mengyuan obtained her undergraduate degree in Biomedical Engineering from Zhejiang University in 2012. While at Zhejiang University, she was actively involved in research in brain-computer interface. She designed and lead an undergraduate research team and completed a 2-year research project to use intra-cranial stimulus to control a rat to stop controllably. From this research experience, she developed great interests in understanding brain structures and functions.

After graduation from Zhejiang University, she decided to continue to pursue her passion about brain-related research in University of Washington, Seattle. There she expanded her knowledge and experience as a inter-disciplinary researcher and has found her time as part of the Biomedical Image Computing Group an exciting and rewarding challenge. Beyond the significant growth in technical field, critical thinking and communication skills, she also acquired the valuable ability to thrive in difficult situations. Beyond her graduate studies, she also enjoyed the abundant nature offered by the beautiful Seattle with her husband and two dogs.