Machine learning for automatic segmentation of neonatal and fetal MR brain images

UMC Utrecht Brain Center

Nadieh Khalili

Machine learning for automatic segmentation of neonatal and fetal MR brain images

Nadieh Khalili

ISBN: 978-90-393-7324-8

Cover design by Nadieh Khalili. Printed by GildePrint.

Machine learning for automatic segmentation of neonatal and fetal MR brain images

Machine learning voor automatische segmentatie van
neonatale en foetale MR-hersenbeelden
(met een samenvatting in het Nederlands)

Proefschrift

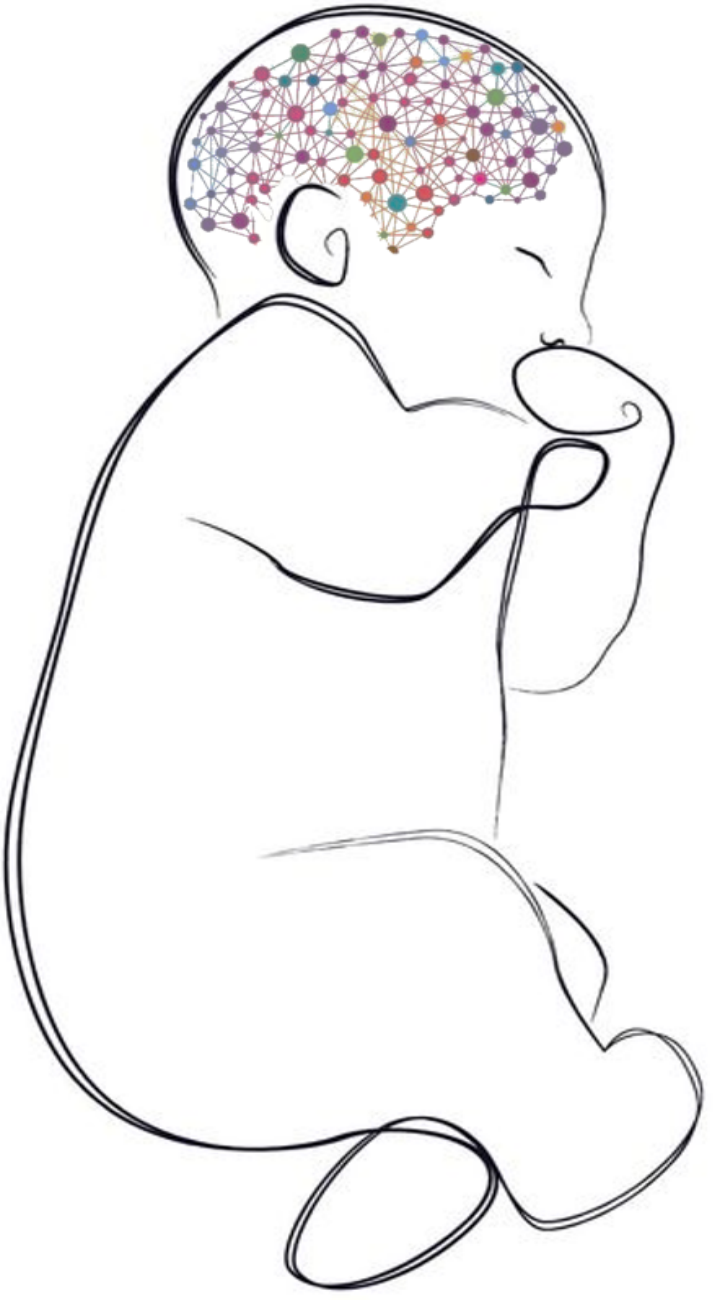
ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. H. R. B. M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op dinsdag 22 september 2020 des middags te 12.45 uur

door

Nadieh Khalili
geboren op 12 juli 1989
te Teheran, Iran

Promotoren: Prof. dr. ir. I. Išgum
Prof. dr. ir. M. A. Viergever
Prof. dr. M. J. N. L. Benders

Contents



Chapter 1

Introduction

1.1 Brain development in infants

Important neurodevelopmental changes occur in the last trimester of pregnancy, i.e., between 30 and 40 weeks of gestation [1–3]. During this rapid brain development period, fetuses and preterm born infants are at higher risk of developing neurodevelopmental problems later in life. Monitoring brain status assists clinicians in detecting impairments and in intervening if needed. Magnetic Resonance Imaging (MRI) is increasingly used to monitor the infant brain both in utero and ex utero, so as to detect injuries that may have a long-term effect on neurocognitive and motor development [4, 5]. However, visual evaluation of images does not always quantify the severity of the injury sufficiently and it is, therefore, not optimal to predict long-term effects. Quantifying brain development, using e.g. volumetric growth and morphology measurements, may aid in establishing injuries more reliably and can be correlated more accurately to cognitive outcome measures [6–8]. To quantify brain development and pathology with MRI, segmentation of fetal and neonatal brain tissues is a prerequisite.

1.2 Fetal MRI Segmentation

Fetal MRI is widely used to investigate brain development in utero. MR images do not only visualize brain tissue classes and identify abnormalities but are also able to quantify brain development using volumetric measurements and cortical folding. Such analysis allows us to investigate brain development in case of antenatal abnormalities and to compare this quantitatively with the development of healthy infants. Fetal MRI visualizes fetal brain, maternal body and infant's body (see Figure 1.2). A prerequisite for quantification is the segmentation of the fetal brain into different tissue classes, such as cortex, white matter and gray matter. Performing this segmentation manually is extremely time-consuming and requires a high level of expertise, not only because of the complex convoluted shapes of the different tissues but also owing to the low image resolution and to fetus motion (see Figure 1.1). Moreover, MR intensity values in tissues are not fully homogeneous in some slices and tissue contrasts are difficult to discern. Consequently, automatic brain tissue segmentation is challenging.

1.3 Neonatal MRI Segmentation

Neonatal MRI analysis is of great clinical interest to quantitatively measure brain development and to thereby aid in diagnosis and treatment decisions. Besides anomaly and defects in newborns, premature birth is the most frequent complication in neonates with an incident of 1 per 8.6 [9]. Premature infants develop the brain ex utero which may lead to brain injury or secondary developmental issues [10, 11]. Obtaining image segmentations manually in a large-scale study is both subjective and time-consuming. Moreover, neonatal MRI frequently suffers from artifacts caused by the infant's mo-

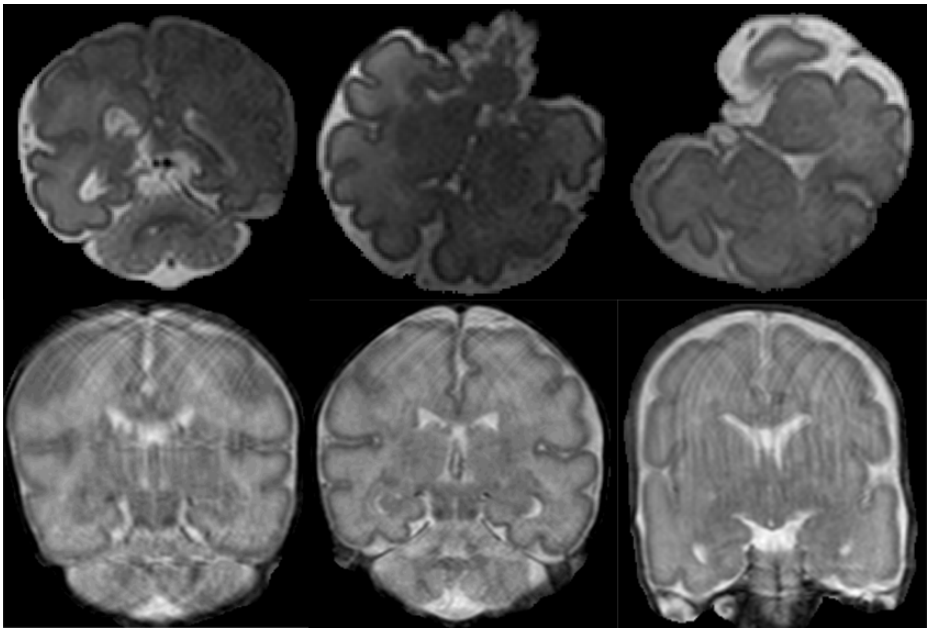


Figure 1.1: Examples of fetal and neonatal MRI with artifacts. The top row shows slices from fetal MRI where intensity inhomogeneity artifacts are visible, especially in the CSF. The bottom row shows slices from neonatal MRI acquired at 30 weeks of PMA with visible motion artifacts caused by the movement of the infant during scanning.

tion that occurs during scanning; examples are shown in Figure 1. A further challenge in neonatal MRI segmentation is anatomical variation due to differences in brain morphology between 30 and 40 weeks postmenstrual age (PMA) or critical abnormalities that alter brain shape significantly. All these causes of large variations and abnormalities in neonatal brain MR images call for automatic segmentation methods.

1.4 Machine learning

Recently, machine learning methods and particularly deep learning achieved excellent segmentation performance in neonatal brain segmentation [12–14]. The major strength of deep learning techniques is their ability to extract features relevant to the task directly from the data. Consequently, there is no need to derive a set of handcrafted features from the image for classification or regression tasks. Instead, a convolutional neural network (CNN) connects the input and output through a number of convolution layers. These convolution layers consist of trainable convolution kernels with weights and biases which are optimized during the training using a loss function and stochastic gradient descent. CNN is the most popular neural network in image processing, classification, and segmentation tasks. Another type of neural net-

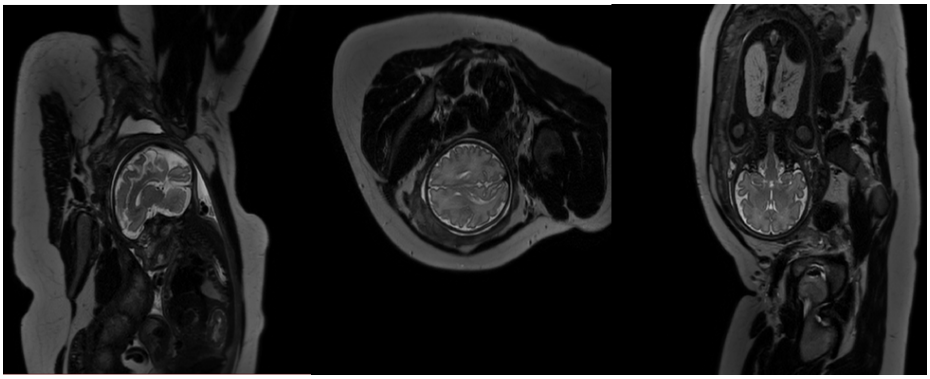


Figure 1.2: Examples of fetal MRI in sagittal (left), axial (middle) and coronal (right) plane. The fetal MRI visualizes fetal brain, maternal body and infant's body.

work recently widely used in image generation is the generative adversarial network (GAN). GAN composed of two networks, a generator, and a discriminator. These two networks are trained against each other, the generator network trains to fool the discriminator, and the discriminator train to discriminates better between real and fake data made by generator. Therefore, in image generation, the network generates more realistic images.

The objective of this thesis is to develop CNN-based automatic methods to segment neonatal and fetal MR brain images.

1.5 Outline of the thesis

This thesis presents machine learning methods for the segmentation of neonatal and fetal MRI. These methods are applied to images of premature infants and infants with abnormalities to quantify brain development:

CHAPTER 2 describes an automatic method to segment intracranial volume (ICV) in neonatal and fetal MRI. ICV segmentation is a step prior to further brain analysis. The method uses a convolutional neural network to segment ICV from a wide variety of data including images visualizing pathology neonatal MRI and scans from different centers and vendors.

CHAPTER 3 presents an automatic method to segment fetal MRI into seven tissue classes using a fully convolutional neural network. In addition, a data augmentation technique that improves segmentation in scans with intensity inhomogeneity artifacts was proposed. The method is validated on neonatal MRI with the same type of image artifacts and demonstrates that the proposed method generalizes to these images.

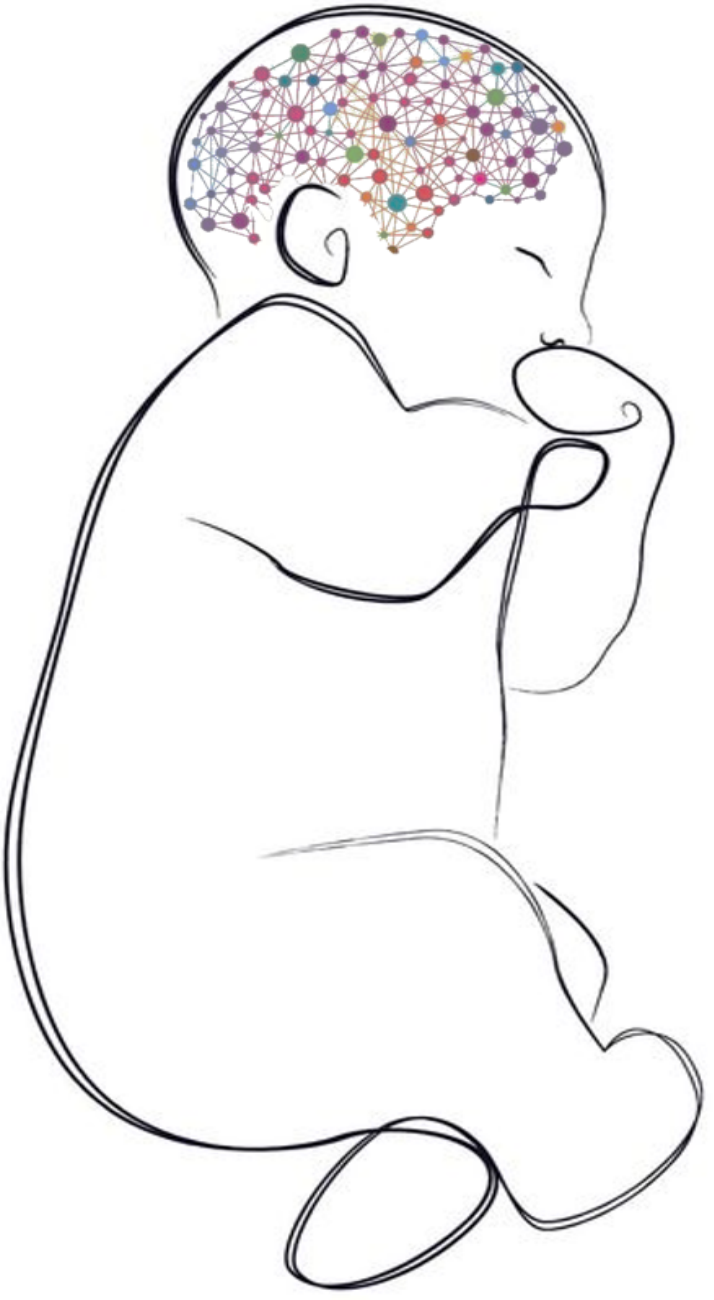
CHAPTER 4 describes a method to segment brain tissue classes in scans with motion artifacts. The proposed method employs a generative adversarial network to transform slices affected by motion artifacts into slices without artifacts, and vice versa. The method was evaluated quantitatively and qualitatively.

CHAPTER 5 describes a study on brain volume quantification from the third trimester of pregnancy through neonatal life in patients with congenital heart disease (CHD). To quantify fetal and neonatal MRI brain tissue volumes, the method described in Chapter 2 and 3 was applied.

CHAPTER 6 describes a study on the effects of the early and late intervention of infants with posthemorrhagic ventricular dilatation (PHVD) of the ventricular volume and any additional brain injury using term-equivalent age MRI. To quantify brain tissue volume in neonatal MRI, the method described in Chapter 3 was applied.

CHAPTER 7 presents methods for brain tissue segmentation in infants with perinatal stroke who are scanned after birth and at the age of three months. The method relies on fully convolutional neural networks to segment brain and ischemic stroke from scans acquired after the acute phase following the stroke. The method was evaluated quantitatively and qualitatively in a larger set.

Finally in **CHAPTER 8** the finding of previous chapters are summarized and the results, as well as future directions, are discussed.



Chapter 2

Automatic extraction of ICV in fetal and neonatal MRI

BASED ON: N. Khalili, E. Turk, M. Benders, P. Moeskops, N. Claessens, R. de Heus, A. Franx, N. Wagenaar, J. Breur, M. Viergever, and I. Išgum. "Automatic extraction of the intracranial volume in fetal and neonatal MR scans using convolutional neural networks", *NeuroImage: Clinical*, vol. 24 (2019), p. 102061.

Abstract

MR images of infants and fetuses allow non-invasive analysis of the brain. Quantitative analysis of brain development requires automatic brain tissue segmentation that is typically preceded by segmentation of the intracranial volume (ICV). Fast changes in the size and morphology of the developing brain, motion artifacts, and large variation in the field of view make ICV segmentation a challenging task.

We propose an automatic method for segmentation of the ICV in fetal and neonatal MRI scans. The method was developed and tested with a diverse set of scans regarding image acquisition parameters (i.e. field strength, image acquisition plane, image resolution), infant age (23-45 weeks post menstrual age), and pathology (posthaemorrhagic ventricular dilatation, stroke, asphyxia, and Down syndrome). The results demonstrate that the method achieves accurate segmentation with a Dice coefficient (DC) ranging from 0.98-0.99 in neonatal and fetal scansregardless of image acquisition parameters or patient characteristics. Hence, the algorithm provides a generic tool for segmentation of the ICV that may be used as a preprocessing step for brain tissue segmentation in fetal and neonatal brain MR scans.

2.1 Introduction

Magnetic resonance imaging (MRI) is a clinically used non-invasive tool for monitoring brain development in fetuses and neonates. The analysis usually comprises of quantification of brain tissue volumes and cortical morphology to extract meaningful information for diagnosis or prognosis [15–21]. Automatic quantification of these indices requires segmentation of brain tissue classes. To allow dedicated analysis within the brain, automatic methods typically perform extraction of the intracranial volume (ICV) prior to further analysis [12, 22].

A number of methods for segmentation of ICV in adult MR scans have been applied to analysis of T1- and T2-weighted neonatal MR images [23–26]. Brain Extraction Tool (BET) [23] is a publicly available tool used as a preprocessing step by many automatic brain segmentation methods [22]. BET iteratively deforms a sphere to fit it on the brain surface using a geometric algorithm. Robust Brain Extraction tool (ROBEX) is another commonly used and publicly available tool for segmentation of the ICV in adult MR images [24]. ROBEX first employs a Random Forest classifier to detect the brain boundary and thereafter uses a point distribution model that ensures a plausible result. Furthermore, Brain Extraction based on non-local Segmentation Technique (BEaST) is a publicly available tool for ICV segmentation [25]. BEaST is a patch-based segmentation method exploiting the similarity between the patches in the region of interest and predefined patches in a library.

Because of the lack of publicly available tools developed for ICV segmentation of neonatal brain MRI, these methods designed to analyze brain MR scans of adults are frequently used to segment the ICV in neonatal scans. Consequently, they generally do not produce highly accurate results when applied to neonatal brain MR scans [27]. Moreover, these methods typically fail when applied to fetal MR scans. Hence, several methods specifically designed to extract the ICV in MR scans of neonates have been proposed. Yamaguchi et al. [27, 28] proposed a method for segmentation of the ICV in brain MRI of neonates and children aged between 36 weeks post menstrual age (PMA) and 4 years. The method uses fuzzy logic and it is applicable to images without severe pathology acquired sagittally. In the first step the intensity distributions of white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), fat, and other tissues visible in the scan are estimated using Bayesian classification and a Gaussian mixture model. Segmentation of brain tissue classes is thereafter performed by means of a fuzzy active surface model using distributions of WM, GM and CSF from the previous step. The qualitative evaluation of this method demonstrated improved performance over BET. Later, Mahapatra et al. [29] proposed a shape model with graph cuts for segmentation of the ICV in neonatal MRI. The shape model is generated by averaging manually labeled images which is afterwards used with graph cut for segmentation. This method was applied to term-born infants imaged at about three weeks of age. Serag et al. [28] proposed an atlas-based segmentation of the ICV. To eliminate the need for rep-

resentative training data i.e. data coming from the same distribution, atlases that are uniformly distributed were selected. The algorithm was applied to T1-weighted and T2-weighted MR scans without visible pathology of preterm infants scanned at term equivalent age. The method showed high segmentation accuracy and it outperformed publicly available tools such as BET and ROBEX.

Similar to methods dedicated to segmentation of the ICV in neonatal MR scans, a number of studies proposed segmentation of the ICV in fetal MRI. Anquez et al. [30] proposed a method that first localizes the eyes and exploits this information to segment the ICV using a graph cut approach guided by shape, contrast, and biometrical priors. The method was applied to scans with unknown fetal orientation and the results demonstrated high segmentation accuracy.

In recent years, convolutional neural networks (CNNs) have become the most popular method for automatic image segmentation in medical images [31]. Several studies investigated different CNN architectures for brain tissue segmentation [32–36] and brain extraction [37–39] in adult MRI. Wachinger et al. [40] proposed a network that combines brain extraction and brain tissue segmentation.

A few studies used CNNs to segment ICV from fetal or neonatal MRI. Rajchl et al. [41] proposed a weakly supervised deep learning approach for ICV segmentation in fetal MRI that combines a convolutional neural network and iterative graph optimization. The network was trained with bounding boxes around the brain as weak labels. The method was applied to fetal MR scans and achieved high segmentation accuracy. In another study, Rajchl et al. [42] investigated the use of crowd sourcing platform for ICV segmentation of fetal MRI using convolutional neural network. Salehi et al. [43] proposed an iterative deep learning segmentation method that uses U-net-like convolutional neural network (Auto-net). In this approach, the fetal brain is segmented from a localized bounding box which was defined manually using ITK-SNAP [44]. In a subsequent study, Salehi et al. [45] evaluated Auto-net on fetal MRI without any preprocessing steps such as defining a bounding box. The method was trained on a very large number of manually annotated fetal MRI and demonstrated accurate segmentation results in fetal scans. Recently, Khalili et al. [46] proposed multi-scale convolutional neural network for ICV segmentation of fetal MRI.

Unlike methods performing ICV segmentation directly, several methods perform brain localization as a step prior to fetal ICV segmentation [47–50]. Recently, Tourbier et al. [51] proposed a pipeline that sequentially performs ICV localization, ICV segmentation and super-resolution reconstruction in fetal MR scans. In this method a template-matching approach, with age as prior knowledge, is used to segment the ICV in fetal MRI. A limitation of template based techniques is that they are typically computationally more expensive than machine learning algorithms. In addition, they have a high chance of failure if representative age-matched templates are not available. Moreover, to segment brain tissue classes, methods employing brain localization require subsequent segmentation of the ICV.

All aforementioned methods were evaluated either on neonatal or fetal MR scans, without visible pathology. To the best of our knowledge, thus far no study proposed a generic method that performs segmentation of the ICV in neonatal and fetal MRI. In this study, we propose a method for automatic segmentation of the ICV in neonatal and fetal T2-weighted MR scans that is robust to imaging parameters (field strength, image acquisition plane, image resolution), and pathology and patient characteristics (posthaemorrhagic ventricular dilatation (PHVD), stroke, asphyxia, Down syndrome). The method employs a convolutional neural network with a U-net architecture [52]. The network was trained with a combination of fetal and preterm born neonatal scans acquired in axial, coronal and sagittal orientation. The age of patients at the time of scanning in the training set ranged from 23 to 35 weeks PMA. The method was evaluated using images of fetuses and infants between 23 weeks PMA and 3 months of age at the time of scanning, ranging from absence of visible pathology to presence of severe pathology such as stroke or PHVD. This work builds upon our preliminary study that described segmentation of the ICV in fetal MRI using a multi-scale convolutional neural network [46].

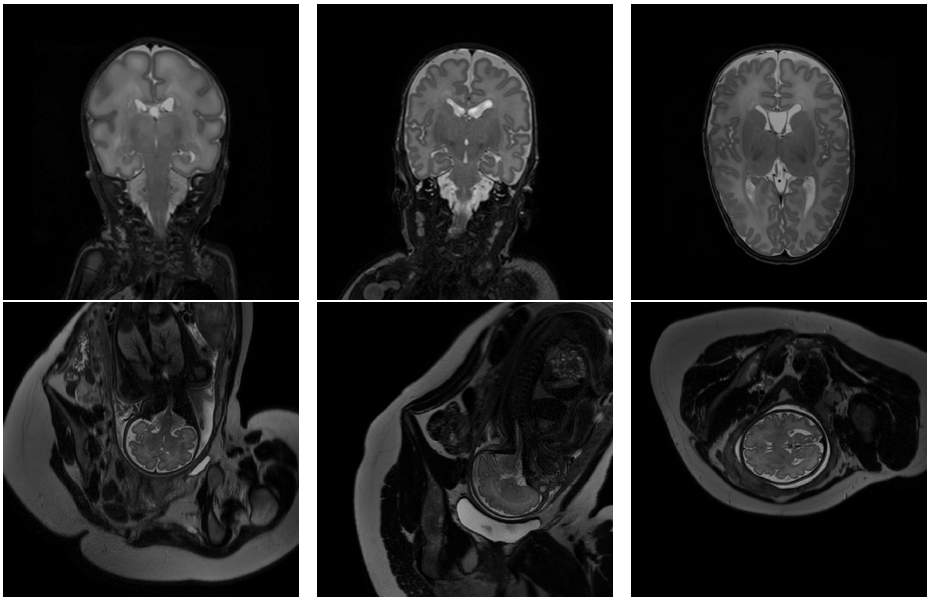


Figure 2.1: Examples of preterm neonatal and fetal MR scans included in the study. Top: coronal MRI acquired at 30 weeks PMA (left), coronal MRI acquired at 40 weeks PMA (middle), axial MRI acquired at 40 weeks PMA (right). Bottom: fetal MRI acquired in coronal (left), sagittal (middle) and axial (right) directions.

2.2 Data

In this study a diverse set of fetal and neonatal T2-weighted MR scans was used. Fetal scans were acquired in axial, sagittal and coronal image planes and did not contain visible pathology. Neonatal images include scans of preterm and term-born infants. The scans were acquired in axial or coronal image planes, and include images without and with pathology. Examples of fetal and neonatal images included in the study are illustrated in Figure 2.1. As shown in the figure, fetal MRIs have a larger field of view that visualizes the entire fetus as well as parts of the maternal body. Moreover, we include scans which were acquired with different scanner-vendors (Philips, Siemens) and field strength (1,5T and 3T). The neuroimaging data were obtained as part of the clinical protocol, written informed consent for use of the clinically acquired data and approval of the experiments and methodology was waived by the institutional review board of the University Medical Center Utrecht, The Netherlands.

2.2.1 Fetal MRI

Two sets of fetal MR scans were used. The first set (Set 1) includes T2-weighted MR scans of fetuses (age: 23-35 weeks PMA). Images were acquired on a Philips Achieva 3T scanner at the University Medical Center (UMC) Utrecht, Utrecht, the Netherlands using a turbo fast spin-echo sequence. The dataset contains 45 scans in total: 17 scans acquired in axial direction, 15 scans in coronal direction and 13 scans in sagittal direction. The images were acquired with voxel sizes of $1.25 \times 1.25 \times 2.5$ mm^3 and reconstructed to $0.7 \times 0.7 \times 1.25$ mm^3 with reconstruction matrix of $512 \times 512 \times 80$. The scans were reconstructed by the scanners algorithm and no further reconstruction (e.g. super-resolution processing) of the acquired images was performed. Furthermore, the proposed approach was applied to the 2D MRI slices without any prepossessing steps such as intensity inhomogeneity or motion correction.

The second set (Set 2) contains publicly available T2-weighted MR scans of 17 fetuses (age range: 29 ± 5 weeks PMA) which present a subset of scans described by Salehi et al. [45]. Scans were acquired on a 3T Siemens Skyra scanner at Boston Childeren's Hospital, Boston, US in axial, sagittal and coronal direction. The scans were acquired with voxel sizes of $1 \times 1 \times 2$ mm^3 with a reconstruction matrix of 256×256; the number of slices varied from 48 to 54.

The third set (Set 3) includes fetal T2-weighted MR scans acquired on Philips Achieva 1.5T scanner at the UMC Utrecht, Utrecht, the Netherlands. The dataset contains 18 scans: 6 scans were acquired in axial direction, 6 in coronal and 6 in sagittal direction. The scans were reconstructed to a voxel size of $1.18 \times 1.18 \times 1.25$ mm^3 and reconstruction matrix of $288 \times 288 \times 80$.

2.2.2 Neonatal MRI

All neonatal scans were acquired on a Philips Achieva 3T scanner at the University Medical Center Utrecht, Utrecht, the Netherlands. We divided the data according to age of the infants at the time of acquisition, image acquisition plane, and presence and type of visible pathology. As shown in Figure 2.1, there are variations in the neonatal scans, especially between 30 and 40 weeks PMA, when the brains exhibit important structural development, including cortical folding, and changes in shape and volume.

PRETERM BORN INFANTS WITHOUT VISIBLE PATHOLOGY This set consists of three different subsets. The first one - 30-weeks coronal MRI - comprises 20 scans of preterm born infants imaged at 30 weeks PMA. The second set - 40-weeks coronal MRI - contains 17 scans of preterm born infants imaged at term equivalent age. The third set - 40-weeks axial MRI - contains 15 scans of preterm born infants imaged at term equivalent age. This set includes all 22 scans from the NeoBrainS12 challenge. Detailed data description is provided in a former study [22].

CROSS-SECTIONAL COHORT A set of 10 T2-weighted MRI scans were taken from a study investigating neonatal brain development that were made shortly after birth (29-43 weeks PMA) [53]. The scans were selected to include images of 10 neonates covering the complete available infant age range. Hence, this set includes preterm and full-term born infants.

INFANTS WITH CONGENITAL HEART DISEASE (CHD) The set consists of 10 T2-weighted MRI scans of 10 patients with critical congenital heart disease (CHD). These infants were scanned before and after univentricular or biventricular cardiac repair using cardiopulmonary bypass within the first 30 days of life [54]. We selected 5 scans made before and 5 scans after surgery, of different patients. The images visualized WM lesions indicating mild to moderate brain injuries. However, the brain morphology was not significantly altered.

INFANTS WITH PHVD A set of 10 T2-weigted MRI scans of 10 infants with germinal matrix-intraventricular hemorrhage (GMH-IVH) and subsequent PHVD requiring intervention were selected randomly from a clinical study on PHVD infants [55]. The infants included in this study received a ventricular shunt that is next to the substantial ventricular dilatation visible in MR images. An example of this is illustrated in Figure 2.2. Note that the the ventricles are substantially enlarged typically resulting in a deformed brain shape. Moreover, these patients often have a temporary ventricular shunt which is visible in a number of scan slices.

INFANTS WITH STROKE This set consists of 10 T2-weighted MRI scans of 10 infants with arterial ischemic stroke [56]. These neonates were treated with 1000 IU/kg rhEPO immediately after diagnosis. A secondary MRI was performed when the patients were 3 months of age. We included 5 primary and 5 secondary scans showing WM degradation. Primary and secondary scans were not showing the same patients. Figure 2.2 illustrates an example of a secondary scan when the stroke-affected area is filled with CSF.

INFANTS WITH ASPHYXIA This set consists of 9 T2-weighted MRI scans of 9 patients with perinatal asphyxia [57]. These scans present diffuse hypoxic-ischemic injury demonstrated as hypointensities in the images that can be present throughout the brain tissue.

INFANTS WITH DOWN SYNDROME (DS) This set consist of 10 T2-weighted MRI scans imaging 10 infants with Down syndrome. In these patients, the brain volume is smaller because of delayed brain growth and gyrification compared with healthy infants [58]. Figure 2.2 illustrates a typical example of a Down syndrome infant demonstrating abnormal shape of the head, the brain and delayed gyrifcation.

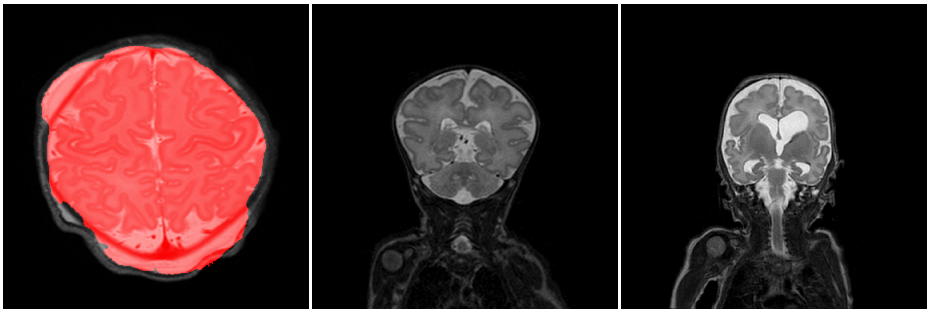


Figure 2.2: Examples of T2-weighted MR scans of preterm born neonates with ischemic stroke (left), Down syndrome (middle), and PHVD (right).

INFANT SCANS WITH ARTIFACTS This set consist of 10 T2-weighted MRI scans acquired in coronal orientation from preterm born infants imaged at term equivalent age (40 weeks of post menstrual age). 5 scans contain intensity inhomogeneity artifacts and 5 scans show motion artifacts.

Details on image acquisition parameters for all sets are listed in Table 2.1.

2.2.3 Reference standard

To establish the reference standard, manual segmentation of the ICV was performed by a trained medical student. Manual annotation was accomplished using in-house developed software by painting ICV voxels in each image slice. ICV included brain, cerebellum and extracranial cerebrospinal fluid. Skull and skin were excluded from the segmentation. We followed the definition of the eight tissue types provided by the NeoBrainS12 challenge for ending point of the brain stem [22]. Note that the reference standard for Set 3 of fetal MRI and infants scans with artifacts was not available, hence the segmentation performance on these two sets was evaluated visually.

To estimate inter-observer variability, three slices of 7 scans were segmented by different observers. Two scans from 30 weeks coronal MRI, three scans from 40 weeks coronal MRI and two scans from 40 weeks axial MRI were selected. Furthermore, the slice representing the middle of the brain and subsequently, the first and last slice on which each tissue was visible were identified.

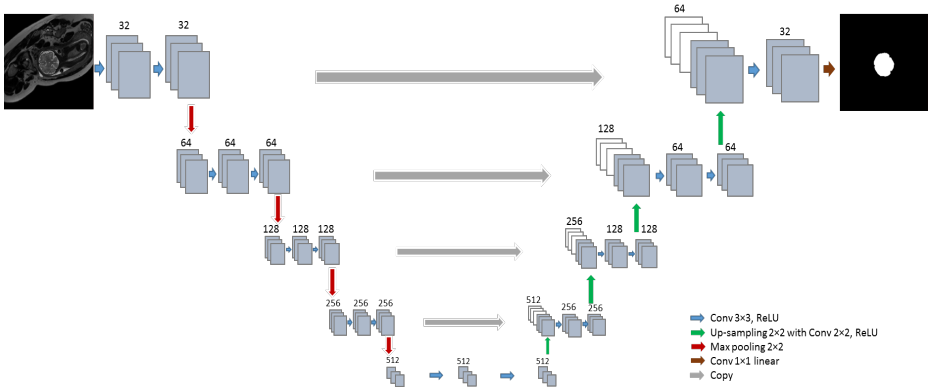


Figure 2.3: Network architecture: The network consists of a contracting path and an expanding path. The contracting path consists of repeated convolution layers followed by max pooling, and the expansion path consists of convolution layers followed by up-sampling.

2.3 Method

Our aim is to train a single network that is able to perform segmentation of the ICV in a diverse set of scans where the diversity comprises differences in field of view, age of the scanned subjects, orientation of image acquisition, image resolution and presence of pathology. Our method employs a fully convolutional network (FCN) with

	Nr	Age	Orientation	Matrix	Voxel size	Training / test
30-weeks coronal	20	30.7±1.0	Cor	384×384×50	0.34×0.34×2.0	3 / 17
40-weeks coronal	15	41.2±0.9	Cor	512×512×110	0.35×0.35×1.2	3 / 12
40-weeks axial	17	41.3±0.8	Ax	512×512×50	0.35×0.35×2.0	3 / 14
Cross-sectional cohort	10	36.9±5.0	Cor	512×512×110	0.35×0.35×1.2	0 / 10
Infants with CHD	10	41.0±1.7	Cor	512×512×110	0.35×0.35×1.2	0 / 10
Infants with PHVD	10	41.0±0.7	Cor	512×512×110	0.35×0.35×1.2	0 / 10
Infants with stroke	10	44.1±6.2	Ax	512×512×50	0.35×0.35×2.0	0 / 10
Infants with asphyxia	9	39.2±1.7	Ax	512×512×50	0.35×0.35×2.0	0 / 10
Infants with DS	10	37.9±5.9	Cor and Ax	512×512×110	0.35×0.35×1.2	0 / 10
Scans with artifacts	10	41.12±0.7	Cor	512×512×110	0.35×0.35×1.2	0 / 10

Table 2.1: Parameters of neonatal MRI scans. For each set the table lists total number of scans (Nr) in a set, average and standard deviation of the infant age at the time of scanning expressed in weeks of PMA (Age), image acquisition (Orientation) as axial (Ax) or coronal (Cor), reconstruction matrix (Matrix), reconstructed voxel sizes expressed in mm and the number of scans used in training and test set (Training/Test).

U-net like architecture [52] since such networks have demonstrated accurate segmentation performance in a number of different segmentation tasks [31]. We have used a smaller version of U-net to avoid over-fitting. The network has a contracting path and an expanding path. The contracting path consists of repeated 3 × 3 zero padded convolutions where each convolution is followed by a rectified linear unit (ReLU). 2 × 2 max pooling layers with stride 2 downsample the feature maps. The number of the feature maps doubles after every two convolutional layers. In the expanding path, upsampling with stride 2 is followed by a 2 × 2 transposed convolution which halves the number of feature channels. The resulting feature maps were concatenated with the corresponding feature map of the contracting path and convolved by two 3 × 3 convolutional layers followed by ReLU. At the final layer, a 1 × 1 convolutions map each component of feature vector to the desired number of classes (Figure 7.2). A softmax function is applied in the last layer to classify ICV and background. As a loss function, cross-entropy between the output layer and the manual segmentation reference is used. For optimization, Nesterov Adam optimizer is applied [59, 60]. In order to increase the mean learning rate, batch normalization [61] is used after each convolutional layer (Convolution, Batch Normalization, ReLU)[61]. The learning rate of Adam optimization is set to 0.0001. The hyper-parameters were tuned using cross-validation on the training set. The training was stopped after 300 epochs when the loss function became stable. The network is trained with 2D slices and batch size is 30 for each iteration. The image intensity were normalized to the range [0, 1023] before feeding them to the network. Data augmentation was applied during the training by random flipping and rotation of 2D slices. The rotations ranged between 0 to 360 degrees to mimic fetal brain angle variations. As all image intensities were normalized between [0, 1023], we did not vary image intensities nor the contrast as an augmentation. We have implemented the network in Keras, an open-source neural-network library written in Python [62].

2.4 Evaluation

The automatic ICV segmentations were evaluated in 3D by means of the Dice coefficient, the mean surface distance and the Hausdorff distance [63] between the manual and automatic segmentations per image per set.

2.5 Experiments and Results

2.5.1 Training with joint neonatal and fetal scans

We performed segmentation in fetal and neonatal MRI scans using a single trained network. The training set consisted of 21 fetal and 9 neonatal scans. Fetal scans in the training contained 7 scans acquired in axial, 7 scans acquired in coronal and 7 scans

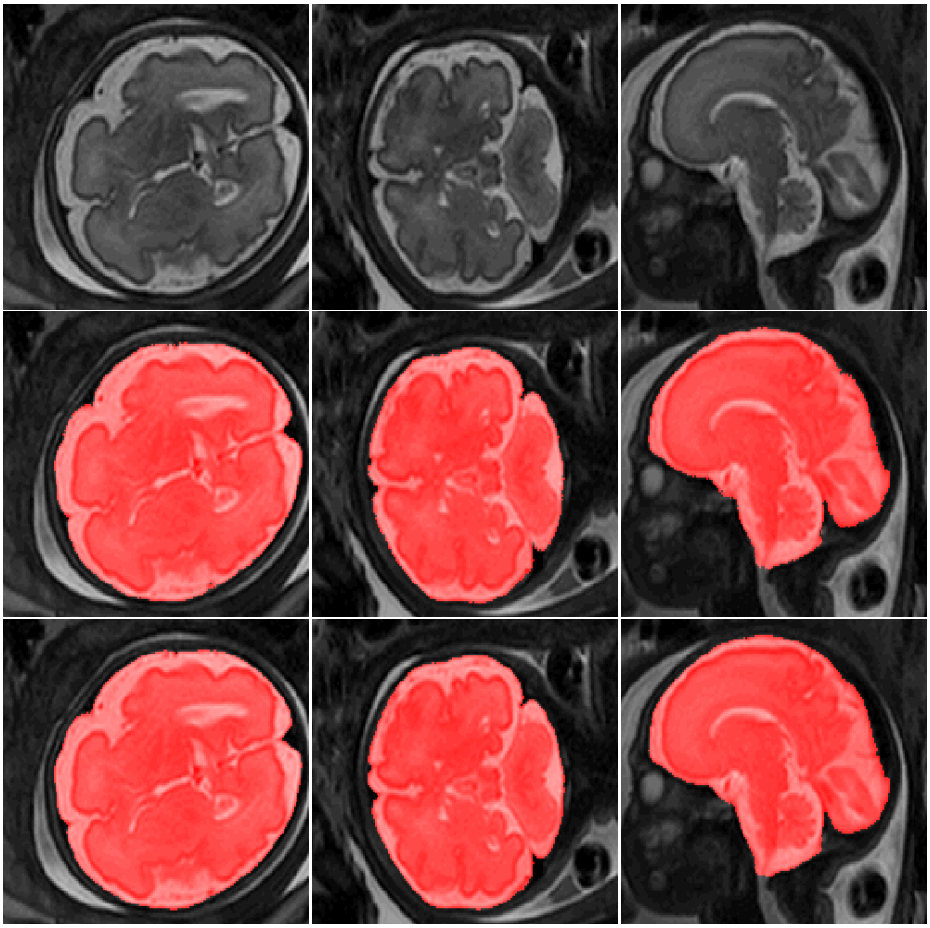


Figure 2.4: Examples of ICV segmentation in slices from fetal scans acquired in axial (left), coronal (middle) and sagittal (right) image planes. The images are selected from the test set. A slice from T2-weighted image (top); segmentation achieved by the proposed method trained with a combination of neonatal and fetal MRI (middle); manual segmentation (bottom).

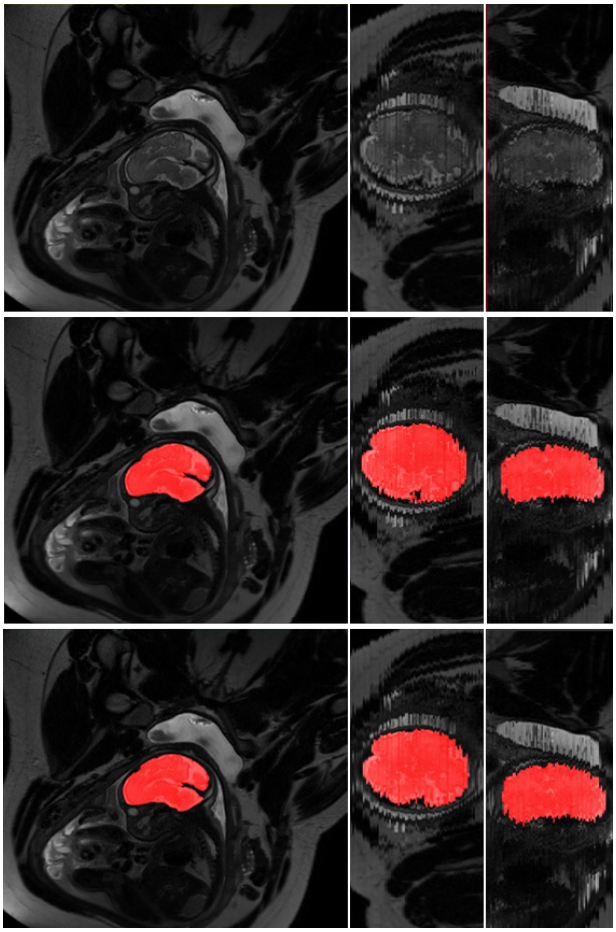


Figure 2.5: Examples of ICV segmentation in a scan acquired in sagittal plane. A slice from T2-weighted fetal MRI scan (top row), segmentation obtained with joint training (middle row) and manual segmentation (bottom row). The first column illustrates the segmentation in the in-plane view. Second and third columns illustrated out-of-plane views. The slices were selected from the test set.

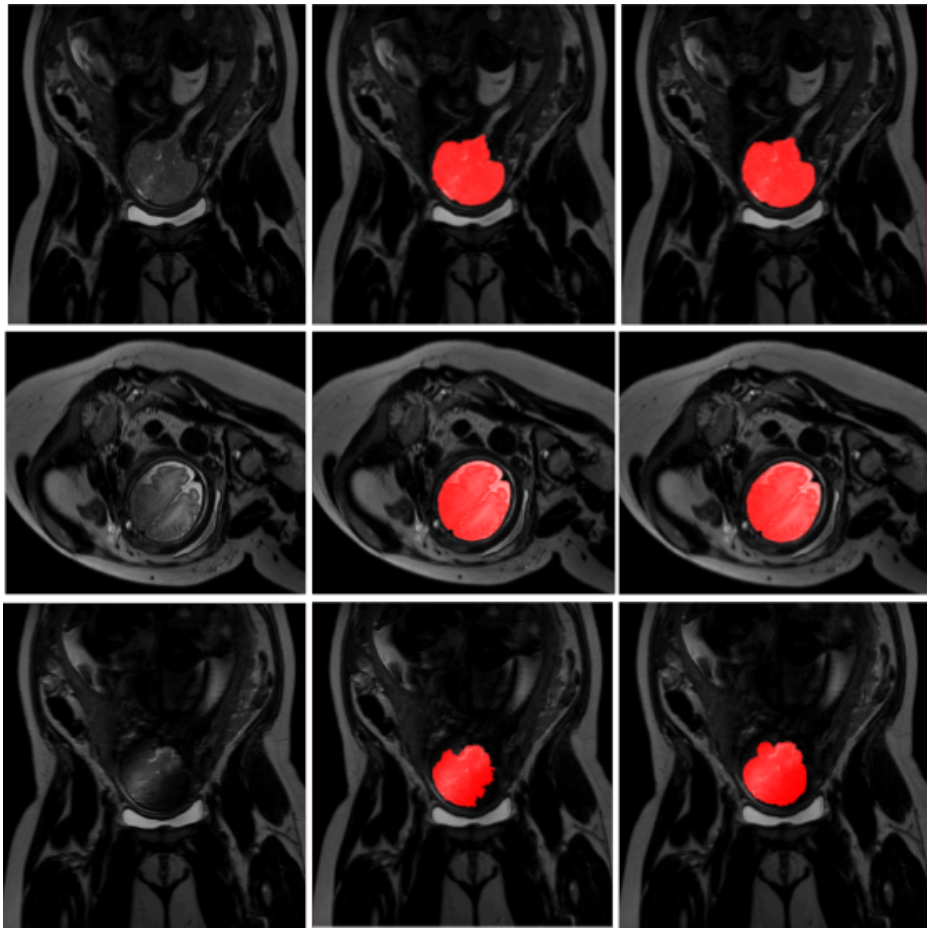


Figure 2.6: Examples of ICV segmentation in slices from fetal scans that visualized intensity inhomogeneity. A slice from T2-weighted image (left); segmentation achieved by the proposed method trained with a combination of neonatal and fetal MRI (middle) and manual segmentation (right). The images were selected from the test set.

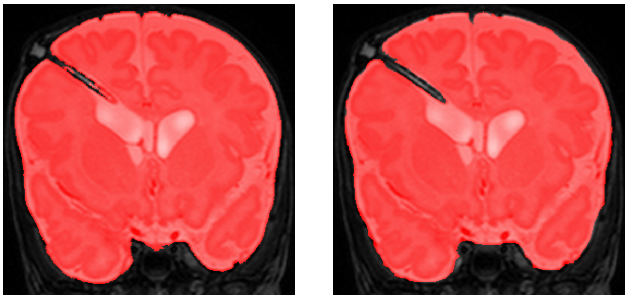


Figure 2.7: Example of ICV segmentation in one test neonate with PHVD on the left compared with manual segmentation on the right.The infants received a temporary ventricular shunt that is visible in some slices. The images were selected from the test set.

Figure 2.8: A slice from a scan of infant with PHVD (left) where the joint training undersegmented cerebellum (middle) compared with reference annotation (right). The cerebellar volume, shape and image intensity are typically different in infants with PHVD from infants without visible pathology. The images were selected from the test set.

	Neonatal and Fetal			Fetal			Representative		
	DC	MSD	HD	DC	MSD	HD	DC	MSD	HD
Set 1	0.976	0.34	**10.58**	**0.980**	**0.32**	16.89	0.978	0.36	15.45
	Neonatal and Fetal			Neonatal			Representative		
	DC	MSD	HD	DC	MSD	HD	DC	MSD	HD
30-weeks coronal	**0.993**	**0.11**	**6.19**	0.988	0.18	7.87	0.992	0.12	10.90
40-weeks coronal	0.993	0.18	**7.98**	**0.994**	**0.14**	8.32	0.993	0.16	9.64
40-weeks axial	**0.988**	**0.22**	**7.60**	0.988	0.24	7.67	0.987	0.44	25.89
Cross-sectional cohort	0.987	0.35	11.13	0.990	**0.19**	**8.88**	**0.991**	0.22	13.94
Infants with CHD	0.987	0.53	17.19	**0.990**	**0.19**	**8.57**	0.985	0.64	30.26
Infants with PHVD	0.987	**0.29**	**14.67**	**0.988**	0.31	15.89	0.986	0.35	17.34
Infants with stroke	0.987	**0.30**	14.68	**0.988**	0.46	**11.09**	0.984	0.58	19.05
Infants with asphyxia	**0.980**	**0.34**	**10.58**	0.970	0.62	15.54	0.963	0.80	16.41
Infants with DS	0.982	0.46	**12.52**	**0.983**	**0.38**	14.36	0.983	0.58	25.87

Table 2.2: Performance of the automatic segmentation expressed by the average Dice coefficient (DC), mean surface distance (MSD) in mm, and Hausdorff distance (HD) in mm. Columns show experiments where the network was trained with: 1) a combination of fetal and neonatal MRI (Fetal and Neonatal) 2) fetal MRI when the test images were from fetuses (Fetal) or neonatal MRI when test images were from neonates (Neonatal) 3) Only a representative set of images. For each test set best results among the three experiments are indicated in bold.

acquired in sagittal imaging orientations (21 scans) of 7 patients from Set 1. Neonatal scans were from preterm born infants without visible pathology. Neonatal scans included in the training consisted of 3 coronal scans acquired at 30 weeks PMA, 3 coronal scans acquired at 40 weeks PMA, and 3 axial scans acquired at 40 weeks PMA. Note that the training and test set were separated per subject. During the training, only in joint training scenario, each batch was balanced between fetal scans, 30 weeks coronal neonatal, 40 weeks axial neonatal and 40 weeks coronal neonatal scans.

The method was tested with the remaining 24 fetal scans from Set 1 that were acquired in axial, coronal and sagittal orientation, and neonatal scans of the remaining 110 patients. The obtained quantitative results are listed in Table 2.2 (first three columns). Figure 2.4 illustrates examples of the obtained ICV segmentations in images acquired in axial, coronal and sagittal image planes. Figure 2.5 illustrates ICV segmentation results in one scan acquired in sagittal imaging plane. The segmentation results are shown in the acquisition plane as well as in planes perpendicular to the acquisition plane. Furthermore, Figure 2.6 illustrates examples of the ICV segmentations in slices with intensity inhomogeneity.

Moreover, Figure 2.7 shows an example of ICV segmentation in a neonate with PHVD. The automatic segmentation excluded the inserted shunt from the brain mask even though PHVD scans were not included in the training data. Figure 2.8 illustrates another example of ICV segmentation in a neonate with PHVD where the cerebellum was undersegmented. It may be observed that in this case cerebellum has voxels of lower intensity than images without visible pathology.

2.5.2 Training with neonatal or fetal scans

Manual annotation in a large set of scans is time-consuming and expensive. Hence, to estimate whether the method performs better on fetal images when trained with fetal images only and whether it performs better on neonatal images when trained with neonatal images only, additional experiments were performed. For this, two separate networks were trained. The first network was trained using only fetal images. This set included scans of 7 fetuses with images acquired in axial, coronal and sagittal directions. The second network was trained using only neonatal images. This training set included images of infants scanned at 30- and 40- weeks PMA acquired in axial and coronal directions. In both experiments, the training images were the same training images that were used in the experiment described in Section 2.4 A when the fetal and neonatal training images were used together in the training. No other changes in the network architecture or training procedure were applied. The obtained results are listed in Table 2.2 (middle three columns).

	DC	MSD	HD
Joint training	0.94±0.02	1.7±0.72	34.5±16.11
Auto-net	**0.98±0.01**	**0.2±0.04**	**10.1±5.45**

Table 2.3: Performance of the proposed method using joint training with 21 fetal and 9 neonatal MRI scans, and performance of the publicly provided Auto-net trained with 260 fetal MRI scans. Both methods were tested on publicly available fetal scans from Set 2. The results are expressed by the average Dice coefficient (DC), mean surface distance (MSD) in mm, and Hausdorff distance (HD) in mm.

	DC	MSD	HD
Joint training	**0.98±0.02**	**0.3±0.36**	**10.6±5.73**
Auto-net	0.87±0.10	3.2±1.84	72.41±50.34

Table 2.4: Performance of the proposed method using joint training with 21 fetal and 9 neonatal MRI scans and performance of the publicly available Auto-net trained with 260 MRI scans. Both methods were tested with fetal images from Set 1. The results are expressed by average Dice coefficient (DC), Mean surface distance (MSD) in mm, and Hausdorff distance (HD) in mm.

2.5.3 Training with representative scans

To evaluate whether it might be advantageous to train the network using representative data only, three instances of the original network were trained. One instance was trained and tested with scans of neonates acquired at 30 weeks PMA coronal, another instance was trained and tested with scans of neonates acquired at 40 weeks PMA coronal, and last instance was trained and tested with scan of neonates acquired at 40 weeks PMA axial. Training images represent subsets of scans used in the experiment where all training data was mixed. The obtained results are listed in Table 2.2 (last three columns).

2.5.4 Second observer evaluation

To evaluate inter-observer variability, we obtained second observer manual annotations for small subset of neonatal data. The evaluation was performed on 3 slices of 7 scans, i.e 21 slices in total and the results were compared to corresponding slices annotated by the first observer. The results are listed in Table 2.6.

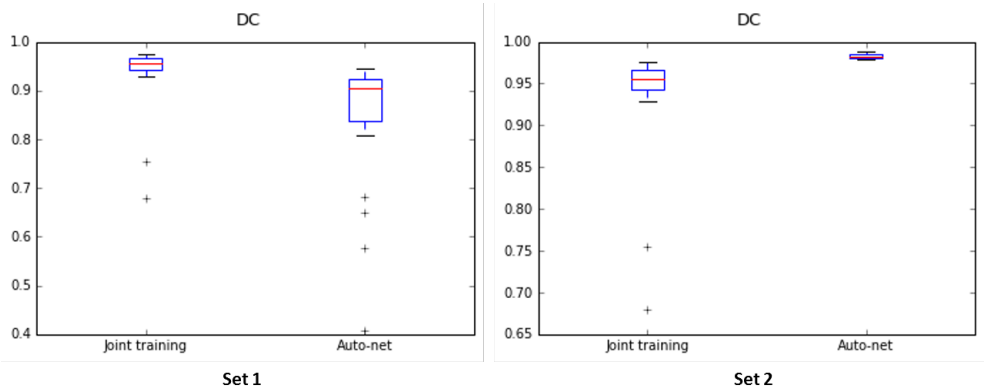


Figure 2.9: Dice coefficients achieved by the proposed method using joint training with 21 fetal and 9 neonatal MRI scans, and by the publicly provided Auto-net trained with 260 fetal MRI scans. Both methods were tested on Set 1 (left) and Set 2 (right).

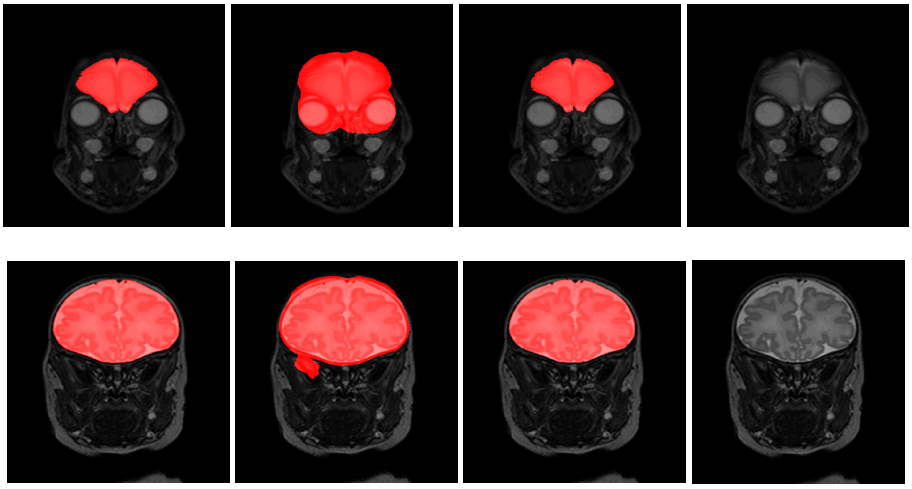


Figure 2.10: Examples of ICV segmentation in neonates acquired at 30 weeks PMA (top row) and 40 weeks PMA (bottom row). Results of the joint training (first column), the result obtained with BET (second column), manual annotation (third column) and the original T2-weighted MRI (last column). The images were selected from the test set.

	Joint training			BET		
	DC	MSD	HD	DC	MSD	HD
30-weeks coronal	0.99	0.11	6.19	0.91	2.05	24.92
40-weeks coronal	0.99	0.18	7.98	0.94	1.92	27.91
40-weeks axial	0.99	0.22	7.60	0.94	1.39	34.63
Cross-sectional cohort	0.99	0.35	11.13	0.93	1.83	26.76
Infants with CHD	0.99	0.53	17.19	0.95	2.36	36.46
Infants with PHVD	0.99	0.29	14.67	0.94	1.72	24.16
Infants with stroke	0.99	0.30	14.68	0.95	1.33	30.80
Infants with asphyxia	0.98	0.34	10.58	0.95	1.33	32.88
Infants with DS	0.98	0.46	12.52	0.95	1.43	14.69

Table 2.5: Performance of the joint training using fetal and neonatal scans for ICV segmentation compared with BET. The results are expressed using the average Dice coefficient (DC), mean surface distance (MSD) in mm, and Hausdorff distance (HD) in mm.

2.5.5 Comparison with state-of-the art methods

The performance of the proposed method was compared with publicly available ICV segmentation tools. Given that BET is frequently used to segment the ICV in premature neonatal images [12], we have applied it to segment images in our test set. The fractional intensity threshold (-f) is empirically set to 0.3. The obtained results are presented in Table 2.5. They demonstrate that BET achieved better performance in neonatal MRI acquired at 40 weeks PMA than in neonatal MRI acquired at 30 weeks PMA. Figure 2.10 shows segmentations obtained with BET and joint training in a slice from a scan acquired at 30 weeks PMA and one acquired at 40 weeks PMA.

Both slices illustrate oversegmentation of the ICV along the whole boundary, which is a frequent error of the BET tool visible in our test set. Quantitative results listed in Table 2.5 show that joint training consistently achieved higher DC and lower HD and MSD than BET.

To investigate robustness of our method to variation in scanner characteristics and patient population, the joint training model was evaluated using publicly available fetal MRI scans from another hospital (Set 2). The results were compared with a publicly available Auto-net [45] [1] model trained on a much larger set of representative fetal scans from the same hospital. Even though, U-net or any fully convolutional neu-

[1] https://bitbucket.org/bchradiology/u-net/src

ral network can take any arbitrary image size but the segmentation performance will likely drop if the images in the training and test set do not have the same resolution. Given that scans in Set 2 have different voxel sizes than our fetal images (Set 1) used in the training, prior to analysis scans from Set 2 were resampled to the resolution of our training images. Furthermore, the images were normalized between [0, 1023]. All obtained results are listed in Table 2.3 and shown in Figure 2.9. Note that this model was only trained on fetal MRI and training data did not include any neonatal MRI data. Therefore, the evaluation was performed on fetal MRI only.

In addition, Auto-net was evaluated on fetal images from our hospital (Set 1). Quantitative results are listed in Table 2.4. As in the previous experiment scans from Set 1 were resampled to the same resolution of the images used to train Auto-net (Set 2). Furthermore, the scans were normalized between [0, 1023]. Figure 2.9 illustrates the segmentation performance in a box plot. Note that even though the scans were resampled to the same resolution in both experiments, the images had different field of view.

2.5.6 Evaluation on scans acquired with 1.5 Tesla scanner

To demonstrate the performance of the proposed method on images acquired with a scanner exploiting a different field strength, the joint training model was evaluated on fetal MRI scans (Set 3) acquired with 1.5 Tesla scanner. We illustrate segmentation results in the three scans without reference standard in Figure 2.11. Visual inspection of the results in these scans reveals that the joint training model produced accurate ICV segmentations in scans with different field strength, although the model was not trained with such scans.

2.5.7 Evaluation on scans with artifacts

To demonstrate the performance of proposed method on scans with intensity inhomogeneity and motion artifacts, the joint training model was evaluated on 5 neonatal scans with intensity inhomogeneity and 5 neonatal scans with motion artifacts. We illustrate segmentation results in the five scans with intensity inhomogeneity in Figure 2.12 and five scans with motion artifacts in Figure 7.8. Visual inspection of the results in these scans reveals that the joint training model produced accurate ICV segmentation in scans with motion artifacts and intensity inhomogeneity.

2.5.8 Second observer evaluation

Manual annotation in a large set of scans is time-consuming and expensive. Hence, to estimate inter-observer variability, second observer performed manual annotations in a small subset of neonatal data. The evaluation was performed on 3 slices of 7 scans,

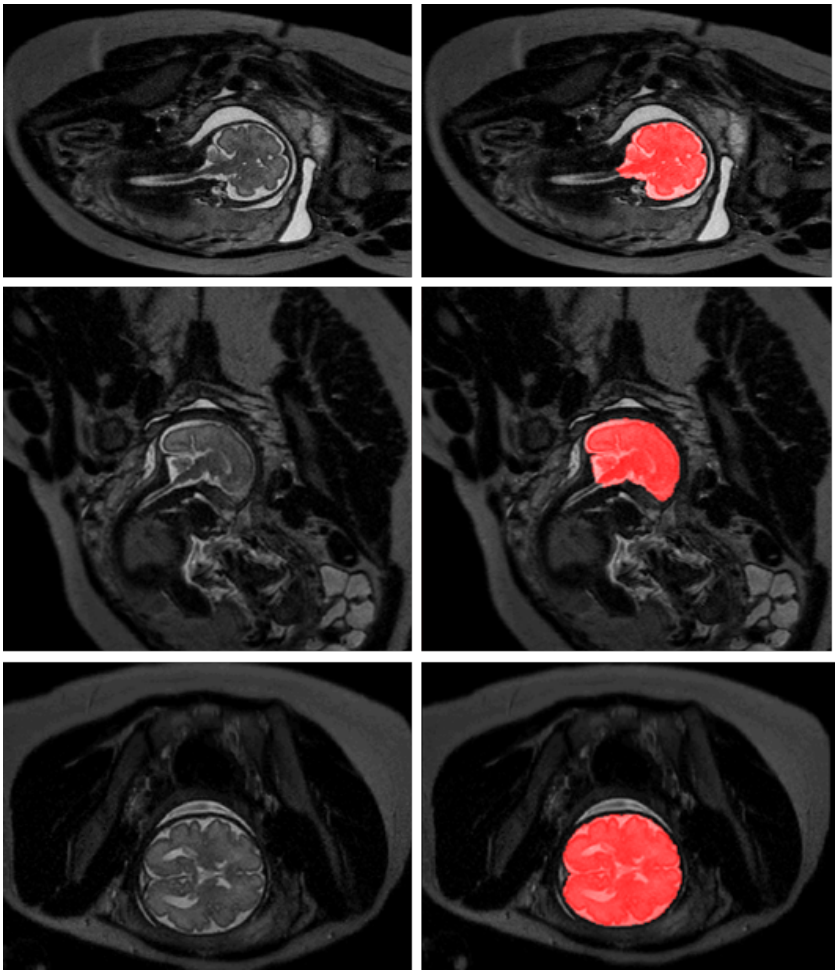


Figure 2.11: Examples of ICV segmentation in slices from fetal scans acquired with 1.5 Tesla scanner in coronal (top), sagittal (middle) and axial (bottom) image planes. A slice from T2-weighted image (left) and segmentation achieved by the proposed method trained with a combination of neonatal and fetal MRIs (right).

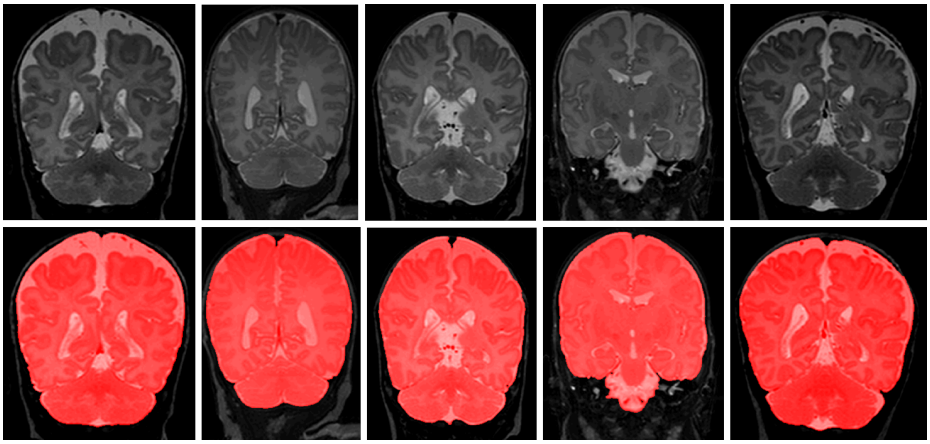


Figure 2.12: Examples of ICV segmentation in 5 neonatal MR scans with intensity inhomogeneity artifacts. A slice from T2-weighted fetal MRI scan (first row); segmentation obtained with joint training (second row).

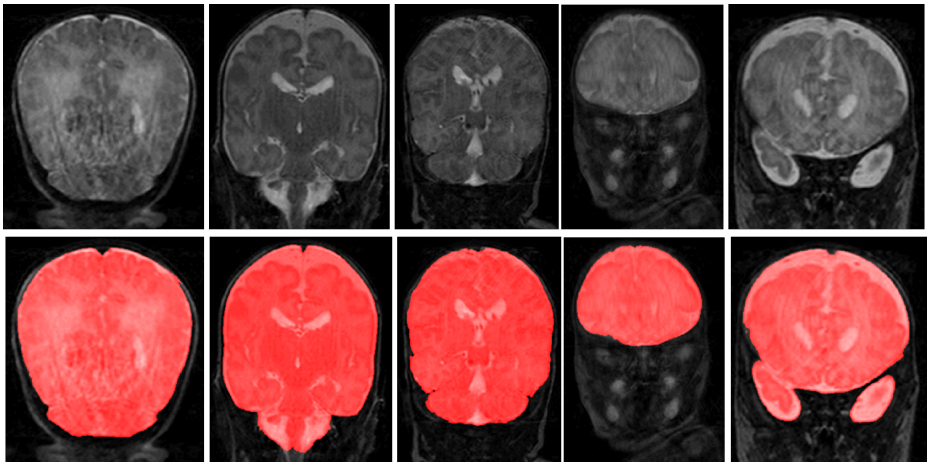


Figure 2.13: Examples of ICV segmentation in 5 neonatal MR scans with motion artifacts. A slice from T2-weighted fetal MRI scan (first row); segmentation obtained with joint training (second row).

	First Observer			Second Observer		
	DC	MSD	HD	DC	MSD	HD
30-weeks coronal	0.993	0.656	11.52	0.983	1.660	16.92
40-weeks coronal	0.992	0.891	9.231	0.994	0.474	6.332
40-weeks axial	0.993	0.828	8.752	0.992	0.979	9.620

Table 2.6: Evaluating joint training segmentation performance with manual segmentation obtained by two different observers. The results are expressed in terms of Dice coefficient (DC), Hausdorff distance (HD) and mean surface distance (MSD). The HD and MSD are expressed in mm. Note that the evaluation was performed in 3 slices in 7 scans, totalling to 21 slices. The segmentations are compared with the segmentations of the first observer in the same slices.

totally in 21 slices. These segmentations were compared with the segmentations of the first observer in the same slices. The results are listed in Table 2.6.

2.6 Discussion and conclusion

An automatic method for segmentation of the ICV in fetal and neonatal brain MR scans was presented. The proposed method employs a fully convolutional network with U-net architecture. It was trained by using a combination of neonatal and fetal MRI and the results demonstrate accurate segmentation of ICV in fetal and neonatal MR scans regardless the orientation of the image acquisition, the age of infants at the time of scanning or the presence of pathology. Unlike previous ICV segmentation methods developed for fetal or neonatal MRI [30, 48, 50], the proposed method does not require brain localization or prior information about the patient age or expected anatomy.

In this study, 2D analysis was applied. Even though 3D analysis regularly allows better exploitation of the available information compared to 2D analysis, 2D analysis was advantageous as it minimized the risk of overfitting and allowed analysis of scans with large slice thickness that led to substantial changes in the anatomy [12, 64, 65]. Moreover, 2D analysis was less influenced by missing and corrupted slices resulting from continuous fetal motion.

Generating manual segmentation is a cumbersome and extremely time-consuming task. The results illustrate that with a small number of available manual segmentation used for training, the network achieves a competitive and robust results in a large test set.Using semi-automatic segmentation comprised of automatic presegmentation and subsequent manual correction to gnerate refrence standard for training purposes could make the process faster. Availability of a large training set could offer possibility

to investigate the impact of the size of the training set on the method performance as well as research towards the requirements regarding characteristics of the training set for employment in the MRIs presenting pathology or artifacts. This may be interesting direction for future research.

Although the network was trained with images containing no visible pathology, the evaluation was performed on a large and diverse set of scans, which includes scans with pathology. The segmentation results in neonatal scans with or without lesions are comparable. Note that large lesions in the brain strongly affecting tissue appearance (infants with stroke), morphological changes (infants with Down syndrome and PHVD), and presence of implants (shunts) that were mostly excluded (PHVD) (see Figure 2.7).

We evaluated the proposed method on scans with artifacts such as intensity inhomogeneity and motion artifacts. The visual inspection demonstrate that even thought scans with artifacts were not in the training, the proposed approach is able to segment ICV.

Furthermore, we investigated whether it is feasible to train a single instance of the network applicable to both fetal and neonatal scans, or whether better performance can be achieved by training a separate network using only fetal or only neonatal scans. The results show that in both cases DC ranges from 0.98 to 0.99.

Moreover, we compared performance using joint training with fetal and neonatal scans against training using representative scans only. The results demonstrate that in both cases accurate segmentation was achieved when evaluating the overlap between automatic and reference segmentations (0.98 to 0.99 Dice coefficient). The results also demonstrate that training with diverse images using fetal and neonatal scans reduced false positive voxels far from the intracranial volume surface leading to lower Hausdorff distances and mean surface distances in all sets. Training with both fetal and neonatal scans indicated the most noticeable improvement in infants with asphyxia. Despite the differences in image acquisition, image orientation, and brain morphology, fetal and neonatal scans share common features that improve the ability of the network to generalize, making it more robust and compensating for the lack of representative data.

To investigate robustness of the proposed method to variations in scanner characteristics and patient population, the method was evaluated using publicly available fetal MRI scans from another hospital (Set 2). The results were compared with a publicly available Auto-net [45] model trained on a much larger set of representative fetal scans from the same hospital. The results show that our model did not outperform the dedicated data-specific approach. Nevertheless, it achieved DC, MSD and HD of 0.94, 1.7 and 34.5 respectively. Similarly, we evaluated Auto-net on fetal scans from our hospital (Set 1). The results demonstrate that the proposed method trained on representative fetal scans from our hospital outperformed Auto-net trained on different data. The two experiments indicate that reasonable performance can be achieved us-

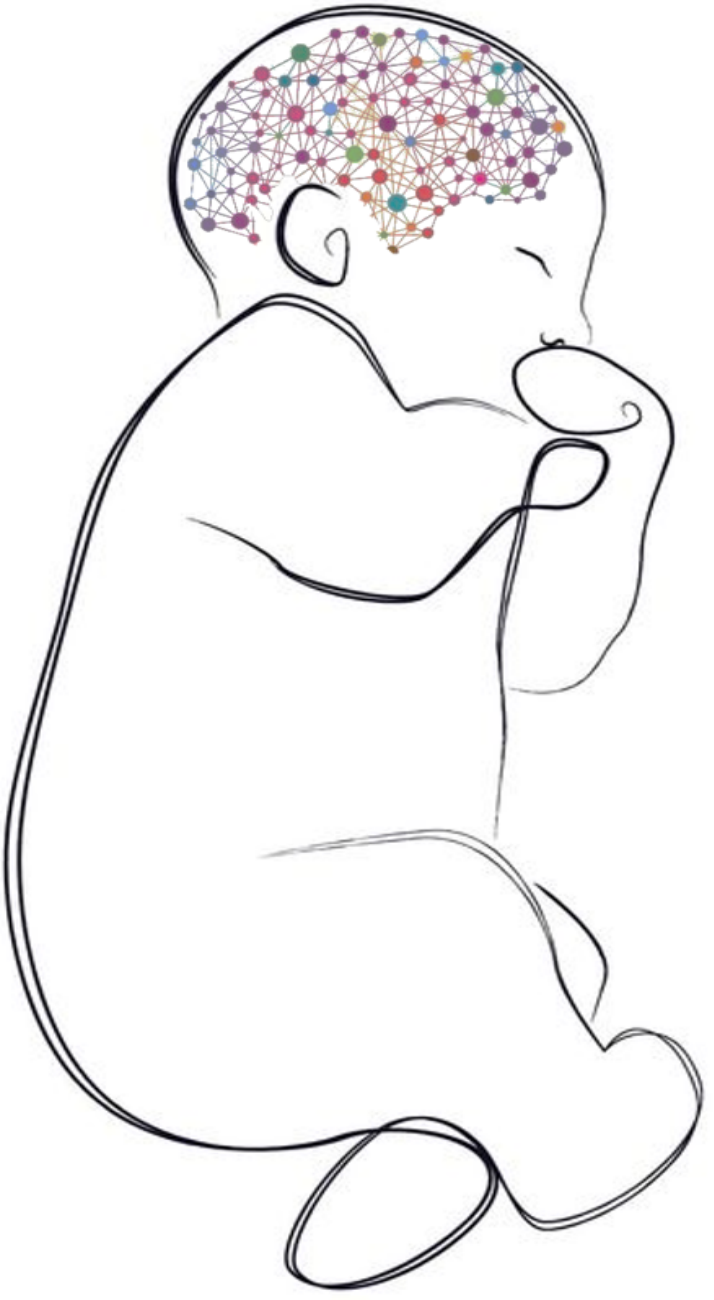
ing different scans but also underline the importance of training with representative data. In future research, investigating interpretability of model using saliency map [66] can demonstrate a better understanding of limitations in network performance.

In addition, the proposed method was compared with the publicly available and widely used BET for the segmentation of neonatal MRIs. Although BET is known to achieve accurate segmentation of ICV in adults, our results demonstrate that it is less suited for neonatal brain. Our dedicated method clearly outperformed BET.

To conclude, this study presented a method for automatic ICV segmentation in neonatal and fetal MRI. Despite the variability among the evaluated scans, the method obtained accurate segmentation results in both fetal and neonatal MR scans. Hence, the algorithm provides a generic tool for segmentation of the ICV that may be used as a preprocessing step for brain tissue segmentation in fetal and neonatal brain MR scans.

2.7 Acknowledgements

This study was sponsored by the Research Program Specialized Nutrition of the Utrecht Center for Food and Health, through a subsidy from the Dutch Ministry of Economic Affairs, the Utrecht Province and the Municipality of Utrecht.



Chapter 3

Automatic brain tissue segmentation in fetal MRI

BASED ON: N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. de Heus, T. Kolk, M. A. Viergever, M. J. Benders, and I. Išgum. "Automatic brain tissue segmentation in fetal MRI using convolutional neural networks", *Magnetic Resonance Imaging*, vol. 64 (2019), pp. 77–89.

Abstract

MR images of fetuses allow clinicians to detect brain abnormalities in an early stage of development. The cornerstone of volumetric and morphologic analysis in fetal MRI is segmentation of the fetal brain into different tissue classes. Manual segmentation is cumbersome and time consuming, hence automatic segmentation could substantially simplify the procedure. However, automatic brain tissue segmentation in these scans is challenging owing to artifacts including intensity inhomogeneity, caused in particular by spontaneous fetal movements during the scan. Unlike methods that estimate the bias field to remove intensity inhomogeneity as a preprocessing step to segmentation, we propose to perform segmentation using a convolutional neural network that exploits images with synthetically introduced intensity inhomogeneity as data augmentation. The method first uses a CNN to extract the intracranial volume. Thereafter, another CNN with the same architecture is employed to segment the extracted volume into seven brain tissue classes: cerebellum, basal ganglia and thalami, ventricular cerebrospinal fluid, white matter, brain stem, cortical gray matter and extracerebral cerebrospinal fluid. To make the method applicable to slices showing intensity inhomogeneity artifacts, the training data was augmented by applying a combination of linear gradients with random offsets and orientations to image slices without artifacts. To evaluate the performance of the method, Dice coefficient (DC) and Mean surface distance (MSD) per tissue class were computed between automatic and manual expert annotations. When the training data was enriched by simulated intensity inhomogeneity artifacts, the average achieved DC over all tissue classes and images increased from 0.77 to 0.88, and MSD decreased from 0.78 mm to 0.37 mm. These results demonstrate that the proposed approach can potentially replace or complement preprocessing steps, such as bias field corrections, and thereby improve the segmentation performance.

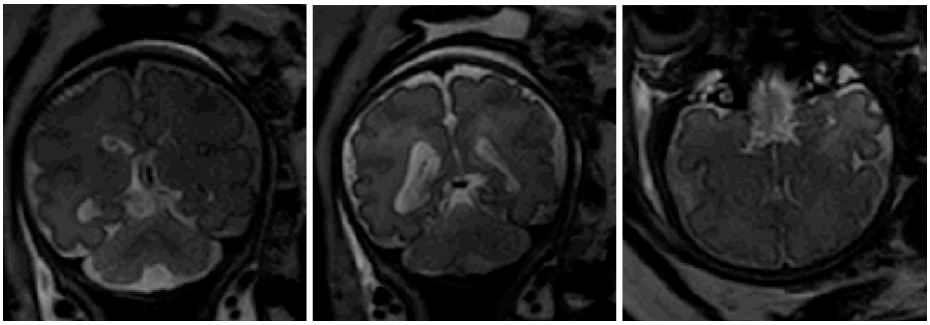


Figure 3.1: Examples of slices with intensity inhomogeneity in T2 weighted MRI of the fetus cropped to the brain. Note that the full slices also visualize a larger part of the fetus as well as the maternal body.

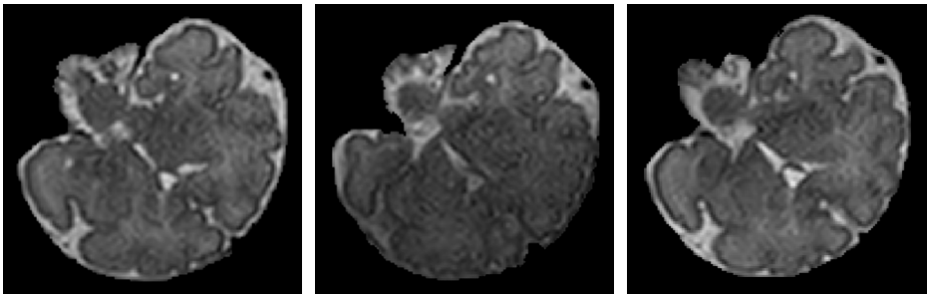


Figure 3.2: Example of a fetal T2 weighted MRI with intensity inhomogeneity (middle) and the slice before (left) and after (right) from the same scan, both without artifacts. Structures outside the fetal cranium have been masked out.

3.1 Introduction

For automatic segmentation of fetal brain tissue in reconstructed MR volumes, Habas et al. [67] proposed a method using an atlas-based expectation maximization (EM) model to segment white matter (WM), gray matter, germinal matrix, and extracerebral cerebrospinal fluid (eCSF). Prior to performing the segmentation, another EM model was used for bias field correction. Gholipour et al. [68] proposed a method for segmentation of the ventricles in fetal MRI. As a preprocessing step, in addition to using volumetric reconstructions, intensity inhomogeneity was corrected using the non-parametric entropy maximization method [69]. Initial segmentation was obtained with the use of STAPLE [70], then the final segmentation was derived with a probabilistic shape model that incorporates intensity and local spatial information. Serag et al. [71] proposed an atlas-based brain segmentation method for both neonatal and fetal MRI. Fetal scans were reconstructed into a single 3D brain volume using the slice-to-volume reconstruction method described in [72] and intensity inhomogeneity was

removed using the N4 algorithm [73]. Thereafter, the fetal brain scans were segmented into cortex, ventricles and hemispheres.

Deep learning methods have recently been very successful and have often outperformed traditional machine learning and model-based methods in medical image analysis [31] including brain MRI [36, 74]. A major strength of these networks is their ability to extract the features relevant for the tasks directly from the data. There is no need anymore to first derive a set of handcrafted features from the image as input to a classifier or model, the networks rather learn themselves to extract and interpret features relevant to the segmentation task. Therefore, deep learning methods often achieve a better performance than traditional machine learning methods with handcrafted features. However, CNNs usually require large sets of diverse training data. To enlarge the size of training set and to ensure robustness to expected variability in the data, some studies use data augmentation techniques such as random rotation, random translation and random noise injection [75, 76]. We therefore hypothesize that, while artifacts such as intensity inhomogeneity are challenging for traditional approaches and therefore normally require preprocessing of the images, CNNs may be able to adapt and become invariant to such artifacts if they are presented enough examples during training. However, manual segmentation of slices with intensity inhomogeneity is much more cumbersome than segmentation of artifact free slices so that a sizable training database is difficult to obtain. We therefore propose to tackle one of the most common artifacts in fetal MRI, namely intensity inhomogeneity, by randomly adding synthetic intensity inhomogeneity to slices for which a corresponding reference segmentation is available. By only mutating the intensity values but not the orientation or shape of structures in the image, the same reference segmentation can be used as ground truth. This tailored data augmentation strategy affects network training only. At inference time, in contrast to previous methods, no complex preprocessing of the image is required.

Furthermore, previous methods focused on segmenting the brain into the three main tissue classes: WM, cortical gray matter and ventricles. However, characteristics of other tissue classes, such as cerebellum (CB) and brain stem (BS), are important to understand and predict healthy or aberrant brain development in preterm infants of similar gestational age as fetuses [77]. The cerebellum is particularly of clinical interest as it is one of the fastest growing brain regions during the last trimester of pregnancy [78].

Another challenge for segmentation of the fetal brain in MRI is the large field of view of these scans. Since the fetus is scanned *in utero*, the images also visualize parts of the maternal and the fetal body, and not only the head of the fetus as would be the case in regular brain MRI. Similar to previous publications [12, 22], we therefore propose to first automatically segment the intracranial volume (ICV) of the fetus to identify the region of interest. A number of studies proposed segmentation of the ICV in fetal MRI [30, 41, 46, 79]. Following our previous work [46], we segment the ICV

directly in the entire image to fully automatically detect a region of interest.

The method we propose performs segmentation of fetal and brain tissues. The method first identifies the ICV from the fetal MRI slices using a convolutional neural network. Subsequently, the identified volume i S, cortical gray matter (cGM) and eCSF in contrast to previous methods which focused on WM, cGM and cerebrospinal fluid only.

The remainder of this paper is organized as follows: in Section 2 the data set used for the method development and evaluation is described, in Section 3 the method for fetal brain segmentation and the simulation of intensity inhomogeneity are described, in Section 4 the evaluation method is given. The performed experiments and their results are presented in Section 5, followed by a discussion of the method and the results in Section 6. Our conclusions are given in the final section.

3.2 Data

3.2.1 Fetal MRI dataset

This study includes T2-weighted MR scans of 12 fetuses (22.9–34.6 weeks post menstrual age). Images were acquired on a Philips Achieva 3T scanner at the University Medical Center (UMC) Utrecht, the Netherlands, using a turbo fast spin-echo sequence. Repetition time (TR) was set to 2793 ms, echo time (TE) was set to 180 ms and the flip angle to 110 degrees. The acquired voxel size was $1.25 \times 1.25 \times 2.5$ mm^3, the reconstructed voxel size was $0.7 \times 0.7 \times 1.25$ mm^3, and the reconstruction matrix was $512 \times 512 \times 80$. The images were not reconstructed using high resolution framework. We included images acquired in coronal plane as they appeared less affected by fetal motion in comparison with images acquired in axial or sagittal plane. The orientation of the fetal brain was determined using fast survey scanning. During the scan, the mother was lying on her left side to reduce the chance of inferior vena cava syndrome. The local ethical board approved the study and parental informed consent was obtained.

The reference standard was defined by manual annotation of all scans in 2D slices by a trained medical student. The brain was segmented into seven tissue classes: CB, BGT, vCSF, WM, BS, cGM and eCSF. Annotation was accomplished by manual pixelwise painting of the brain tissues in each coronal image slice using an in-house developed software. The labeling of each of the seven classes was indicated by a color overlay (Figure 3.3). The software allowed the user to zoom-in, zoom-out and scroll through the slices during the manual segmentation. The manual segmentation protocol was identical to the protocol described by Išgum et al. [22] for neonatal brain tissue segmentation. The ICV was defined as the union of all manually segmented tissue classes.

In total 15 slices (1.5%), 7 in the training set and 8 in the test set, were too distorted

by severe motion artifacts to be manually annotated. In total 126 of the remaining, manually annotated slices (26.2%), 32 in the training set and 94 in the test set, were identified as affected by intensity inhomogeneity that hampered manual annotation.

3.2.2 Neonatal MRI dataset

While we propose a method aimed specifically at segmentation of fetal MRI scans, the proposed segmentation approach and especially the data augmentation technique that simulates intensity inhomogeneity (detailed in the following section) might be useful for brain segmentations in MRI scans. Therefore, to evaluate the applicability of this technique to a different MRI data set, we additionally included nine brain MR scans of preterm born infants.

T2-weighted MR images were acquired on a Philips Achieva 3T scanner at the University Medical Center Utrecht, the Netherlands. The images were made with Turbo Field Echo (TFE) and Turbo Spin Echo (TSE) sequences with TR set to 4847 ms and TE set to 150 ms. The scans were acquired at 40 weeks of post menstrual age in the coronal plane. The acquired voxel size was $0.35 \times 0.35 \times 1.2$ mm^3 and the reconstruction matrix was $512 \times 512 \times 110$. The reference standard was defined by manual annotation of an expert into seven tissue types (CB, BGT, vCSF, WM, BS, cGM and eCSF) in MR scans without visible intensity inhomogeneity artifacts. Manual annotation was performed using the same protocol as described above for fetal MRI. These scans are part of the NeoBrainS12 segmentation challenge [80] and did not show intensity inhomogeneity artifacts.

The remaining four scans showed intensity inhomogeneity and they were used as a test set. However, manual reference segmentations in these scans were not available and, consequently, we evaluated the segmentation performance in this set only by visual inspection.

3.3 Method

To simplify the brain tissue segmentation and allow the segmentation method to focus on the fetal brain only, the fetal ICV is first automatically extracted. Subsequently, the identified ICV is automatically segmented into seven tissue classes. An overview of this pipeline is shown in Figure 3.4. The same network architecture, described in Section 3.3.1, was used for ICV extraction and brain tissue segmentation.

3.3.1 Brain segmentation

Extraction of ICV and its subsequent segmentation into seven brain tissue classes are achieved with two fully convolutional networks (FCN) with identical U-net architecture [52] trained with 2D slices. Each network is trained independently to perform its specific task. The U-net architecture consists of a contracting path and an expanding

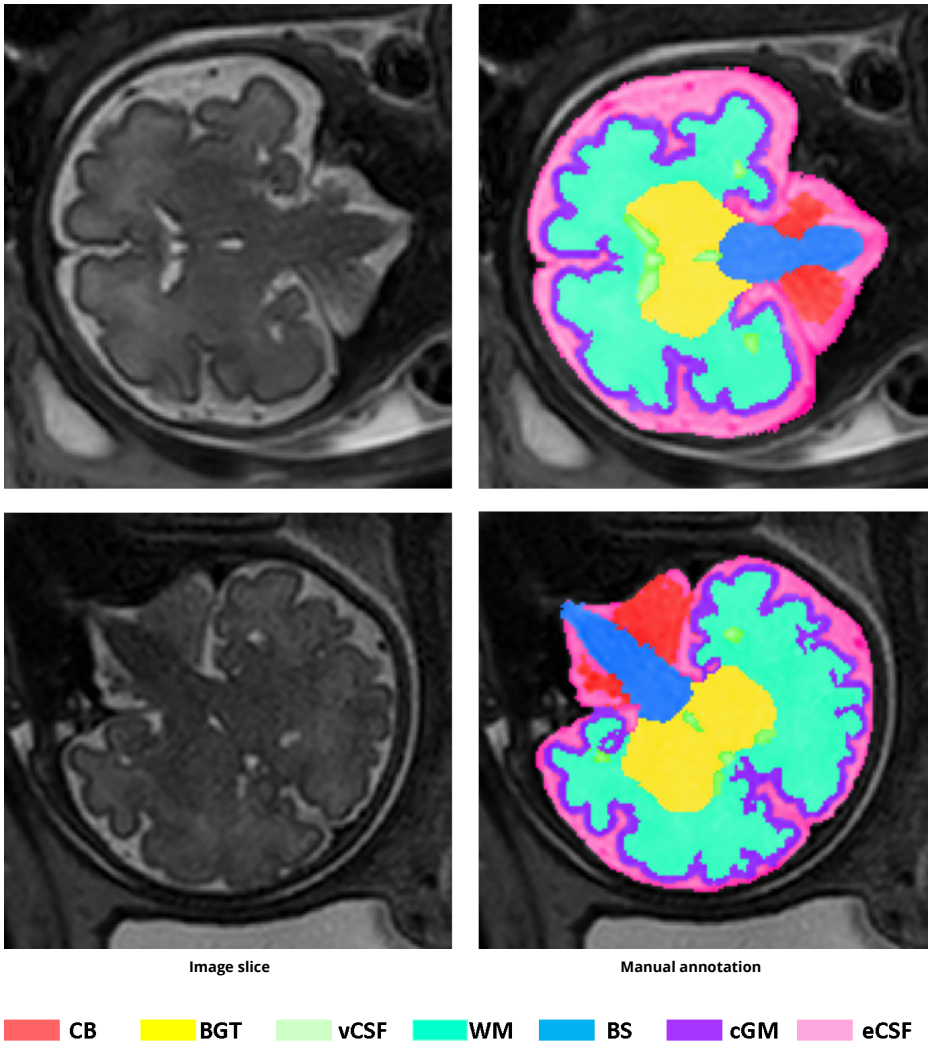


Figure 3.3: Examples of manual reference segmentation in coronal fetal MRI, showing slices cropped to the region of interest (first column) and the slices overlaid with the manual segmentations (second column).

path. The contracting path consists of repeated 3 × 3 convolutions followed by rectified linear units (ReLUs). A 2 × 2 max pooling downsamples the features. The number of feature channels doubles after every two convolutional layers. In the expansion path, an up-sampling is followed by a 2 × 2 convolution which halves the number of feature channels. The results are concatenated with the corresponding contraction path and convolved by two 3 × 3 convolutional layers followed by a ReLU. At the final layer,

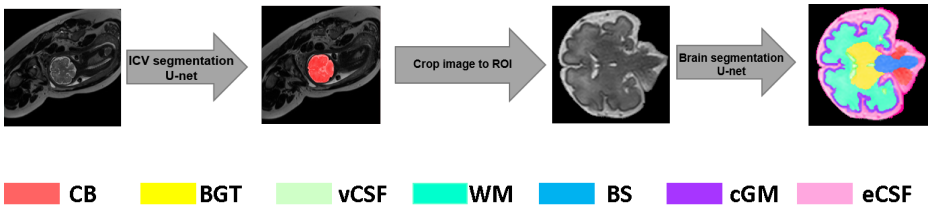


Figure 3.4: Proposed pipeline for automatic tissue segmentation method. The fetal ICV is automatically segmented in the original slice. Thereafter, the image is automatically cropped to the region of interest (ROI) so that the tissue segmentation can be restricted to the fetal ICV only.

one 1×1 convolutional layer maps each component of the feature vector to the desired number of classes. Batch normalization [61] is applied after all convolutional layers to allow for faster convergence. The network architecture is illustrated in Figure 7.2.

Both networks were trained using stochastic gradient descent with back propagation. We optimized both networks using the Adam optimizer with Nesterov momentum [59, 60] using a fixed learning rate of 0.0001. Standard data augmentation techniques, namely random flipping and rotation, were used during training to increase the variation of the training data and to mimic different orientations of the fetal brain. The slices were flipped in horizontal and vertical direction with 50% probability and were rotated with a rotation angle randomly chosen between 0 and 360 degrees.

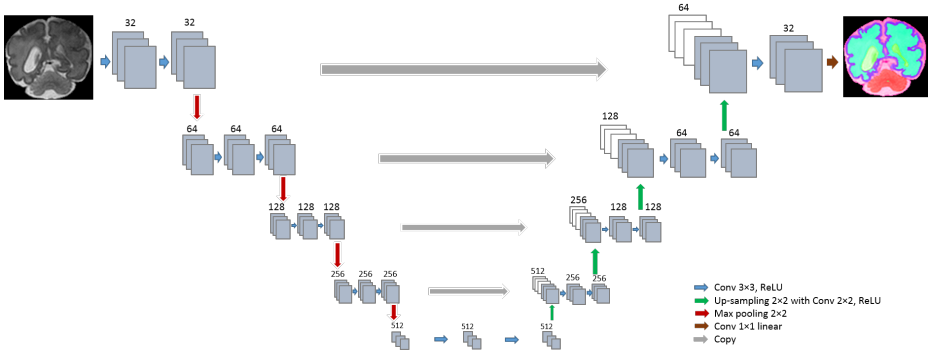


Figure 3.5: Network architecture: The network consists of a contraction path and an expansion path. The contraction path consists of repeated convolution layers followed by max pooling, and the expansion path consists of convolution layers followed by upsampling.

To first identify the intracranial area in the image, a U-net was trained by minimizing the cross-entropy between network output and manual segmentation for all pixels in each slice in each training batch. The convolutional neural network was trained

with batches of 12 slices in each iteration. Given that the network performs voxel classification, ICV segmentation may result in small isolated clusters of false positive voxels outside the ICV. These were removed by discarding 3D connected components smaller than 3 cm^3. This threshold was empirically chosen in this study. It was chosen large enough to remove small false positive clusters of voxels, and also small enough to prevent removing any parts of the brain that are not fully connected in 3D. The latter is often the case in scans with substantial motion artifacts where the signal is lost in one slice and consequently, the ICV segmentation is not fully connected in 3D.

The segmented intracranial fetal areas were further segmented into seven tissue classes using another, separately trained, U-net. Each pixel in the image was classified as either CB, BGT, vCSF, UWM, BS, cGM, eCSF or background. In contrast to the network for ICV segmentation, this network was trained by maximizing the Dice coefficient between network output and manual segmentation. This was done to achieve robustness against an imbalance of samples from the different classes. This network was trained with batches of 18 slices in each iteration. We have implemented the network in Keras, an open-source neural-network library written in Python [62].

3.3.2 Intensity inhomogeneity augmentation (IIA)

To make the network segmenting brain tissue classes robust to intensity inhomogeneity artifacts, we trained this network with slices containing simulated intensity inhomogeneity artifacts. The artifacts were simulated by applying a combination of linear gradients with random offsets and orientations to a slice without intensity inhomogeneity artifacts (I):

$$Z = I \times ((X + x_0)^2 + (Y + y_0)^2), \tag{3.1}$$

where X and Y are 2D matrices with integer values from zero to the size of the image in x and y direction, respectively. The offsets x_0 and y_0 control the balance between the x and y components and were randomly chosen from different ranges (x_0: [43, 187]; y_0: [-371, 170]). The optimal ranges were found with a random hyperparameter search. Additionally, the gradient patterns were randomly rotated between 0 to 360 degree to mimic intensity inhomogeneity in various directions. These random components, offsets and rotation, result in inhomogeneity patterns that allow the network to become invariant to the location and orientation of regions with low and decreasing contrast. The intensities in both the original slices as well as the slices with simulated intensity inhomogeneity were normalized to the range $[0, 1023]$ before feeding them to the network. Figure 3.6 shows examples of two slices from a fetal MRI scan with added synthetic intensity inhomogeneity.

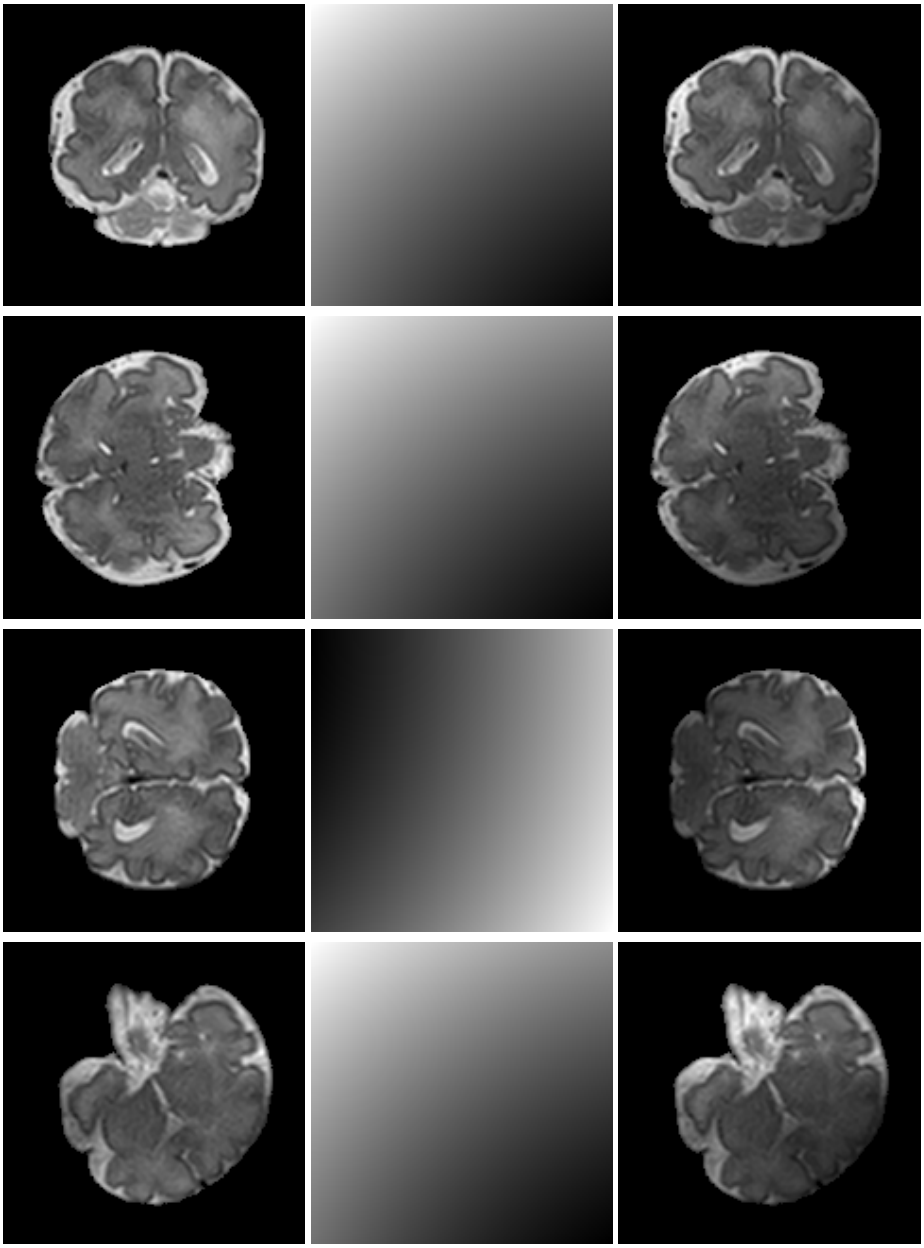


Figure 3.6: Example of coronal slices with simulated intensity inhomogeneity. Original slice (first column), simulated intensity inhomogeneity pattern (second column) and the image slice after adding the synthetic intensity inhomogeneity artifact (third column).

	Training set	Training set description
Set 1 (fetal)	6 scans	Excluding slices with intensity inhomogeneity artifacts
Set 2 (fetal)	6 scans	Including slices with intensity inhomogeneity artifacts
Set 3 (neonatal)	3 scans	Intensity inhomogeneity not visible in any slice

Table 3.1: The fetal MRI dataset was randomly divided into a training and a test set each containing 6 scans. We defined two training sets with fetal data: Set 1 contained 6 scans, but without those slices in which intensity inhomogeneity was visible. Set 2 contained the entire slices of 6 scans. The neonatal MRI dataset was divided into a training set with three scans and a test set with the six remaining scans (Set 3). The test set in all three sets contained 6 scans.

3.4 Evaluation

The automatic brain tissue segmentation was evaluated by means of the Dice coefficient (DC) for volume overlap and the mean surface distance (MSD) between manual reference segmentation and automatically obtained segmentation. In the fetal MRI scans, these metrics were calculated in 2D, i.e., per slice, and were then averaged across all slices. In the neonatal MRI scans, following previous work [12], these metrics were calculated in 3D.

3.5 Experiments and Results

In our experiments, we first evaluated the overall segmentation performance of the proposed pipeline with respect to the different tissue classes. To evaluate the influence of the proposed intensity inhomogeneity augmentation technique with the standard augmentation techniques, the segmentation performance before and after applying intensity inhomogeneity augmentation was evaluated. Furthermore, we evaluated whether this augmentation technique is able to generalize to different data, i.e., whether it leads to similar performance improvements in neonatal brain segmentation. The fetal MRI dataset was randomly divided into a training and a test set. Each set contained 6 scans. The neonatal MRI dataset was divided into a training set with 3 scans and a test set with the remaining 6 scans for which the manual reference segmentations of two scans were available. The training and test sets are listed in Table 3.1.

3.5.1 Segmentation performance

The performance of the proposed method using standard augmentation and the proposed IIA as described in Section 3.3 was evaluated. Slices with intensity inhomogeneity artifacts resulting from image acquisition were excluded from the training data. The average performance in the six test scans is listed in Table 3.2 for each of the seven tissue classes (Set 1). The average DC ranged from 0.80 for CB to 0.94 for eCSF and the average MSD ranged from 0.62 mm for CB to 0.18 mm for BS. Furthermore,

to evaluate whether IIA improves the performance differently in slices with artifacts than in artifact free slices, we compared the performance on slices with clearly visible intensity inhomogeneity artifacts and slices without visible artifacts. These results are also listed in Table 3.2. As shown, the automatic segmentations were less accurate on slices with intensity inhomogeneity strong enough to hamper manual annotation compared with slices without visible intensity inhomogeneity. Figure 3.7 and Figure 3.8 illustrate the segmentation performance on slices with intensity inhomogeneity and without visible intensity inhomogeneity, respectively. Note that these results were obtained with networks trained without any slices with intensity inhomogeneity artifacts resulting from the image acquisition, but slices with simulated intensity inhomogeneity were used for training. Excluding slices with intensity inhomogeneity from the training set is a more realistic training scenario because manual segmentation of such slices is cumbersome and reference segmentations for such slices might therefore not be available. However, extending the training set with image slices affected by real intensity inhomogeneity artifacts but in which manual annotation was still feasible could potentially further improve the performance. We therefore trained the networks also including slices with intensity inhomogeneity artifacts resulting from image acquisition (Set 2). The quantitative results are listed in Table 3.3. As shown in the table, in this experiment, we also separately evaluated the performance and impact of IIA in slices with visible and without visible intensity inhomogeneity artifacts.

Finally, we compared the performance of the proposed fetal brain tissue segmentation method with the performance of previous methods (Table 3.4). The performance of the proposed method was comparable to the performance of other methods, even though it performs a finer segmentation into seven tissue classes instead of only four [67] or three [71] tissues. The performance of previous methods is taken from the literature. Hence, the methods have been evaluated using different data set and thus this comparison can provide an indication only.

3.5.2 Comparison of data augmentation techniques

To evaluate the influence of the proposed IIA as well as the influence of the standard data augmentation techniques used in this study (random flipping and random rotation), the following experiments using fetal scans were performed. In this experiment using fetal scans, all slices with intensity inhomogeneity were removed from the training set (Set 1). First, a network was trained without any data augmentation to serve as baseline for the comparison. Second, a network was trained using random flipping of the training slices to augment the training data. Third, a network was trained using random flipping and random rotation of the training slices. Finally, as presented in Section 3.5.1, a network was trained using random flipping and random rotation, and additionally with IIA, i.e., randomly simulated intensity inhomogeneity. In this last experiment, all slices were manipulated with IIA during the training. Results of this

			CB	BGT	vCSF	WM	BS	cGM	eCSF	Mean
All test slices	IIA	DC	0.802	0.889	0.875	0.922	0.930	0.829	0.943	0.884
		MSD	0.620	0.414	0.470	0.384	0.181	0.318	0.188	0.368
	Without IIA	DC	0.688	0.807	0.724	0.849	0.850	0.672	0.820	0.773
		MSD	0.995	0.726	1.331	0.875	0.253	0.769	0.549	0.785
Slices with II	IIA	DC	0.694	0.901	0.807	0.899	0.947	0.782	0.877	0.844
		MSD	0.813	0.372	0.807	0.565	0.286	0.418	0.416	0.525
	Without IIA	DC	0.467	0.572	0.485	0.744	0.704	0.492	0.678	0.592
		MSD	1.724	1.340	3.742	1.656	0.629	1.275	1.000	1.624
Slices without II	IIA	DC	0.754	0.914	0.854	0.918	0.926	0.829	0.923	0.874
		MSD	0.428	0.406	0.729	0.426	0.159	0.351	0.241	0.392
	Without IIA	DC	0.719	0.837	0.802	0.902	0.933	0.758	0.870	0.832
		MSD	0.536	0.502	0.675	0.595	0.208	0.446	0.403	0.481

Table 3.2: Performance of fetal brain tissue segmentation into seven tissue classes when the network is trained on slices without intensity inhomogeneity resulting from image acquisition. The network is evaluated on the entire test set and additionally on only the slices with intensity inhomogeneity artifacts (94 slices) and on the slices without visible intensity inhomogeneity artifacts (357 slices). The segmentation performance with intensity inhomogeneity augmentation (IIA) used during training is compared with the performance of the same network without IIA. The results are expressed as the mean Dice coefficient (DC) and the mean surface distance (MSD) in mm.

			CB	BGT	vCSF	WM	BS	cGM	eCSF	Mean
All test slices	IIA	DC	0.794	0.931	0.874	0.919	0.946	0.835	0.944	0.892
		MSD	0.715	0.428	0.434	0.381	0.195	0.307	0.182	0.377
	Without IIA	DC	0.778	0.887	0.851	0.923	0.931	0.821	0.940	0.876
		MSD	0.771	0.501	0.521	0.440	0.193	0.356	0.195	0.425
Slices with II	IIA	DC	0.719	0.906	0.816	0.901	0.934	0.790	0.878	0.849
		MSD	0.735	0.426	0.703	0.613	0.278	0.398	0.413	0.509
	Without IIA	DC	0.722	0.877	0.842	0.905	0.941	0.801	0.885	0.853
		MSD	0.951	0.535	0.705	0.521	0.267	0.351	0.397	0.532
Slices without II	IIA	DC	0.749	0.905	0.884	0.927	0.926	0.856	0.934	0.883
		MSD	0.489	0.438	0.388	0.343	0.150	0.272	0.212	0.327
	Without IIA	DC	0.731	0.923	0.888	0.933	0.933	0.850	0.932	0.884
		MSD	0.520	0.402	0.363	0.360	0.141	0.290	0.215	0.327

Table 3.3: Performance of fetal brain tissue segmentation into seven tissue classes when the network is trained with slices with intensity inhomogeneity resulting from the image acquisition. The network is evaluated on the entire test set and additionally on only slices with intensity inhomogeneity (94 slices) and on the slices without intensity inhomogeneity (357). The segmentation performance with intensity inhomogeneity augmentation (IIA) in the training is compared with the performance of the same network without IIA in the training. The results are expressed as the mean Dice coefficient (DC) and the mean surface distance (MSD) in mm.

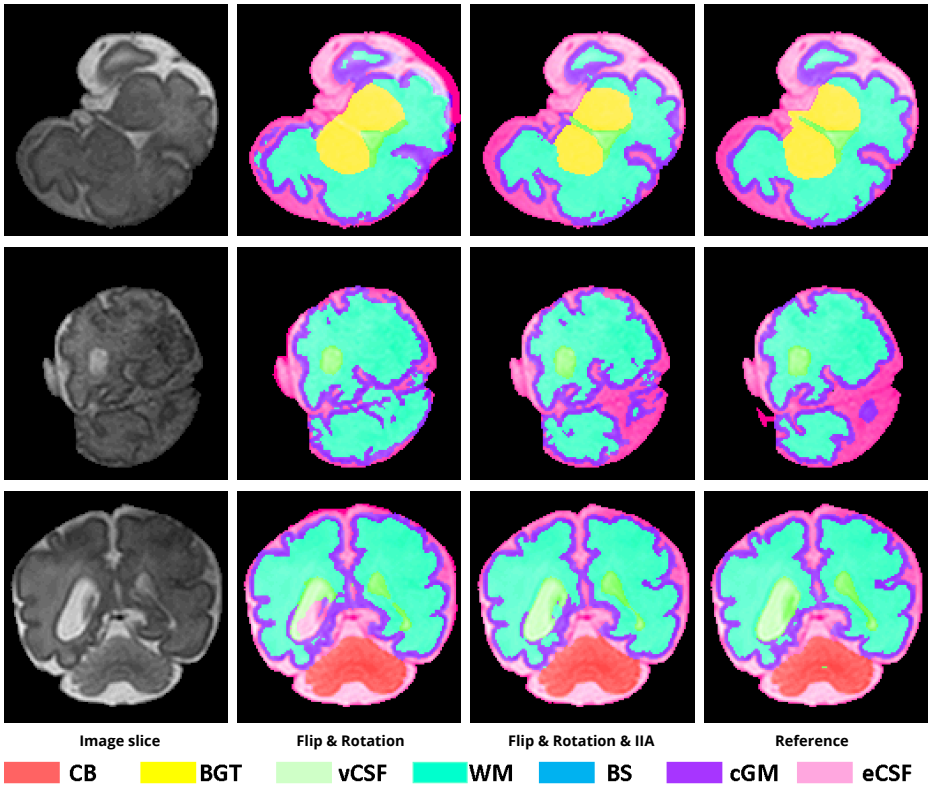


Figure 3.7: Examples of automatic brain tissue segmentation in slices with visible intensity inhomogeneity. A slice from T2-weighted fetal MRI scan with visible intensity inhomogeneity (first column); segmentation obtained with network only using flipping and rotation augmentation (second column); segmentation obtained with network using IIA (third column); manual reference segmentation (fourth column).

last experiment are listed in Table 3.2.

The achieved average DC and MSD for each scan in the test set are shown in Figure 3.9. The performance improved in all scans the more data augmentation was used and especially further improved when IIA was added in addition to the standard data augmentation techniques. IIA largely reduced performance differences between different scans, which standard augmentation techniques were not able to achieve. The performance improvement was particularly large in a scan in which nearly all slices showed intensity inhomogeneity (yellow marker in Figure 3.9). With IIA, the segmentation performance reached an accuracy comparable to that achieved on scans with fewer artifacts.

Additionally, the achieved average DC and MSD for each tissue class are shown in Figure 3.10. Overall, all augmentation methods improved segmentation performance.

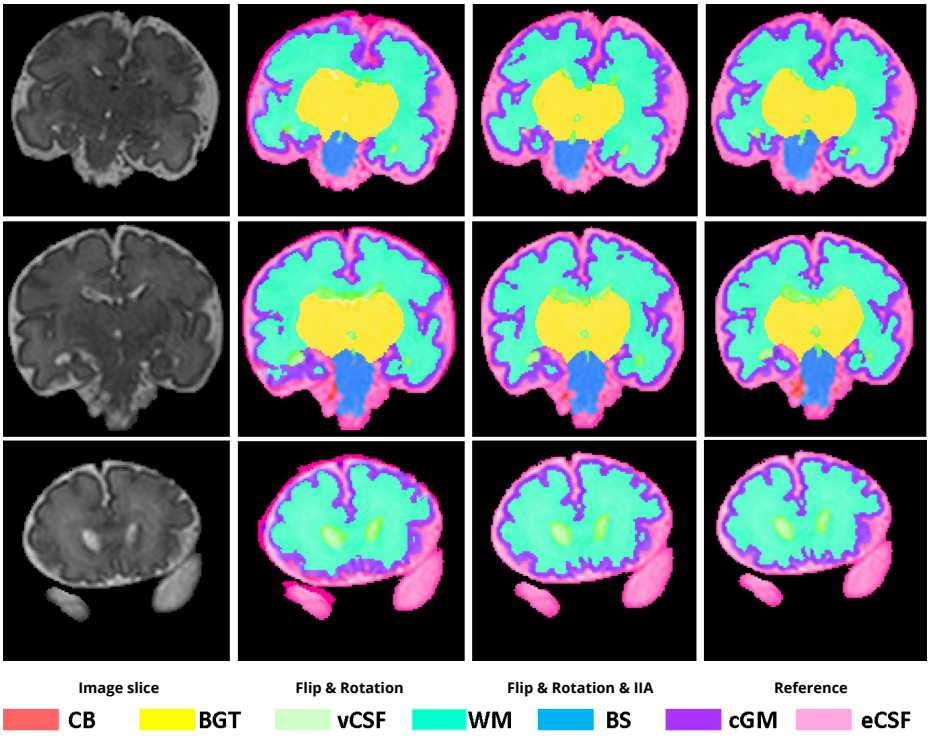


Figure 3.8: Examples of automatic brain tissue segmentation in fetal images acquired in the slices without visible intensity inhomogeneity. A slice from T2-weighted fetal MRI scan with visible intensity inhomogeneity (first column); segmentation obtained with network using only flipping and rotation augmentation (second column); segmentation obtained with network additionally using IIA (third column); manual reference segmentation (fourth column).

Adding augmentation based on random flipping and rotation of the image slices improved the segmentation of CB, BS, vCSF, BGT noticeably but not the segmentation of WM, cGM and eCSF. IIA was able to further improve the segmentation performance in all tissue classes including WM, cGM and eCSF. Overall, CB segmentation performance benefited the most from all augmentations where the DC increased from 0.3 to 0.8. Furthermore, augmentation based on random rotation severly influenced performance in segmentation of BGT, BS and CB.

3.5.3 Impact of extent of intensity inhomogeneity augmentation

To assess the impact of the proportion of slices in each training batch with simulated intensity inhomogeneity on the network performance, we trained the network with different percentage of artifact free slices to slices with simulated intensity inhomogeneity. In this experiment, all slices with intensity inhomogeneity were removed

	CB	BGT	vCSF	WM	BS	cGM	eCSF
Proposed Method	0.794	0.931	0.874	0.919	0.946	0.835	0.944
Hebas et al. [67]	-	-	0.900	0.900	-	0.820	-
Serag et al. [71]	-	-	0.920	0.900	-	0.840	-

Table 3.4: Segmentation performance of the proposed method and of other methods evaluated with Dice coefficient. Performance of previous methods is taken from the literature. Hence, this comparison can be used as indication only.

from the training set (Set 1). We varied number of slices with added synthetic intensity inhomogeneity from 0% to 100%. Figure 3.11 shows the obtained results. Training with only simulated slices led to slightly worse performance compared with a mix of original and manipulated slices. Even having only 20% of the slices with synthetic intensity inhomogeneity already improved the performance substantially. There was, however, no marked improvement for larger percentages of slices with synthetic intensity inhomogeneity, the performance was comparable for percentage of 20% to 100%.

3.5.4 Evaluation of IIA on neonatal brain segmentation

Intensity inhomogeneity is an artifact that occurs in various of MRI scans, albeit often to lesser extent than in fetal MRI. To asses whether the proposed augmentation technique-IIA-is also able to improve the performance of a segmentation task in other MR images, we trained the brain tissue segmentation network to perform segmentation in neonatal brain MRI scans. Following previous work [12] to limit the number of voxels considered in the classification, brain masks were generated with BET [23]. The network was trained with three scans and tested with the remaining two scans (Set 3). Like in the previous experiments, the network was first trained using standard data augmentation, i.e., using random flipping and rotation of the training slices as described in Section 3.3, and subsequently the network was trained additionally with IIA. The obtained segmentation results are listed in Table 3.5. As the two scans in the test set did not show any intensity inhomogeneity artifacts, we additionally qualitatively evaluated the segmentation performance on four scans with visible intensity inhomogeneity artifacts for which manual reference segmentations were not available. We illustrate segmentation results in the four scans without reference standard in Figure 3.12. In [12] these scans were not analyzed due to the presence of artifacts. Visual inspection of the results in these scans reveals that the segmentation was more accurate when IIA was used, particularly in BGT, BS, vCSF and eCSF.

3.6 Discussion

We presented a pipeline for automatic segmentation of the fetal brain into seven tissue classes in MRI. The method consists of two fully convolutional networks with identical

			CB	BGT	vCSF	WM	BS	cGM	eCSF	Mean
All test slices	With IIA	DC	0.857	0.883	0.776	0.825	0.765	0.500	0.600	0.744
		MSD	0.919	0.677	0.751	0.385	0.510	0.375	0.625	0.643
	Without IIA	DC	0.858	0.854	0.737	0.819	0.776	0.510	0.596	0.736
		MSD	1.465	1.359	0.617	0.393	1.383	0.372	0.635	0.889

Table 3.5: Performance of neonatal brain segmentation into seven tissue classes. The segmentation performance with IIA is compared with the performance of the same network without IIA. The results are expressed as the mean Dice coefficient (DC) and the mean surface distance (MSD) in mm.

U-net architectures. The first network extracts ICV and the second network performs segmentation of the brain into seven tissue classes. The results demonstrate that segmentation using the proposed data augmentation with simulated intensity inhomogeneity artifacts leads to accurate segmentations of the brain tissue classes. Moreover, we demonstrated that the method performs accurate segmentation while trained using manual reference segmentation only in slices without artifacts that occur during image acquisition. In other words, we showed that the proposed data augmentation is able to compensate for the lack of training data in which performing manual annotations is cumbersome.

Using the proposed data augmentation technique, we were able to achieve state-of-the-art segmentation performance with a substantially lower number of training scans. Our method was trained with only 6 fetal scans while previous methods used 20 [68] and up to 80 fetal scans [28]. Given that manual annotation of a fetal brain MR scan into 7 tissue classes requires about 40 hours, reducing the number of the manually annotated training scans substantially reduces the required manual annotation effort and associated costs.

Intensity inhomogeneity is a frequently occurring artifact in MRI and often hampers automatic image analysis due to diminishing contrast between different tissues. We demonstrated that the described method based on convolutional neural networks can become more robust to these artifacts by training with data augmented with simulated random intensity inhomogeneities. This can potentially replace or complement prepossessing steps, such as bias field corrections or volumetric reconstructions that would require acquisition of additional MR data. Simulating artifacts instead of manually annotating compromised data for training supervised methods is beneficial as manual reference annotations can be obtained more easily and with higher accuracy for artifact-free data.

Standardly used data augmentation techniques, random flipping and rotation of the image slices, improved the segmentation performance in CB, BS, BGT, and vCSF considerably but had little impact on other tissue types. Introducing training data with simulated intensity inhomogeneity further improved the segmentation performance in all tissue classes, including WM and cGM, which are often challenging to separate due to low inter-tissue contrast when intensity inhomogeneity artifacts are present.

Moreover, a frequent mistake of the automatic segmentation method when training without IIA was mistaking vCSF for eCSF and vice versa. Using IIA helped to overcome this issue in many cases, presumably by forcing the network not to focus on the intensity values only but additionally on other intensity invariant information such as shape and context.

Moreover, evaluating augmentation techniques per scan showed that the MSD reduces with adding random flipping and rotating augmentation in all scans and DC improves in all scans except one (indicated with a yellow marker in Figure 3.9). Retrospective visual inspection revealed that this scan has intensity inhomogeneity in nearly all slices. Even though the intensity inhomogeneity is not severe in all slices, the automatic segmentation was still severely affected. Experiments show that adding IIA in the training helped to overcome this issue and increased the segmentation performance in all scans.

Furthermore, our experiments illustrate that training the network with IIA increases the segmentation performance even in slices without visible intensity inhomogeneity. IIA makes the network more robust to intensity variations in MRI, forcing the network not to focus only on the tissue intensity for assigning a label.

In the current study, the segmentation method was evaluated on fetal brain MRI acquired in the coronal plane. Since the presented method is entirely supervised it can be readily applied to fetal MRI acquired in axial or sagittal plane if manually annotated training data is available.

We relied on the U-net architecture for ICV and brain tissue segmentation in fetal MRI. However, the proposed IIA can be used for data augmentation regardless of the network architecture or even with supervised methods not based on convolutional neural networks. We did not evaluate IIA with other architectures but it would likely improve the segmentation performance of other networks with different architecture as supervised CNNs regularly profit from large and diverse training data.

Additionally, we evaluated IIA on neonatal MRI. The visual inspection shows a substantial improvement on slices with artifacts. In the images without visible artifacts the quantitative results showed slight improvement when IIA is applied. The performance of the segmentation in the neonatal brain scans is lower than obtained by our previous method [12] as here presented network was not specifically adjusted for segmentation of neonatal brain tissues. However, the results clearly demonstrate the benefit of training with IIA. IIA could be readily applied to the segmentation method presented in [12] enabling the multi-scale CNN to segment neonatal brain scans with intensity inhomogeneity artifacts as shown in Figure 3.12.

In this study, 2D analysis was applied since fetal MR images has inter-slice motion due to the 2D MR acquisition. Additionally, in a few slices severe motion artifacts occurred. These slices were excluded from the training and test set as manually segmenting them for evaluation would be hardly feasible. Generating such slices that are heavily affected by motion artifacts in fetal MRI could be an interesting direction for

future work.

We have trained the proposed method with representative data, i.e. in a supervised manner. Training with non-representative, in addition to the representative data, using transfer learning would allow increasing the training sample size. This could be addressed in future work and it could potentially further improve segmentation performance.

3.7 Conclusion

We presented an automatic method for brain tissue segmentation in fetal MRI into seven tissue classes using convolutional neural networks. We demonstrated that the proposed method learns to cope with intensity inhomogeneity artifacts by augmenting the training data with synthesized intensity inhomogeneity artifacts. This can potentially replace or complement preprocessing steps, such as bias field corrections, and help to substantially improve the segmentation performance.

Acknowledgment

This study was sponsored by the Research Program Specialized Nutrition of the Utrecht Center for Food and Health, through a subsidy from the Dutch Ministry of Economic Affairs, the Utrecht Province and the Municipality of Utrecht. Furthermore, we thank Nienke Heuvelink for her help with creating the manual fetal brain segmentations.

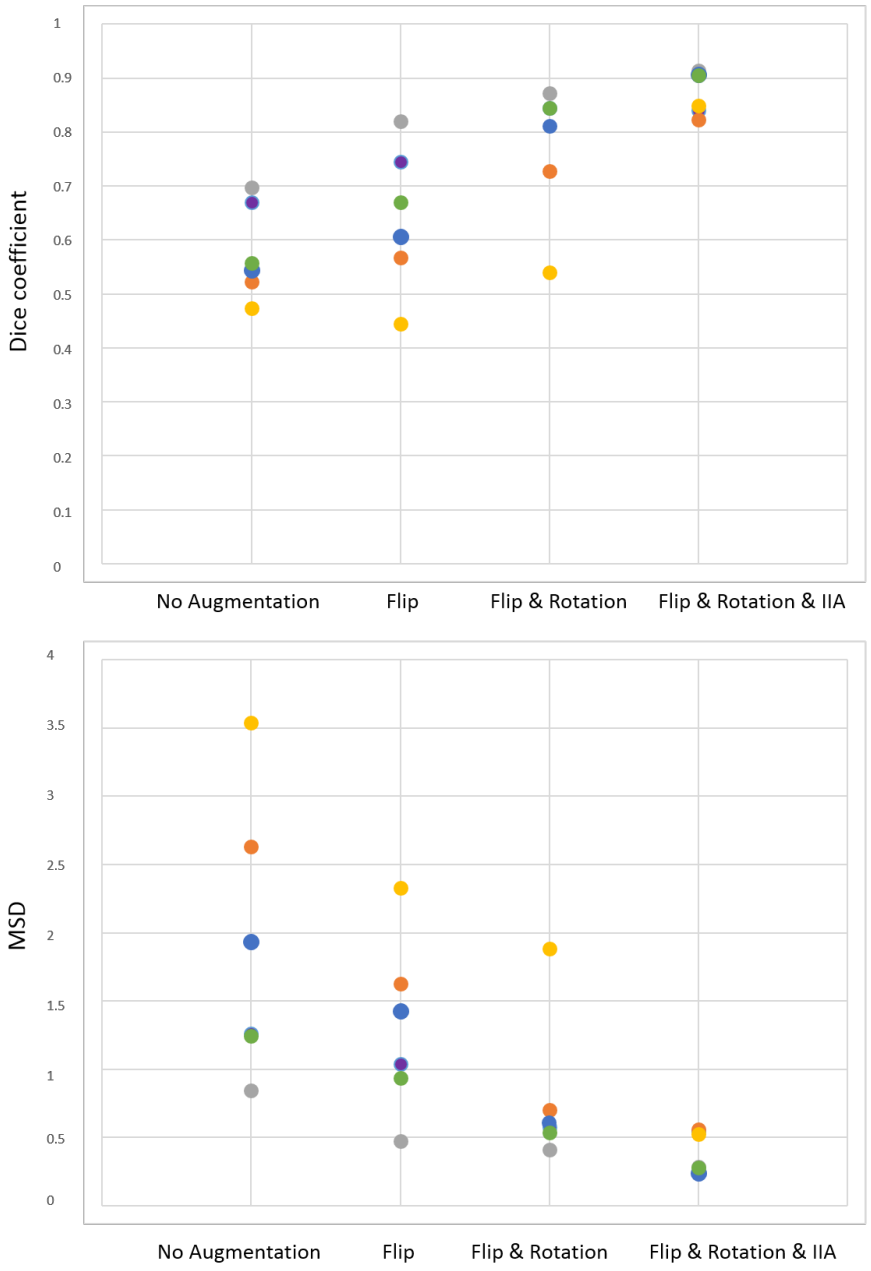


Figure 3.9: Performance of automatic fetal brain tissue segmentation into seven tissue classes when the network is trained without any augmentation, with flipping slices as augmentation, with flipping and rotating slices as augmentation, and with flipping, rotating and IIA. The results are expressed as the mean Dice coefficient and the mean surface distance (MSD) in mm. Each marker corresponds to the mean performance across all tissue classes in one of the six test scans.

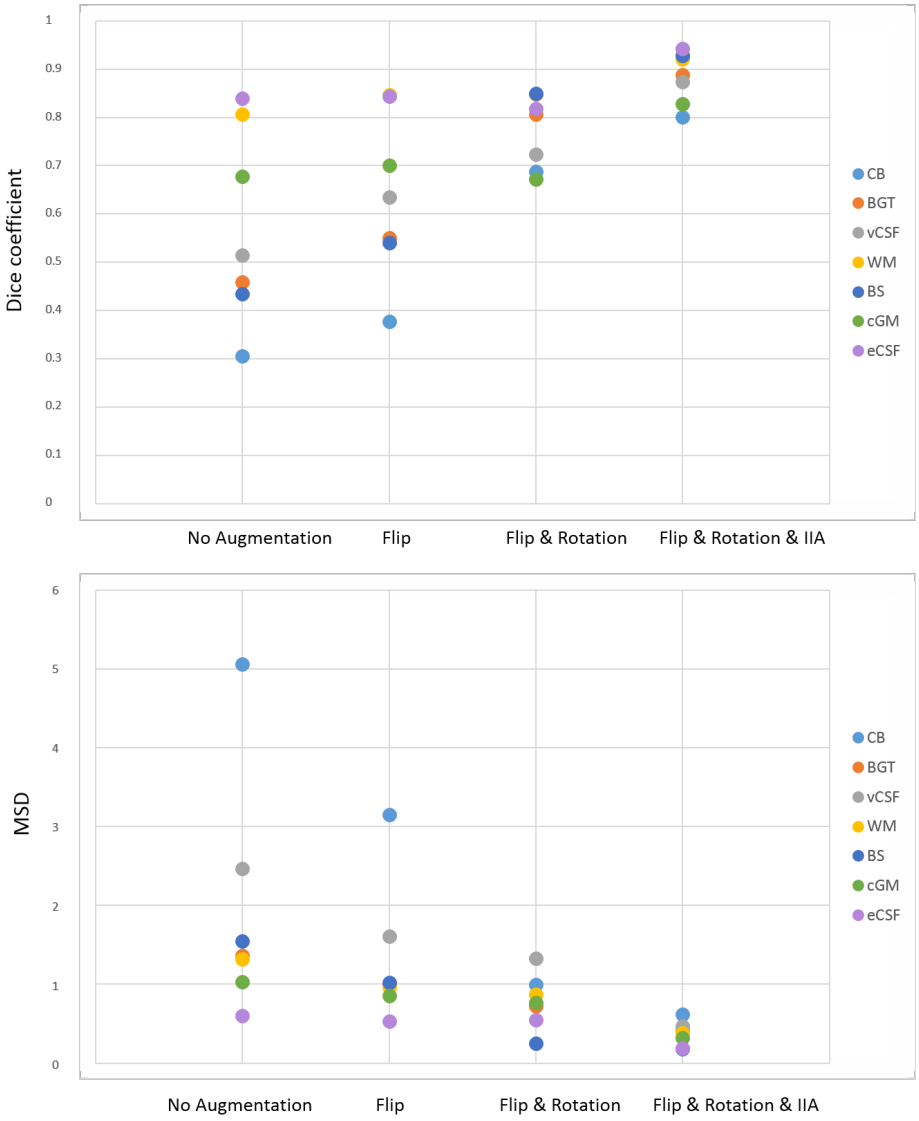


Figure 3.10: Performance of automatic fetal brain tissue segmentation for each of the seven tissue classes when the network is trained without any augmentation, with flipping augmentation, with flipping and rotating augmentation, and with flipping, rotating and IIA. The results are expressed as the mean Dice coefficient and the mean surface distance (MSD) in mm. Each marker corresponds to the mean performance across all six test scans for one of the seven tissue classes.

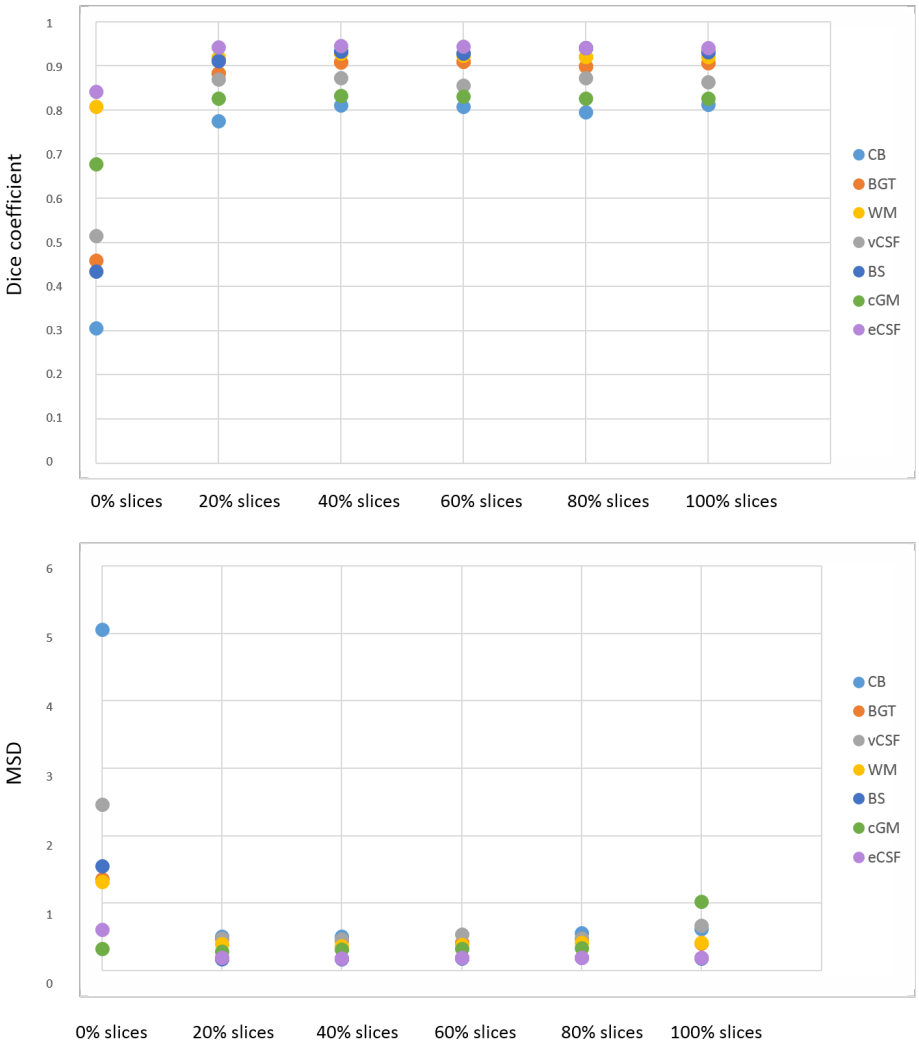


Figure 3.11: Performance of fetal brain tissue segmentation into seven tissue classes with different proportion of slices with synthetic intensity inhomogeneity in each training batch. The results are expressed as the mean Dice coefficient (DC) and the mean surface distance (MSD) in mm. Each marker corresponds to the mean performance across all six test scans for a different tissue class.

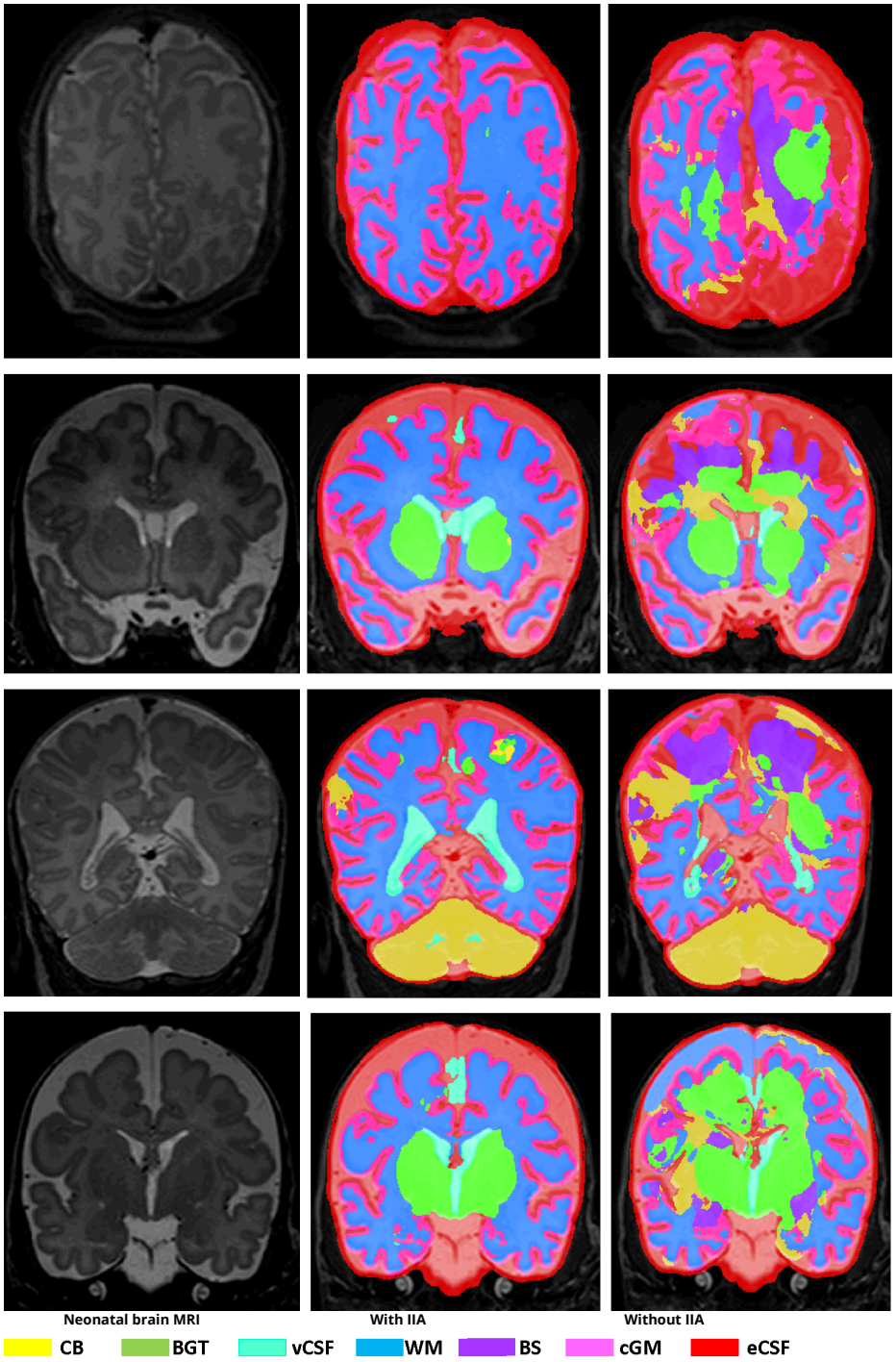
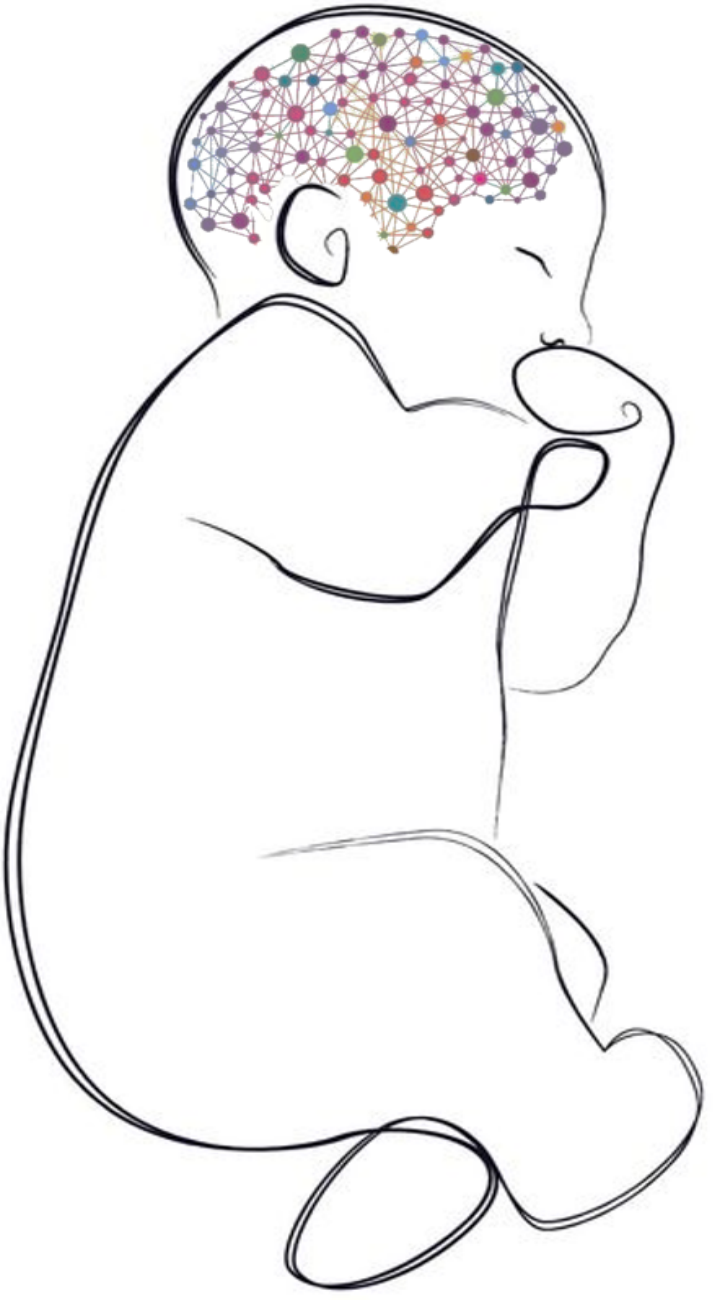


Figure 3.12: Examples of brain tissue segmentation in neonatal MRI with intensity inhomogeneity artifacts. A slice from T2-weighted fetal MRI scan (first column); segmentation obtained with network using rotation, flipping and IIA (second column); segmentation obtained with network only using flipping and rotation of the slices as training data augmentation (third column)



Chapter 4

GAN-based segmentation of motion affected neonatal brain MRI

BASED ON: N. Khalili, E. Turk, M. Zreik, M. A. Viergever, M. J. N. L. Benders, and I. Išgum. "Generative adversarial network for segmentation of motion affected neonatal brain MRI", *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019,* Cham: Springer International Publishing, 2019, pp. 320–328.

Abstract

Automatic neonatal brain tissue segmentation in preterm born infants is a prerequisite for evaluation of brain development. However, automatic segmentation is often hampered by motion artifacts caused by infant head movements during image acquisition. Methods have been developed to remove or minimize these artifacts during image reconstruction using frequency domain data. However, frequency domain might not always be available. Hence, in this study we propose a method for removing motion artifacts from the already reconstructed MR scans. The method employs a generative adversarial network trained with a cycle consistency loss to transform slices affected by motion into slices without motion artifacts, and vice versa. In the experiments 40 T2-weighted coronal MR scans of preterm born infants imaged at 30 weeks postmenstrual age were used. All images contained slices affected by motion artifacts hampering automatic tissue segmentation. To evaluate whether correction allows more accurate image segmentation, the images were segmented into 8 tissue classes: cerebellum, myelinated white matter, basal ganglia and thalami, ventricular cerebrospinal fluid, white matter, brain stem, cortical gray matter, and extracerebral cerebrospinal fluid. Images corrected for motion and corresponding segmentations were qualitatively evaluated using 5-point Likert scale. Before the correction of motion artifacts, median image quality and quality of corresponding automatic segmentations were assigned grade 2 (poor) and 3 (moderate), respectively. After correction of motion artifacts, both improved to grades 3 and 4, respectively. The results indicate that correction of motion artifacts in the image space using proposed approach allows accurate segmentation of brain tissue classes in slices affected by motion artifacts.

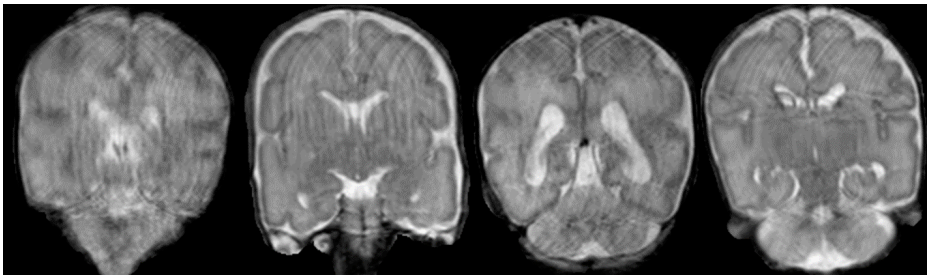


Figure 4.1: Examples of coronal slices from T2-weighted MRI acquired in preterm born infants at 30 weeks postmenstrual age affected by motion artifacts. Structures outside the neonatal cranium have been masked out.

4.1 Introduction

Important brain development occurs in the last trimester of pregnancy including brain growth, myelination, and cortical gyrification [1]. Magnetic resonance imaging (MRI) is widely used to non-invasively assess and monitor brain development in preterm infants. In spite of ability of MRI to visualize the neonatal brain, motion artifacts caused by the head movement lead to blurry image slices or slices with stripes (see Figure 4.1). These artifacts hamper image interpretation as well as brain tissue segmentation.

To enable the analysis of images affected by motion artifacts, most studies perform the correction in the frequency domain (k-space) prior to analysis [81, 82]. However, k-space is typically not stored and hence, not available after image reconstruction. Recently, Duffy at al. [83] and Paware et al. [84] proposed to use convolutional neural networks (CNNs) to correct motion-corrupted MRI from already reconstructed scans. CNNs were trained to reconstruct simulated motion artifacts that were modelled with a predefined formula. This enforces the network towards an assumed distribution of artifacts. However, in practice, it is difficult to estimate the real distribution motion. Alternatively, a CNN could be trained to generate an images without motion artifacts from images with such artifacts. However, this would require training with paired scans, which are rarely available. To solve this, recently cycleGAN has been proposed to train CNNs for image-to-image transformation with unpaired images [85].

In this study, we propose to employ a cycleGAN to generate MR slices without motion artifacts from slices affected by motion artifacts in a set of neonatal brain MR scans. The cycleGAN transform slices affected by motion artifacts into slices without artifacts, and vice versa. To demonstrate the effectiveness of the proposed method, we applied the motion correction to motion affected slices and subsequently, we propose to augment the segmentation training data from the cycleGAN that synthesizes slices with artifacts from slices without the artifacts. We demonstrate that the proposed correction for motion artifacts improves image quality and allows accurate automatic segmentation of brain tissue classes in brain MRI of infants. We also show that the

proposed data augmentation further improves segmentation results.

4.2 Data

This study includes 80 T2-weighted MRI scans of preterm born infants scanned at average age 30.7 ± 1.0 weeks postmenstrual age (PMA). Images were acquired on a Philips Achieva 3T scanner at University Medical Center Utrecht, the Netherlands. The acquired voxel size was 0.34×0.34 mm^2 and the reconstruction matrix was $384 \times 384 \times 50$. The scans were acquired in the coronal plane. In this data set, 60 scans had visible motion artifacts in most of slices and 20 scans had no visible motion in any slice. The reference segmentation of 10 scans out of 20 scans without motion artifacts were available. The scans were manually segmented into 8 tissue classes: cerebellum (CB), myelinated white matter (mWM), basal ganglia and thalami (BGT), ventricular cerebrospinal fluid (vCSF), white matter (uWM), brain stem (BS), cortical gray matter (cGM), and extracerebral cerebrospinal fluid (eCSF)

4.3 Method

Motion artifacts in the neonatal brain MR hamper the diagnostic interpretability and precise automatic segmentation of the brain tissue classes. To address this, we propose to correct motion artifacts in the reconstructed MR scans using a cycleGAN. Thereafter, to evaluate whether the corrected images are suitable for segmentation of brain tissues, a CNN architecture was trained to segment the brain into eight tissue classes. Furthermore, to improve segmentation performance, we augment the training data by synthesizing images with motion artifacts from the images without artifacts using the cycleGAN.

4.3.1 Artifact correction network

CycleGAN has been proposed to train image-to-image translation CNNs with unpaired images. Given that obtaining paired scans with and without motion artifacts is difficult, cycleGAN was employed to transform slices affected by motion to slices without motion artifacts and vice versa (Figure 4.2). The network architecture consists of two cycles, motion correction and motion generation cycle. The motion correction cycle consists of three networks. Motion correction network (MC) transforms slices affected by motion to slices without motion artifacts. Motion generation network (MG) reconstructs the generated slices without motion artifacts to the original image slices. A discriminator CNN discriminates between generated and real slices without motion artifacts Dis_{MC}. While the discriminator distinguishes between generated and real slices without motion artifacts, the generator tries to prevent it by generating images which are not distinguishable for the discriminator. Similarly, motion generation cycle

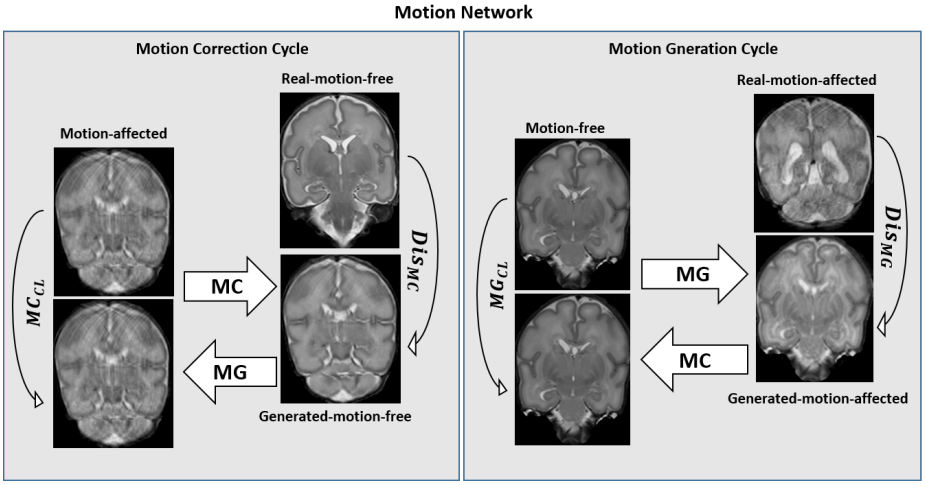


Figure 4.2: The CycleGAN consists of two cycles: motion correction and motion generation. In the motion correction cycle, first network is trained to transform slices affected by motion into slices without motion artifacts (MC), the second network is trained to transform the generated slices without motion artifact back to the original slices (MG), and the third network discriminates between real and synthesized slices without motion artifacts (Dis_{MC}). In the motion generation network, motion was added to the slices without motion artifacts (MG), motion correction network transforms generated slices to the original slices (MC), and the discriminator network discriminates between real and fake slices affected by motion artifacts (Dis_{MG})

transforms slices without motion artifacts to slices affected by motion. The network architecture in both cycles is identical. The generator contains 2 convolutions layers with stride of 2, 9 residual blocks [86], and 2 fractionally strided convolutions with stride proposed in [87]. The discriminator networks have a PatchGAN [88], which classifies 70 × 70 overlapping image patches as fake or real. Two adversial losses [89] were used in both motion correction network and motion generation network. Furthermore, cycle consistency loss in motion correction network (MC_{cl}) and motion generation network (MG_{cl}) were weighted by λ and were added to adversial losses.

4.3.2 Segmentation Network

To assess segmentation performance in images affected by motion artifacts, a CNN with Unet-like architecture was trained to segment images into eight tissue classes. The segmentation network consists of a contracting path and an expanding path. The contracting path consists of 10 3×3 convolution layers followed by rectified linear units (ReLUs). Every two conlution layers the features were downsampled by 2 × 2 max pooling and the feature channels were doubled in following scheme 32, 64, 128, 256, 512. In the expanding path, an up-sampling is followed by a 2 × 2 convolution which halves the number of feature channels. The results are concatenated with the corresponding

contracting path and convolved by two 3 × 3 convolutional layers followed by a ReLU. At the final layer, one 1 × 1 convolutional layer maps each component of the feature vector to the desired number of classes. Batch normalization is applied after all convolutional layers to allow for faster convergence. The network was trained with 3D patches of 256 × 256 × 3 voxels. The network was trained by minimizing the average of Dice coefficient in all classes between the network output and manual segmentation.

4.4 Evaluation

To quantitatively evaluate the proposed method, motion is synthesized in images using the motion generation network. Thereafter, the performance of the segmentation network was evaluated using the Dice coefficient (DC), Hausdorf distance (HD) and mean surface distance (MSD) to evaluate between manual reference and automatically obtained segmentations. The evaluation was performed in 3D.

To evaluate the proposed method on images with real motion artifacts, the images and the corresponding automatic segmentations before and after motion correction were qualitatively evaluated using 5-points Likert scale. The image quality was scored on a scale from 1 to 5, where 1 indicates uninterpretable images with severe motion artifacts, and 5 indicates excellent image quality. Similarly, automatic segmentations were scored 1 when the segmentation failed and 5 when the segmentation was very accurate.

4.5 Experiments and Results

Prior to analysis, the intracranial brain volume was extracted from all scans using BET [23]. To train the artifact correction network 15 scans without motion artifacts and 20 scans with motion artifacts were selected for training. The remaining 5 scans without motion artifacts and 40 scans with motion artifacts were used for testing. From scans without motion artifacts, 700 slices with no visible motion artifacts were selected. Similarly, from the scans with motion artifacts, 714 slices with visible motion artifacts were selected. The network was trained with a batch size of 4. Adam [59] was used to minimize the loss function for 100 epochs with a fixed learning rate of 0.00005. λ was set to 10.

To segment the brain into eight tissue classes, the segmentation network was employed. The network was trained with 5 scans without motion artifacts selected from the 15 training scans used to train the motion correction network. The segmentation network was trained with a batch size of 6. Adam was used to minimize the loss function for 200 epoch and the learning rate was set to 0.0001.

In the experiments, we performed quantitative evaluation of the proposed method through the evaluation of the brain tissue segmentation. First, to determine the upper limit of the segmentation performance, images without artifacts were segmented

		CB	mWM	BGT	vCSF	(u)WM	BS	cGM	eCSF	Mean
Original Scans	DC	0.90	0.53	0.89	0.84	0.94	0.84	0.67	0.83	0.80
	HD	44.92	32.97	39.06	23.08	17.25	42.57	18.47	8.60	28.36
	MSD	0.36	1.85	0.56	0.36	0.20	0.56	0.21	0.23	0.54
Motion Synthesized	DC	0.87	0.38	0.87	0.77	0.90	0.81	0.62	0.75	0.75
	HD	52.27	53.80	42.93	33.70	21.33	48.18	21.53	22.43	37.02
	MSD	0.62	4.10	1.04	1.32	0.77	0.92	0.55	1.00	1.29
Motion Corrected	DC	0.90	0.47	0.89	0.83	0.94	0.83	0.68	0.85	0.79
	HD	45.06	41.93	33.58	22.84	18.25	39.19	18.57	8.90	28.54
	MSD	0.46	2.07	0.55	0.35	0.20	0.41	0.21	0.16	0.55
Motion Augmented	DC	0.88	0.45	0.88	0.80	0.92	0.81	0.63	0.80	0.77
	HD	40.19	27.42	28.43	19.27	14.98	30.85	15.03	11.79	23.49
	MSD	0.46	1.84	0.61	0.39	0.27	0.48	0.27	0.24	0.57
Motion Corrected & Augmented	DC	0.91	0.48	0.89	0.84	0.94	0.84	0.67	0.84	0.80
	HD	45.62	34.52	26.83	17.77	14.40	35.93	17.18	7.63	24.99
	MSD	0.45	1.89	0.44	0.29	0.19	0.42	0.20	0.17	0.51

Table 4.1: Performance of brain tissue segmentation into eight tissue classes. The evaluation of segmentation was performed 1) on scans without motion artifact (original scans) 2) on the same scans with synthesized motion using motion generation network (motion synthesized) 3) on the scans that synthesized motion were corrected using motion correction network (motion corrected). The segmentation network was retrained with motion augmented scans using motion generation network. The evaluation of segmentation was performed 4) on the scans with synthesized motion using motion generation network (motion augmented) 5) on the scans that synthesized motion were corrected using motion correction network (motion corrected and augmented)

(Table 4.1, top row). Second, we aimed to evaluate performance in the images with artifacts. However, motion artifacts are prohibitive for accurate manual annotation and manual annotations were not available for such images. Hence, the motion generation network was used to synthesized from the images without artifacts, for which manual segmentations were available (Table 4.1, second row). Third, using motion correction network, the artifacts were removed from the images with synthesized artifacts and those were subsequently segmented (Table 4.1, third row). In the previous experiments, the segmentation network was trained only with images without motion artifacts, as only those were manually labelled. However, we hypothesized that the performance would improve when the segmentation would be trained with both types of images. Hence, to obtain images affected by motion that can be used for training, similar to the second experiment, we synthesized training images using motion generation network. In the fourth experiment, we evaluated segmentation network trained with augmented training data, i.e. images with and without motion artifacts on images with synthesized motion artifacts (Table 4.1, fourth row). Finally, images with synthesized artifacts were corrected as in the third experiment, and training data for the segmentation was augmented as in the fourth experiment (Table 4.1, bottom row). The results show that correction of motion artifacts using motion correction network improves the performance (Table 4.1, second vs. third row). Moreover, results demonstrate that the performance of the segmentation network improves when the training data is augmented (Table 4.1, second row vs fourth row and third vs. bottom row).

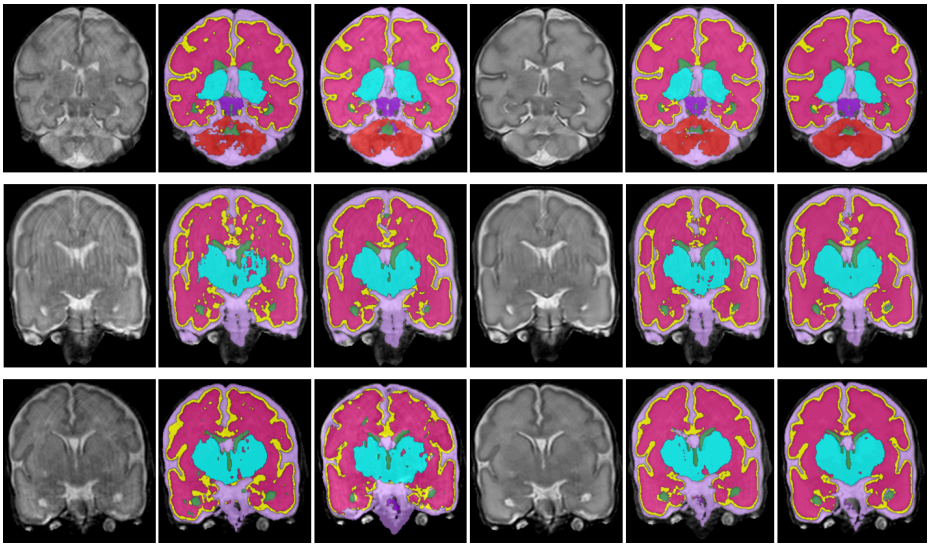


Figure 4.3: Examples of slices affected by motion artifacts and the corresponding tissue segmentation in neonatal MRI. 1st column: A motion affected slice; 2nd column: Automatic segmentation when the network was trained on slices without motion artifacts; 3rd column: Automatic segmentation, network trained on slices with augmented motion; 4th column: A motion corrected slice; 5th column: Automatic segmentation result on the corrected slice; 6th column: Automatic segmentation results on the corrected slice when the network was trained with data augmentation.

To qualitatively evaluate the performance of the motion correction network, 40 scans affected by motion artifacts were corrected using motion correction network. Subsequently, the segmentation network trained with the proposed data augmentation was used to segment the corrected images. The image and segmentation qualitative scoring before and after motion correction were performed. The median score of image quality and segmentation quality after motion correction increased from grade 2 to grade 3 and from grade 3 to grade 4 respectively. Furthermore, the first interquartile score of image quality and segmentation quality raised from 1 to 3 and 2 to 3 respectively. The third interquartile score of image quality and segmentation quality raised from 3 to 4 and 3.5 to 4 respectively. Figure 4.3 shows examples of images and corresponding segmentations before and after motion correction. This shows that the motion correction network reduces motion artifacts and hence, improves quality of the images and corresponding segmentations. Moreover, the figure shows that our proposed motion augmentation further improves automatic segmentations.

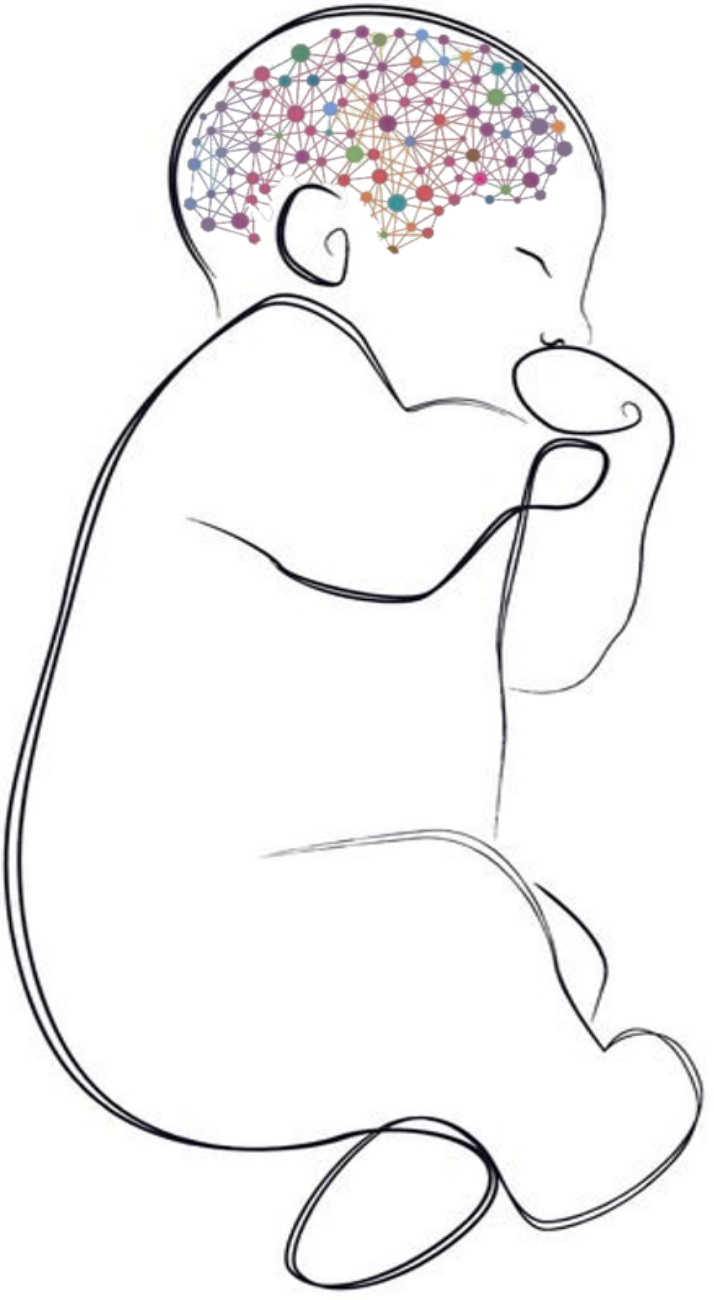
4.6 Discussion and conclusion

We presented a method to correct for correction of motion artifacts in reconstructed brain MR scans of preterm infants using a cycleGAN. We demonstrate that the pro-

posed artifact correction generates images that are more suitable for (automatic) image segmentation, the images were subsequently segmented into eight tissue classes. Additionally, we show that training the segmentation network with the proposed data augmentation further improves segmentation performance.

Unlike previous methods that performed motion correction in the frequency domain (k-space), the proposed method corrects motion artifacts in already reconstructed scans. Given that k-space is often not available after scans have been reconstructed and stored, the proposed method allows correction.

Quantitative results show that correction of motion artifacts and training the segmentation network with augmented data achieved best results. Furthermore, the qualitative evaluation shows that the proposed method improves image quality and segmentation in scans affected by motion.



Chapter 5

Brain and CSF volumes in fetuses and neonates with CHD

BASED ON: N. Claessens, N. Khalili, I. Išgum, H. Ter Heide, T. Steenhuis, E. Turk, N. Jansen, L. de Vries, J. Breur, R. de Heus, and M. Benders. "Brain and CSF volumes in fetuses and neonates with antenatal diagnosis of critical congenital heart disease: a longitudinal MRI study", *American Journal of Neuroradiology*, vol. 40 (2019), pp. 885–891.

Abstract

BACKGROUND AND PURPOSE: Fetuses and neonates with critical congenital heart disease are at risk of delayed brain development and neurodevelopmental impairments. Our aim was to investigate the association between fetal and neonatal brain volumes and neonatal brain injury in a longitudinally scanned cohort with an antenatal diagnosis of critical congenital heart disease and to relate fetal andneonatal brain volumes to postmenstrual age and type of congenital heart disease.

MATERIALS AND METHODS: This was a prospective, longitudinal study including 61 neonates with critical congenital heart disease undergoing surgery with cardiopulmonary bypass<30 days after birth and MR imaging of the brain; antenatally (33 weeks postmenstrualage), neonatal preoperatively (first week), and postoperatively (7 days postoperatively). Twenty-six had 3 MR imaging scans; 61 had at least1 fetal and/or neonatal MR imaging scan. Volumes (cubic centimeters) were calculated for total brain volume, unmyelinated white matter,cortical gray matter, cerebellum, extracerebral CSF, and ventricular CSF. MR images were reviewed for ischemic brain injury.

RESULTS: Total fetal brain volume, cortical gray matter, and unmyelinated white matter positively correlated with preoperative neonatal total brain volume, cortical gray matter, and unmyelinated white matter (r = 0.5–0.58); fetal ventricular CSF and extracerebral CSF correlated with neonatal ventricular CSF and extracerebral CSF (r = 0.64 and 0.82). Fetal cortical gray matter, unmyelinated white matter, and the cerebellum were negatively correlated with neonatal ischemic injury (r = −0.46 to −0.41); fetal extracerebral CSF and ventricular CSF were positively correlated with neonatal ischemic injury (r = 0.40 and 0.23).Unmyelinated white matter:total brain volume ratio decreased with increasing post-menstrual age, with a parallel increase of cortical gray matter:total brain volume and cerebellum:total brain volume. Fetal ventricular CSF:intracranial volume and extracerebral CSF:intracranial volume ratios decreased with increasing postmenstrual age; however,neonatal ventricular CSF:intracranial volume and extracerebral CSF:intracranial volume ratios increased with postmenstrual age.

CONCLUSIONS: This study reveals that fetal brain volumes relate to neonatal brain volumes in critical congenital heart disease, with anegative correlation between fetal brain volumes and neonatal ischemic injury. Fetal brain imaging has the potential to provide early neurologic biomarkers.

5.1 Introduction

Preoperative and postoperative ischemic brain injury is prevalent in neonates with critical congenital heart disease (CHD) undergoing open-heart surgery with the use of cardiopulmonary bypass, with an incidence over 50%.[90] White matter injury (WMI) and stroke are the most common types of ischemic brain injury, and especially WMI acquired in the neonatal period is associated with neurodevelopmental disabilities throughout school-age.[91, 92] In addition, delayed brain growth with suggested antenatal onset has been reported in individuals with critical CHD.[93, 94]

Antenatal studies have shown fetuses with critical CHD to demonstrate progressively reduced brain growth when compared to healthy fetuses.[95] Smaller brain volumes in fetuses with critical CHD are thought to be the result of reduced antenatal cerebral oxygen delivery and cerebral oxygen consumption, a consequence of the anatomical changes by the cardiac defect.[94, 96, 97] The extent in which delayed fetal brain growth is transient or persists in the neonatal period remains undescribed. Revealing the association between fetal and neonatal delayed brain growth and acquired brain injury provides important information for the potential use of fetal brain imaging as an early biomarker of brain abnormalities and the future implementation of interventions improving brain growth.

Using a prospective longitudinal design, we investigated the correlation between fetal brain and CSF volumes and neonatal (preoperative and postoperative) brain and CSF volumes and acquired neonatal (preoperative or postoperative) ischemic brain injury using MRI. Our secondary aim was to examine trajectories of brain and CSF volumes in relation to postmenstrual age (PMA) over the third trimester of gestation and first weeks after birth.

Figure 5.1: Fetal and neonatal brain tissue segmentation. Coronal T2-weighted segmented fetal and neonatal image of the same individual with transposition of the great arteries at 30 weeks of gestation and before the operation.

5.2 Materials and Methods

This was a prospective, longitudinal cohort study.

5.2.1 Study population

Between May 2016 and December 2017, fetuses with an antenatal diagnosis of critical CHD requiring open heart surgery with the use of cardiopulmonary bypass at 30 days of life and referred to the Wilhelmina Children's Hospital Utrecht, a tertiary level hospital, underwent longitudinal MR imaging of the brain at 3 time points, antenatally (around 33 weeks PMA), neonatal preoperatively (within the first week of life), and neonatal postoperatively (around 7 days after the operation; range, 5–10 days). For this study, we excluded individuals with a genetic syndrome confirmed by antenatal karyotype or microarray (such as trisomy 21 or 22q11 deletion) or cerebral congenital malformations (such as corpus callosum agenesis). The local medical ethics board approved the study, and parental informed consent for the use of clinical data for research purposes was obtained.

5.2.2 MR Imaging Acquisition

Fetal MRI was performed in the third trimester of gestation, around 33 weeks PMA (range 29-35 weeks). All fetuses were scanned in a 3.0 T MR system (Philips Medical Systems, Best, Netherlands). Mothers were positioned comfortably in the MR scanner and no sedation was given. The following sequences were used for this study: T2 weighted imaging in all three directions (TE/TR=180/55860ms, FOV=360x360, slice thickness=2mm), axial snap inversion recovery and axial diffusion weighted imaging. Neonatal MRI acquisition All fetuses and neonates were scanned on a 3T MR imaging system (Achieva, Philips Healthcare, Best, the Netherlands). Fetal and neonatal MR images were reviewed by 2 independent researchers for the presence of congenital anomalies, parenchymal hemorrhage, ischemic brain injury, and ventriculomegaly (atrial diameter, 10 mm).[98, 99] Details are provided in the On-line Appendix.

5.2.3 Quantitative 3D volumetric analysis

Fetal and neonatal coronal T2-weighted images were automatically segmented into different tissue classes: extracerebral CSF (ECSF), ventricular CSF (VCSF), unmyelinated white matter (UWM), cortical gray matter (CGM), cerebellum, basal ganglia and thalamus, brain stem, and myelinated white matter (5.1), using a further adjusted approach for neonatal MR images.[13] Volumes were calculated for all 8 tissue classes. All segmentations were manually checked for quality; no manual adjustments to the segmentations were made. Because basal ganglia, brain stem, and myelinated white matter were difficult to distinguish reliably on fetal T2-weighted images, these tissues were not taken into account for this study. Intracranial volume (ICV) was calculated as the sum of all 8 tissues classes; total brain volume (TBV), as the sum of UWM, CGM, cerebellum, basal ganglia and thalamus, brain stem, and myelinated white matter.

5.2.4 Statistical analysis

For statistical analysis, R statistical and computing software, Version 3.5.0 (http://www.r-project.org/) was used. Clinical variables were predominantly not normally distributed; therefore, nonparametric tests were performed and medians (25th/75th centiles) were presented for continuous data. Counts (percentage) were presented for categoric data. To test the differences in baseline characteristics between individuals with and without fetal MR imaging, we used Mann-Whitney U and Fisher exact tests. Univariable linear and nonlinear quadratic regression analysis was performed to examine the association between fetal brain and CSF volumes and neonatal (preoperative and postoperative) brain, and between CSF volumes and acquired neonatal brain injury, with correction for PMA at the time of scanning. Correlation coefficients with 95% confidence intervals were presented. Mixed-model analysis (R, nlme package 3.1–137) was performed to test the difference in brain and CSF volumes between different types of CHDs (fixed effect), including PMA (fixed effect), with the brain or CSF volume as dependent factors and individual subjects as random factors. A P value .007 was considered significant with correction for multiple comparisons.

5.3 Results

5.3.1 Study population

In the study period, 71 neonates with an antenatal diagnosis of critical CHD were born, of whom 8 died without any MR imaging of the brain. Sixty-three neonates with an antenatal diagnosis of critical CHD had cardiac surgery within 30 days after birth, of whom, 2 were excluded because of a genetic syndrome, leaving 61 neonates eligible for this study. Thirty-one neonates underwent all 3 MR imaging scans;however, 5 had poor fetal imaging quality, leaving 26 with all 3 MR imaging scans for the primary study aim. All 61 neonates were included for the secondary study aim:26plus an additional 21 with 2 MR imaging scans and 14 with 1 MR imaging scan (flowchart in On-line Figure). Baseline characteristics are presented in 5.1.

5.3.2 Fetal and Neonatal MR Imaging: Conventional Analysis

Two fetuses showed unilateral ventriculomegaly (atrial diameter, 10–15 mm). None of the fetuses showed parenchymal hemorrhage or ischemic brain injury.

At preoperative neonatal MR imaging, focal infarction was seen in 2 (8%) and WM injury (WMI) in 2 (8%). Punctate cerebellar hemorrhage was seen in 1 (4%). Cumulatively (at preoperative and post operative MR imaging),focal in farction was present in 7 (28%) and WMI in 9 (38%). Hypoxic-ischemic watershed injury was not seen in any neonates.

Characteristics	
Maternal characteristics(n = 29; fetal MRI)	
Maternal age (yr)	30.5 (27.8/35.3)
Maternal body mass index	22.8 (20.8/23.9)
Maternal smoking	3 (12%)
Placental weight (g)	455 (375/510)
Patient characteristics (n = 61; totalstudy population)	
Female	18 (30)
CHD groups	
SVP	21 (34)
TGA	18 (30)
LVOTO	22 (36)
Antenatal CHD diagnosis	61 (100)
Twin	1 (2)
Genetic disorder	0 (0)
Additional congenital anomaly	3 (5)
Gestational age (wk)	39.1 (38.4/40.1)
Birth weight (g)	3320 (2970/3670)
Birth weight (zscore)	-0.4 (-0.8/0.3)
Cesarean delivery1	8 (31)
Neonatal death	6 (10)
MRI characteristics	
Fetal, PMA (wks)	33.4 (32.7/34.1)
Preoperative, days after birth	5 (3/6)
Preoperative, PMA (wks)	40.1 (39.1/41.0)
Postoperative, days after surgery	8 (7/9)
Postoperative, PMA (wks)	42.0 (40.6/43.0)

Table 5.1: Baseline and MRI characteristics of the study population. Median with interquartile range or number with percentage are shown. No significant differences were seen between the primary and secondary study population.

		Neonatal		
		Preoperative volume	Postoperative volume	Ischemic brain injury
Fetal volumes	TBV*	●	●	○
	UWM*	●		□
	CGM*	●	■	□
	CB*	■		□
	UWM:TBV†		●	■
	CGM:TBV†	■		
	CB:TBV†	●	●	■
	TBV:ICV†	●	■	
	VCSF†	●		
	ECSF†	●	●	■
	VCSF:ICV†	■	●	
	ECSF:ICV†	●	■	

■Moderate correlation (closed square r = 0.3 to 0.5; open square r = -0.5 to -0.3)

● Strong correlation (closed circle r = 0.5 to 0.8; open circle r= -0.8 to -0.5)

Table 5.2: Association between fetal and neonatal MRI. Strength of correlation of fetal and neonatal preoperative volumes (first column); fetal and postoperative neonatal volumes (second column); fetal volumes and neonatal ischemic brain injury (third column). *Result of linear regression analysis with correction for postmenstrual age at scan. †Result of quadratic regression analysis.

5.3.3 Fetal Brain and CSF Volumes: Clinical Factors

Fetal UWM was positively associated with fetal CGM (r 0.87); fetal UWM, with fetal cerebellum (r 0.60); and fetal CGM, with fetal cerebellum (r 0.56). Fetal UWM:TBV (proportion of TBV occupied by UWM) was negatively associated with fetal CGM: TBV (r 0.51). Fetal VCSF and ECSF were negatively associated with fetal UWM:TBV (r0.39 and 0.53, respectively).

Fetal brain and CSF volumes (corrected for PMA at scanning) were not associated with maternal body mass index, maternal age, maternal smoking, placental weight, or birth weight z score (all P values .007 with Bonferroni correction for multiple comparisons). None of the mothers of the primary study population had preeclampsia or gestational diabetes.

5.3.4 Brain and CSF Volumes: Correlation between Fetal and Neonatal MR Imaging

Fetal volumes of CGM, UWM, and TBV correlated with preoperative neonatal volumes of CGM, UWM, and TBV, respectively(5.2). Postoperatively, the association of fetal CGM and TBV with neonatal CGM and TBV, respectively, was still seen. Fetal VCSF, ECSF, VCSF: ICV, and ECSF:ICV correlated with preoperative neonatal VCSF, ECSF, VCSF:ICV, and ECSF:ICV, respectively. Postoperatively, the associations between fetal and neonatal MR imaging were still seen for ECSF and VCSF: ICV (5.2 and On-line Table).

5.3.5 Fetal Brain and CSF Volumes: Correlation with Neonatal Ischemic Brain Injury

Baseline characteristics were not different between neonates with and without ischemic brain injury (data not shown). Both fetuses with antenatal ventriculomegaly had moderate-severe WMI before the operation (100%).

The fetal TBV, CGM, UWM, and cerebellum were negatively associated with the presence of neonatal (either preoperatively or postoperatively acquired) ischemic brain injury (5.2), in which fetal ECSF was positively associated with neonatal ischemic brain injury. The rate of preoperative ischemic brain injury was low and consequently was not examined separately in relation to fetal brain and CSF volumes.

5.3.6 Fetal and Neonatal MRI: brain volumes, PMA and type of CHD

For the secondary study aim, the total study population of 61 subjects was analyzed. ICV, TBV, UWM, CGM, and the cerebellum all showed a linear association with PMA (5.2). The proportion of TBV occupied by UWM decreased with time, with a parallel increase of occupation of the TBV by CGM and the cerebellum. The proportions

of VCSF: ICV and ECSF:ICV decreased during the fetal period with increasing PMA; however, it increased again in the neonatal period (5.3).

By means of mixed-model analysis, no differences in brain or CSF volumes were seen among CHD groups (all F-test P values .007, with correction for multiple comparisons).

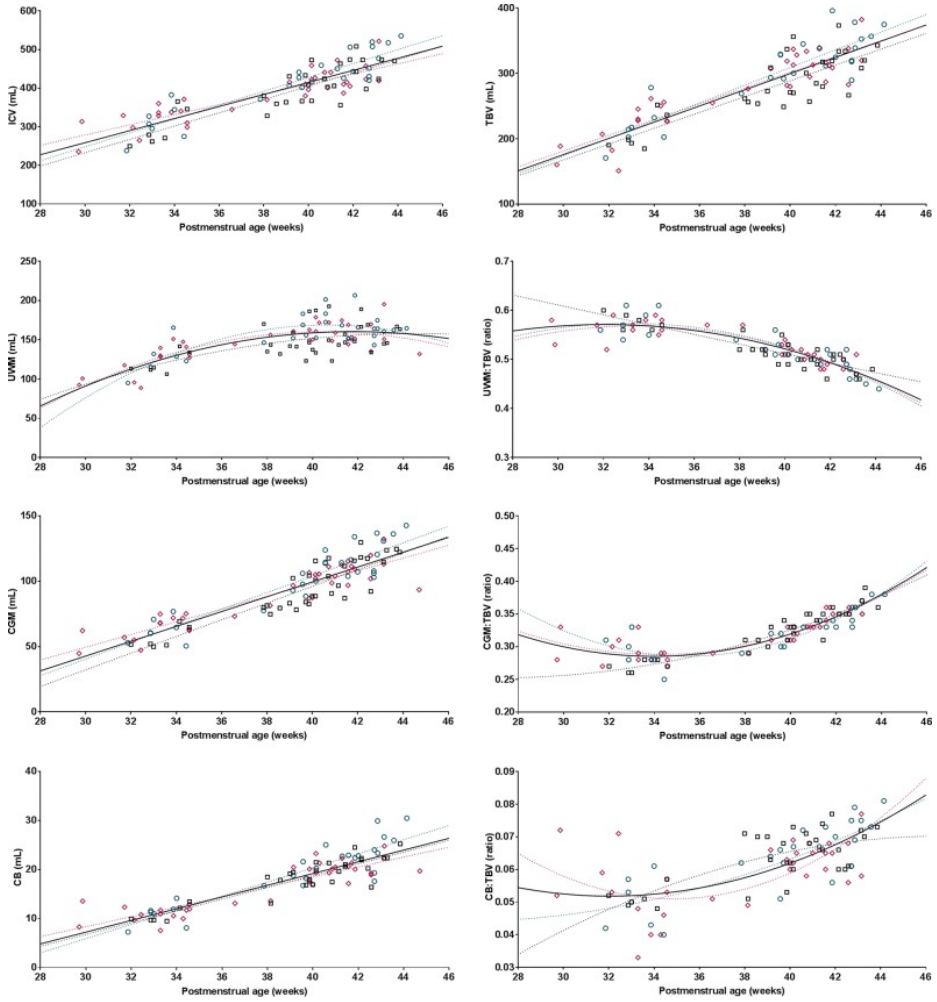


Figure 5.2: Fetal and neonatal brain volumes in relation to postmenstrual age. Plotted are brain volumes (y-axis) in relation to postmenstrual age atscanning (x-axis) for transposition of the great arteries (pink diamonds), left ventricle outflow tract obstruction (green circles), and single-ventricle physiology (black squares).Either the regression or quadratic line was fitted for each congenital heart defect separately (dotted lines),also as 1 line combining all subjects (solid black line). R^2 values reflecting the goodness of fit: ICV, 0.76; TBV, 0.76; UWM, 0.53; CGM, 0.81; cerebellum, 0.79; CGM:TBV, 0.78; UWM:TBV, 0.77; cerebellum:TBV, 0.48.

5.4 Discussion

Previous studies have reported smaller brain volumes in fetuses and neonates with critical CHD compared with healthy subjects. Linking longitudinal fetal and neonatal brain and CSF volumes, for the first time, this study reveals the strong relationship between brain volumes and postmenstrual age in individuals with critical CHD. Additionally,the results show a negative correlation between fetal brain volumes and neonatal ischemic brain injury. Revealing the association between fetal brain imag-

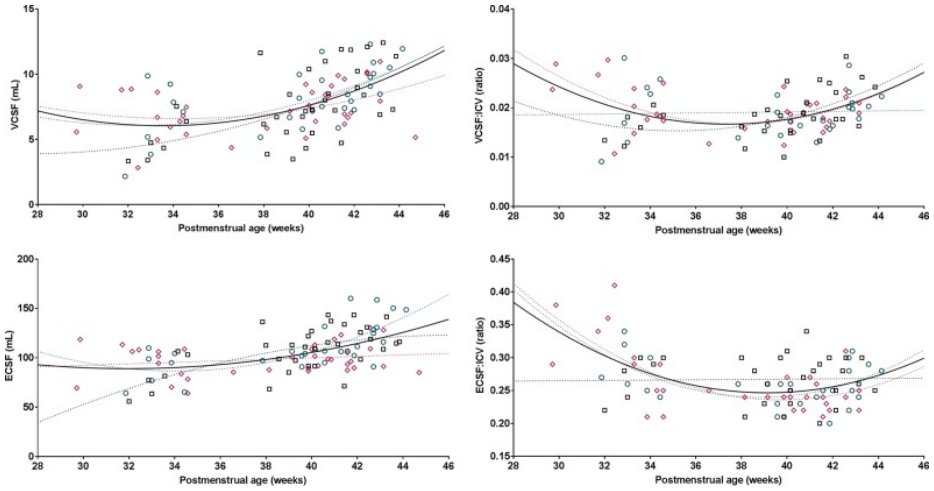


Figure 5.3: Fetal and neonatal CSF volumes in relation to postmenstrual age. Plotted are CSF volumes (y-axis) in relation to postmenstrual age atscanning (x-axis) for transposition of the great arteries (pink diamonds), left ventricle outflow tract obstruction (green circles), and single-ventricle physiology (black squares). Either the regression or quadratic line was fitted separately for each congenital heart defect (dotted lines),also as 1 line for all subjects (solid black line). R^2 values reflecting the goodness of fit: VCSF, 0.27; ECSF, 0.27; VCSF:ICV, 0.14; ECSF:ICV, 0.24.

ing parameters and neonatal brain abnormalities provides important information for the potential use of fetal brain imaging parameters to estimate neonatal neurologic findings in critical CHD.

In neonates with critical CHD, volumes of the largest brain structures occupying most of the total brain volume, ie, UWM, CGM, and cerebellum, correlated with the volumes measured in the same individuals around 33 weeks of gestation. The findings of this study support the hypothesis that neonatal brain volumes are the result of antenatal brain growth and strengthen the possibility of detecting delayed brain growth at an early stage, even before birth. Inourstudy, the association of fetal UWM and CGM was stronger with preoperative neonatal UWM and CGM than with postoperative UWM and CGM. This finding suggests that growth of UWM and CGM after birth is dependent on other neonatal factors than antenatal brain growth alone.

In our study, the association between fetal and neonatal age was less strong for the cerebellum than it was for UWM and CGM. However, the proportion of fetal TBV occupied by the cerebellum was strongly associated with the proportion of preoperative and postoperative neonatal TBV occupied by the cerebellum. The cerebellum is the brain structure with the highest increase in volume between 30 and 40 weeks PMA in extremely preterm infants.[100] Previous studies have shown that in the third trimester of gestational age, cerebellar volume was comparable between fetuses with and without CHD[101, 102] whereas at term age, a reduction of 20% in the cerebellum has been observed in neonates with critical CHD.[93]

Together with the relative reduction in brain volumes with increasing gestation, a relative increase in CSF spaces has been described in fetuses with critical CHD compared with healthy fetuses.[95] Larger CSF spaces are seen as an expression of brain immaturity.[103] In healthy fetuses,CSF volumes plateaulate in the third trimester,[104] and the proportion of ICV occupied by CSF decreases during the third trimester from 0.4 to 0.1.6 In our study, the proportion of ICV occupied by VCSF and ECSF was higher than 0.1 throughout the antenatal and neonatal period for all neonates with critical CHD. Absolute and proportional CSF volumes did decrease during the third trimester of gestation, to increase again in the neonatal period. In addition, the findings of our study revealed that increased absolute and proportional CSF volumes at fetal age did correlate with increased neonatal CSF volumes. This increase in CSF spaces seems abnormal and might be a sign of further brain underdevelopment or the result of brain tissue loss. Studies in older children with single-ventricle physiology have shown an association between larger CSF spaces and poorer neurodevelopment.[105]

Although it is becoming well-established that WM, CGM, and cerebellar volumes are smaller in fetuses and neonates with various types of critical CHD compared with healthy controls,[19, 93, 95, 101, 102, 106] the contribution of these findings to neurodevelopment later in childhood remains largely unknown. Smaller neonatal brain volumes are associated with abnormal neonatal neurobehavior [107] and lower 6-year intelligence in children with critical CHD.[92] In adolescents with critical CHD, decreased TBV is strongly correlated with poorer cognitive and executive functioning.[108]

In our study, ischemic and hemorrhagic brain injury was not seen in any of the fetuses. This finding supports the common thought that acquired brain injury has its onset after birth, in the neonatal intensive care and surgical period. The rate of preoperative and new postoperative neonatal ischemic injury in our population is comparable with that in previous reports.[90] All individuals included in this study had antenatal CHD diagnosis, which has been suggested as a protective factor for preoperative neonatal ischemic brain injury.[91] In neonates with critical CHD, WMI and focal infarction are the most common forms of acquired brain injury, and especially WMI has been shown to be associated with poorer motor and cognitive outcomes in the long term.[91, 92]

The results of this study revealed a correlation of smaller fetal TBV, UWM, CGM, and cerebellum with acquired neonatal ischemic brain injury (ie, moderate-severe WMI or focal infarction) in neonates with an antenatal diagnosis of critical CHD. Absolute fetal volumes of UWM, CGM, and the cerebellum (corrected for PMA) showed a stronger correlation with acquired neonatal brain injury than the proportion of TBV occupied by any of these brain structures. Previous studies have also suggested a link between neonatal brain underdevelopment and acquired neonatal brain injury.[109, 110] However, our study cannot reveal whether there is a causal relationship between smaller brain volumes and acquired brain injury. Brain underdevelopment might be accompanied by an increased vulnerability of the brain to disturbances in cerebral blood flow and cerebral oxygen delivery during intensive care admission and cardiac surgery.

All critical CHD groups have reduced oxygen delivery compared with healthy controls, and this reduction in brain oxygen delivery is directly associated with smaller fetal and neonatal brain volumes.[94, 96] Oxygen delivery is dependent on cerebral blood flow and blood oxygen content. As a result of the cardiac defect, cerebral blood flow is most severely altered in single-ventricle physiology and left-ventricle outflow tract obstruction, whereas cerebral blood oxygen content is lowest in transposition of the great arteries.[111] Despite these differences in antenatal cerebral circulatory disturbances, we found no differences in growth trajectories of brain volumes among these CHD groups, reflecting most critical cardiac defects. Smaller fetal TBV[95] and lower neonatal brain growth rate[112] have been described in single-ventricle physiology compared with other cardiac defects; however, most neonatal studies have shown comparable reductions in the volumes of UWM and CGM among transposition of the great arteries and single-ventricle physiology.[19, 93] All individuals with critical CHD are at risk of reduced antenatal brain growth; however, other patient-related factors than the cardiac defect itself might determine the degree of brain underdevelopment.

This study is one of the first to explore the correlation between fetal and neonatal brain volumes using quantitative longitudinal MR imaging. This study provides important information for future research investigating whether there is a causal relationship between brain underdevelopment and brain injury, fetal brain underdevelopment as an early biomarker for neurodevelopmental outcomes, and potential strategies to improve antenatal brain growth.

This study has several limitations. First, postnatal CHD diagnosis increases the risk of preoperatively acquired ischemic brain injury, especially when low cardiac output syndrome is present. However, this population cannot be included in fetal MR imaging studies and was therefore not examined in this study. Second, absolute brain and CSF volumes are dependent on the MR imaging protocol and postimaging processing methods and are therefore difficult to compare among studies. Third, the subgroup sample size (by heterogeneity of CHD populations) limited the possibility of performing subgroup analysis. The subgroup analysis of fetal-neonatal brain volumes by PMA

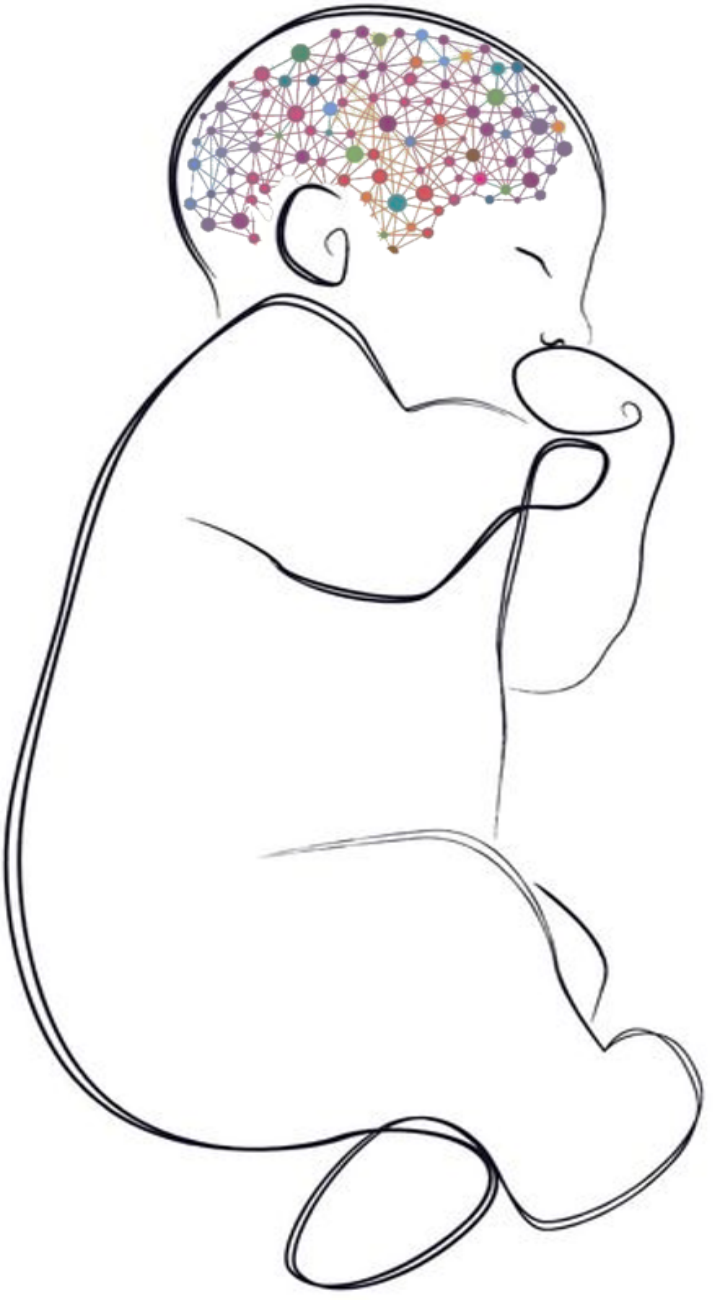
is at risk of type 2 errors, potentially not showing differences among the groups.

5.5 Conclusions

Linking longitudinal fetal and neonatal brain and CSF volumes,this study reveals the strong relationship between brain volumesover postmenstrual age in individuals with critical CHD. Addi-tionally, the results show a negative correlation between fetalbrain volumes and neonatal ischemic brain injury. Combiningthese findings, this study suggests that fetal MR imaging can beused as an early biomarker to estimate neonatal neurologic find-ings in critical CHD.

5.6 ACKNOWLEDGMENTS

The authors would like to thank Gabrielle van Iperen, RaymondStegeman, and Felix Haas for their contributions to the brainimaging program, and the Congenital Heart Disease Life Spanstudy group of the Wilhelmina Children's Hospital in Utrecht,which includes the Department of Neonatology, Department ofPediatric Intensive Care, Department of Obstetrics, Departmentof Pediatric Cardiology, Department of Pediatric CardiothoracicSurgery, Department of Pediatric Anesthesiology, Department ofRadiology, Department of Medical Psychology and Child Devel-opment, and Exercise Center.



Chapter 6

Assessment of brain injury and brain volumes after PHVD

BASED ON: M. N. Cizmeci, N. Khalili, N. H. Claessens, F. Groenendaal, K. D. Liem, A. Heep, I. Benavente-Fernández, H. L. van Straaten, G. van Wezel-Meijler, S. J. Steggerda, J. Dudink, I. Išgum, A. Whitelaw, M. J. Benders, L. de Vries, and the ELVIS study group. "Assessment of brain injury and brain volumes after posthemorrhagic ventricular dilatation: a nested substudy of the randomized controlled elvis trial", *The Journal of Pediatrics,* vol. 208 (2019), pp. 191–197.

Abstract

Objective: To compare the effect of early and late intervention for posthemorrhagic ventricular dilatation (PHVD) on additional brain injury and ventricular volume using term-equivalent age magnetic resonance imaging (TEA-MRI).

Study Design: In the ELVIS (Early versus Late Ventricular Intervention Study) trial 126 preterm infants ≤34 weeks gestation with PHVD were randomised to low threshold (LT, ventricular index (VI) > p97 and anterior horn width (AHW) > 6mm) or higher threshold (HT, VI > p97 + 4mm and AHW > 10mm). In 88 (80%) with a TEA-MRI, the Kidokoro Global Brain Abnormality Score and the frontal and occipital horn (FOH) ratio were measured. Automatic segmentation was used for volumetric analysis.

Results: TEA-MRI was obtained in 44 in the LT group and 44 in the HT group. The total Kidokoro score of the infants in the LT group was lower than in the HT group (median (interquartile range): 8 (5-12) vs 12 (9-17), respectively; p<0.001). There were more infants in the LT group with a normal or mildly increased score versus more infants in the HT group with a moderately or severely increased score (46% vs. 11% and 89% vs. 54%, respectively; p=0.002). The FOH ratio was lower in the LT group (0.42 (0.34-0.63) vs 0.48 (0.37-0.68), respectively; p=0.001). Ventricular CSF volumes could be calculated in 47 infants and were smaller in the LT group (p = 0.03).

Conclusion: Our findings demonstrate more brain injury and larger ventricular volumes in the HT group. These results support the positive effects of early intervention for PHVD.

6.1 Introduction

Over the past decades, substantial developments in obstetric and neonatal care have resulted in a significant increase in the survival of premature infants. Along with mortality, a further aim has been to reduce the major morbidities and improve neurodevelopmental outcome. However, due to the increased survival rate of extremely preterm infants, germinal matrix-intraventricular hemorrhage (GMH-IVH) continues to be a serious complication of preterm birth [113, 114]. Posthemorrhagic ventricular dilatation (PHVD) occurs in approximately 30-50% of the preterm infants after a severe hemorrhage and increases the risk of neurocognitive and motor impairments [4]. Adverse effects of PHVD on the developing newborn brain include white matter injury and decreased volumes of deep gray matter and cerebellum, indicating the importance of timely intervention to restrict these problems as much as possible [55]. After the use of temporizing methods, overall conversion to a permanent shunt varies from 20-65% depending on the time of onset of the intervention [115]. Given the high rates of infection, dysfunction and life-long dependence after ventriculoperitoneal (VP)-shunt insertion, it would be beneficial if a treatment could reduce the risk of shunt requirement [116, 117]. Removing the hemorrhagic cerebrospinal fluid (CSF) by lumbar punctures or taps from a ventricular reservoir may reduce the need for VP-shunt placement, since removal of CSF that contains blood components, protein, and cytokines might re-establish normal CSF circulation [118]. Although optimum timing of intervention continues to be a matter of debate in the neonatal literature, there is accumulating evidence showing the beneficial effects of early intervention on ventricular dilatation and outcomes [119, 120]. In a recent randomized controlled trial (the ELVIS, Early versus Late Ventricular Intervention Study) no significant difference was found for the need for VP-shunt in those treated before or after crossing the 97th centile +4 mm line of the graph of Levene [121]. However, only a small number of infants in both study arms had a VP-shunt inserted, the lowest number reported in the literature so far (19-23%) [115]. The aim of the present nested substudy was to compare the extent of injury in different brain regions, and brain volumes on term-equivalent age magnetic resonance imaging (TEA-MRI) in patients randomized to the early or late intervention group.

6.2 Patients and Methods

6.2.1 Patients

A total of 126 infants participated in the ELVIS trial, a RCT conducted between 2006 and 2016 to compare the effects of low versus high threshold treatment in preterm infants of ≤34 weeks' gestational age with progressive PHVD. Infants were eligible for the RCT when they had an IVH grade III, with or without a periventricular hemorrhagic infarct (PVHI) according to Volpe [122]. They were randomly allocated to

either low threshold (LT) group (intervention when an increase in ventricular width according to Levene [121] above the p97 line showing an increase towards the p97 + 4mm line but without crossing the p97 + 4mm line, and an increase in diagonal anterior horn width according to Davies et al. [123] above 6 mm towards 10 mm, but not above 10 mm) or high threshold (HT) group (intervention once the ventricular width crossed the p97 + 4mm line and the anterior horn width was above 10 mm). Antenatal and perinatal factors including gestational age, birth weight, sex, the severity of hemorrhage, and timing and type of intervention, and postmenstrual age at MRI day were collected for each patient from the patient files and/or hospital database. Approval from the Research Ethics Boards at each center and informed written parental consent were obtained for all of the patients and for the control infants participating in the study before enrollment into the study.

6.2.2 MRI Acquisition

In all centers MR images were acquired around TEA. A 3.0 Tesla MR system (Philips Healthcare, Best, The Netherlands) using a sense head coil was available at three centers (University Medical Center Utrecht (UMCU), University Medical Center Leiden and Isala Hospital, Zwolle) and from 2014 onwards at Southmead Hospital, Bristol. Until April 2014, a 1.5-Tesla MR system (GE Signa Excite HD system, USA) was used in Bristol. University Medical Center Groningen (SonataVision, Siemens, Germany), University Hospital Puerta del Mar, Cadiz (Magnetom Symphony, Siemens, Germany), Radboud University Nijmegen Medical Centre (Magnetom Symphony, Siemens, Germany), University of Rotterdam (GE Signa Excite HD system, USA) and University of Lisbon (Philips Healthcare, Best, The Netherlands) used a 1.5 Tesla MR system. All participating centers used conventional axial 3D T1-weighted imaging and T2-weighted imaging and followed a predefined MRI protocol according to their institutional guidelines during the study period. Only the high-quality images that were suitable for scoring and volumetric measurements were included in the study.

6.2.3 Assessment of Brain Injury

An investigator with more than 20 years of experience in reading neonatal MRIs (LSdV) who was blinded to the infant's clinical information, and the allocated arm of the RCT, assessed the images. Ventricular measurements (ventricular index (VI) and anterior horn width (AHW)) were performed as described by Levene [121] and Davies et al [123]. The frontal and occipital horn (FOH) ratio was obtained by measuring the widest distances across the frontal horns and the occipital horns, and the average of these measurements was then divided by the largest biparietal diameter as defined by Kulkarni et al. [124]. To evaluate the intraobserver reliability of the measurements, 15 studies from 15 random patients were assessed and the intra-class correlation coefficient (ICC) was calculated. For the assessment of brain injury, a validated scoring

system for evaluating cerebral white matter (WM), cortical gray matter (GM), basal ganglia and thalami (BGT), and cerebellum abnormalities was used. The measurements were corrected for postmenstrual age, and a global brain abnormality score was calculated as the sum of the regional total scores and classified as normal (total score 0-3), mild (total score 4-7), moderate (total score 8-11) and severe (total score 12 or more), as defined by Kidokoro et al. [11].

6.2.4 Assessment of Brain Volumes

Automatic segmentation of cerebral MRIs was applied on axial or coronal T2-weighted images for computerized volume analysis. The images were segmented into 8 regions: cerebellum, myelinated white matter (mWM), BGT, ventricular cerebrospinal fluid (vCSF), unmyelinated WM, brain stem, cortical GM, and extracerebral cerebrospinal fluid (eCSF), as described by Moeskops et al. [13]. Quality of automatic segmentations was established by visual evaluation. Images with low-quality segmentations were excluded from further analysis, and high-quality images were manually edited when deemed necessary prior to further analysis. Subsequently, volumetric measurements of the segmented tissues were obtained by multiplying the number of segmented voxels per tissue by the voxel size. Thereafter, contours were drawn around the structure of interest on consecutive slices through the brain. Both porencephalic cysts and cysts following PVHI but not communicating with the lateral ventricles were included in ventricular volume measurements. The relative volumes of the brain regions were calculated by dividing the volume of the area of interest by total intracranial volume, which includes brain tissues and ventricular and extra-ventricular CSF spaces (Figure 6.1).

6.2.5 Statistical Analysis

Statistical analyses of the data were performed using the Statistical Package for the Social Sciences v21.0 program (SPSS Inc., Chicago, Ill., USA). The continuous variables were presented as mean ± standard deviation (SD) and median (interquartile range, IQR) depending on their distribution. The categorical values were presented as frequency and percentage. Fisher's exact and Chi-squared tests were used to compare categorical variables among groups. Mann-Whitney U test was used to compare non-parametric variables and student t-test was used for the comparison of variables that showed normal distribution. Logarithmic transformation was used to obtain a Gaussian distribution of the non-normally distributed volumetric measurements. Observed associations controlled for the grade of IVH by using multiple regression. To evaluate the reliability of measurements, intra-class correlation coefficient was calculated and classified as good for $0.8<ICC<0.9$ and excellent for $ICC>0.9$. Statistical significance was set at $p < 0.05$.

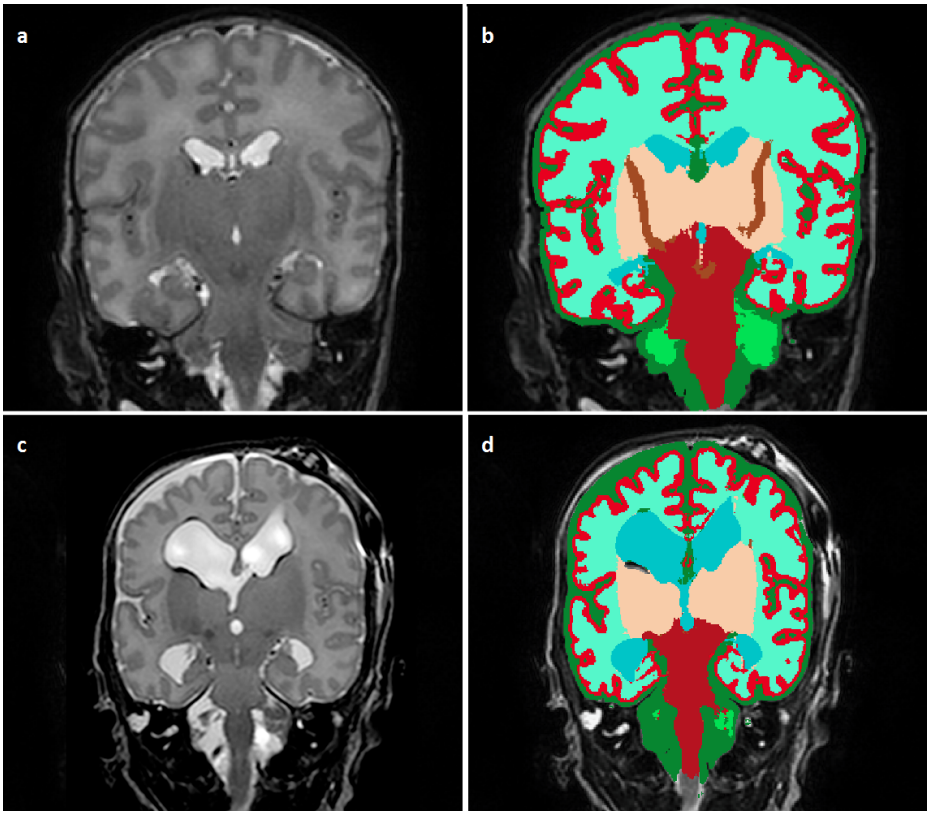


Figure 6.1: T2-weighted coronal MRIs obtained at TEA show A, mildly enlarged ventricular CSF volumes in a preterm infant in the low-threshold group, B, the same infant after automatic segmentation of the MRI into 8 regions for volumetric analysis, C, severely enlarged ventricular CSF volumes in a preterm infant in the high-threshold group, D, automatic segmentation of the image in C.

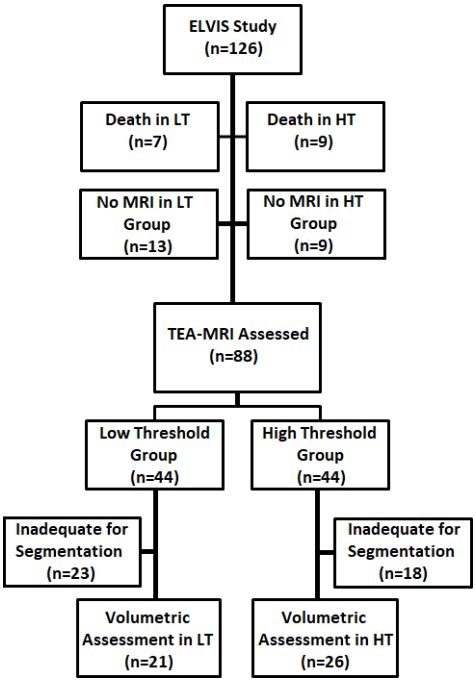


Figure 6.2: Flowchart of patient allocation and subsequent MRI assessments at TEA.

6.3 Results

6.3.1 Study Population

During the 10-year study period, 126 infants were enrolled into the ELVIS cohort of whom 38 were not eligible for inclusion in the present study. Of these ineligible infants, MRI was not available due to death in 16 and was not performed around TEA in 22 infants. The main reasons for not obtaining an MRI were transfer back to the referring hospital where no MR device was present, and not being able to transfer the infant again to the study site exclusively for imaging reasons at term-equivalent age (n = 17). Two had a very early MRI only and two had an MRI well beyond the neonatal period, and one had an MRI in a level 2 hospital with insufficient quality. This resulted in a final sample of 88 infants being enrolled in the study. Of these infants, 44 were in the LT group and 44 were in the HT group (Figure 6.2 online). No statistically significant differences between the LT and HT groups were observed in terms of gestational age, sex, birth weight and postmenstrual age at the time of MRI. Infants who were not included in the present study due to not having MRI were similar to those included with respect to demographic and clinical variables. Characteristics of the participants in whom MRI was completed are presented in 6.1.

6.3.2 Frontal and Occipital Horn Ratio and Kidokoro Score

The intra-observer reliability showed an excellent correlation for the measurements (ICC = 0.94). Median ventricular measurements, including VI and AHW ($p < 0.001$ for both), and the FOH ratio were lower in the LT group ($p = 0.001$). The total Kidokoro score of the infants in the LT group was also lower than that of the HT group ($p < 0.001$). The subgroup analyses were performed after excluding infants with PVHI. Data regarding these measurements and comparisons are presented in 6.1. When the groups were compared in terms of severity of the Kidokoro score, there were more infants in the LT group with a normal or mildly increased score and more infants in the HT group with a moderately or severely increased score ($p = 0.002$). Observed associations persisted after controlling for the grade of IVH. A linear correlation between the Kidokoro score and FOH ratio was found ($r = 0.62$, $p < 0.001$) and average FOH ratio increased by 0.06 for every point increase in the Kidokoro score (95% confidence interval (CI): 0.05-0.08).

6.3.3 Kidokoro Subscores

In the cerebral white matter (WM) evaluation, statistically significant differences were observed between the groups in myelination delay, thinning of the corpus callosum and dilatation of the lateral ventricles subscores. Furthermore, a trend towards biparietal volume reduction in the HT group was seen ($p = 0.07$). The groups were different with regards to cerebral WM subscore ($p = 0.001$). In the cortical GM evaluation, infants in the HT arm showed increased extra-cerebral spaces ($p<0.001$) and a trend towards delayed gyral maturation ($p = 0.07$). The cortical and BGT subscores were lower in the LT group ($p < 0.001$). The groups were similar in terms of cerebellar signal abnormalities and volume reduction ($p = 0.8$ and $p = 0.4$, respectively). The subscore analysis of the infants are tabulated in 6.2.

6.3.4 Brain and CSF Volumes on TEA-MRI

Brain and CSF volumes could be calculated in a total of 47 infants, of which 21 were in the LT and 26 in the HT group (Figure 6.2 online). No statistically significant differences in unadjusted brain and CSF volumes were observed in relation to PHVD (6.3). When the relative volumes of the brain regions were compared after normalization of the variables with logarithmic transformation, ventricular CSF volumes of the LT group were lower than that of the HT group ($p = 0.03$). Unmyelinated WM volumes of the LT and HT groups were not significantly different ($p = 0.3$). Combination of the WM and GM volumes showed a trend towards higher values in the LT group when compared with the HT group ($p = 0.06$). The analyses were performed after excluding infants with PVHI ($n = 4$ in the LT and $n = 11$ in the HT), which resulted in a final sample of 17 infants in the LT and 15 in the HT group. GM volume showed a trend towards

	Low-threshold group (n = 44)	**High-threshold group (n = 44)**	***P* value**
Gestational age at birth, weeks	28.1 ± 2.4	27.8 ± 2.7.	.6*
Birth weight, g	1176 ± 361	1175 ± 404	.9*
Sex			
Male	23 (52)	26 (59)	.5†
Female	21 (48)	18 (41)	
Day of enrollment	9 (6-10)	9 (6-12)	.9‡
Postmenstrual age at MRI, weeks	41.0 (40.4-42.7)	40.9 (40-41.7)	.3‡
GMH-IVH grade			
III	30 (68)	25 (57)	.3†
III + PVHI	14 (32)	19 (43)	
Reservoir inserted	28 (64)	23 (52)	.3†
VP-shunt inserted	9 (20)	12 (27)	.4†
Duration between VP shunt and TEA-MRI, days	4 (-1-28)	10 (-10-29)	.8‡
Ventricular measurements on MRI, mm	13.4 (12.6-15.1)	15.9 (14.5-18.8)	<.001‡
Ventricular width			
AHW	6.6 (5.3-10.3)	10.6 (8.4-13.5)	<.001‡
FOH ratio	0.42 (0.4-0.46)	0.48 (0.43-0.51)	.001‡
Total Kidokoro score	8 (5-12)	12 (9-17)	<.001‡
Infants with grade III	7 (5-9)	10 (8-12)	<.001‡
Infants with grade III+ PVHI	13 (7-19)	16 (15-19)	<.001‡
Kidokoro score severity	.002‡		
Normal	3(7)	0(0)	
Mild	17 (39)	5 (11)	
Moderate	12 (27)	13 (30)	
Severe	12 (27)	26 (59)	

Table 6.1: Clinical characteristics of the study population and MRI assessments

GMH-IVH, Germinal matrix hemorrhage-intraventricular hemorrhage.

Significant values are given in bold.

Values are mean±SD, median (IQR), or number (%).

**t* test.

†χ^2 test.

‡Mann–Whitney *U* test

a higher value (p = 0.06), and combination of the WM and GM volumes was significantly higher in the LT group (p = 0.03). There were no differences between groups in other regions of interest. The FOH ratio was positively associated with ventricular CSF volumes (β [95% CI]: +145 [72; 218], p < 0.001).

6.4 Discussion

In this nested substudy of our recently published randomized controlled ELVIS trial5 of preterm infants with PHVD, infants who were in the LT group had lower global brain abnormality scores and had lower regional total subscores of the cerebral WM, cortical GM and BGT on TEA-MRI. When the total Kidokoro scores were stratified according to the severity, there were significantly more infants with normal or mildly increased scores in the LT group, and significantly more infants with moderately or

	Low-threshold group (n = 44)	**High-threshold group** (n = 44)	***P* value**
Cystic lesions			.2*
None	27 (61)	18 (41)	
Focal unilateral	3 (7)	1 (2)	
Focal bilateral	2 (5)	3 (7)	
Extensive unilateral	11 (25)	20 (45)	
Extensive bilateral	1 (2)	2 (5)	
Focal signal abnormality			.6*
None	33 (75)	28 (63)	
Focal punctate	6 (14)	7 (16)	
Extensive punctate	3 (7)	6 (14)	
Linear	2 (4)	3 (7)	
Myelination delay			.01*
PLIC and corona radiata	0 (0)	0 (0)	
Only PLIC	32 (73)	21 (48)	
Minimal—no PLIC	12 (27)	23 (52)	
Thinning of the corpus callosum			.02*
None	39 (89)	27 (62)	
Partial	2 (4)	5 (12)	
Global	3 (7)	11 (26)	
Dilated lateral ventricles			.01*
Both sides < 7.5 mm	12 (27)	3 (7)	
7.5mm ≤ 1 side < 10 mm	9 (20)	5 (11)	

7.5 mm ≤ both sides < 10 mm or 1 side ≥ 10 mm	13 (30)	13 (30)	
Both sides ≥ 10 mm	10 (23)	23 (52)	
Volume reduction			.07*
cBPW ≥ 77 mm	26 (59)	18 (41)	
77 mm > cBPW ≥ 72 mm	13 (29)	13 (29)	
72 mm > cBPW ≥ 67 mm	2 (5)	10 (23)	
67 mm > cBPW	3 (7)	3 (7)	
Cerebral white matter subscore			.001†
Median (min-max)	4 (1-13)	7 (1-14)	
Cortical gray matter signal abnormality			.5*
None	44 (100)	43 (98)	
Focal unilateral	0 (0)	0 (0)	
Focal bilateral	0 (0)	1 (2)	
Extensive unilateral	0 (0)	0 (0)	
Extensive bilateral	0 (0)	0 (0)	
Gyral maturation			.07*
Delay <2 weeks	4 (9)	1 (2)	
2 ≤ delay < 4 weeks	18 (41)	11 (25)	
Delay ≥ 4 weeks	22 (50)	32 (73)	
Increased extracerebral space			<.001*
IHD < 4 mm	24 (55)	1 (2)	
4mm ≤ IHD < 5 mm	8 (18)	8 (18)	
5mm ≤ IHD < 6 mm	5 (11)	14 (32)	
IHD ≥ 6 mm	7 (16)	21 (48)	
Deep gray matter signal abnormality			.2*
None	40 (91)	38 (87)	
Focal unilateral	0 (0)	4 (9)	
Focal bilateral	1 (2)	1 (2)	
Extensive unilateral	3 (7)	1 (2)	
Extensive bilateral	0 (0)	0 (0)	
Deep gray matter volume reduction			.06*
cDGMA ≥ 9.5 mm	336 (82)	38 (86)	
9.5 mm3 > cDGMA ≥ 8.5 mm	36 (14)	2 (5)	
8.5 mm3 > cDGMA ≥ 7.5 mm	30 (0)	4 (9)	
7.5 mm3 > cDGMA	2 (4)	0 (0)	
Cerebellum signal abnormality			.8*
None	26 (59)	24 (55)	

Punctate unilatera	l7 (16)	9 (21)	
Punctate bilateral	6 (14)	8 (18)	
Extensive unilateral	2 (4)	2 (4)	
Extensive bilateral	3 (7)	1 (2)	
Cerebellum volume reduction			.4*
cTCD ≥ 50 mm	30 (68)	23 (52)	
50 mm > cTCD ≥ 47 mm	7 (16)	11 (25)	
47 mm > cTCD ≥ 44 mm	3 (7)	5 (11)	
44 mm > cTCD	4 (9)	5 (11)	
Cortical gray matter and deepgray matter subscore			<.001†
Median (min-max)	4 (0-14)	6 (2-14)	

Table 6.2: Kidokoro subscore analysis of the study population
cBPW, Biparietal width corrected for gestational age; *cDGMA*, deep gray matter area correctedfor gestational age; *cTCD*, transcerebellar diameter corrected for gestational age; *IHD*, inter-hemispheric distance; *PLIC*, posterior limb of internal capsule. Significant values are given in bold.
*Fisher exact test.
†Mann-Whitney *U* test

severely increased scores in the HT group, even though at the time of randomization the number of infants with a grade III hemorrhage or PVHI was similar in the study arms. Infants in the HT group also demonstrated more delay in myelination and more often partial or global thinning of the corpus callosum. Moreover, lower FOH ratios, VI and AHW at TEA and smaller ventricular CSF volumes were found in infants in the LT group. In the subgroup analysis, after excluding infants with PVHI, combination of the WM and GM volumes was significantly higher in the LT group ($p = 0.03$). Using a structured scale assessment together with the quantification of the ventricular dilatation acquired at TEA, we were able to identify injury in specific regions of the brain, demonstrating the possible beneficial effects of early intervention after the onset of PHVD. The pathogenesis of PHVD is a complex process determined by both direct injury and secondary inflammatory interactions [10, 125–128]. To address the net effects of PHVD on brain lesions in different regions, an objective structured scale assessment was used in combination with volumetric analysis in the present study. This approach enabled us to determine the correlation between ventricular size and the extent of brain injury. The smaller ventricular CSF volumes together with the lower global brain abnormality scores as well as lower regional total subscores of the major regions of the brain in the LT group indicates the possible beneficial effects of early intervention as we found that almost half (46%) of the infants in the LT group had normal or mildly increased Kidokoro scores compared with only 11% in the HT group. In infants with PHVD, expanding ventricles might cause atrophy of the adjacent brain tissue as a result of compression by CSF under pressure [129, 130]. By using

a manual segmentation technique, Jary et al. [129] calculated cerebral, thalamic and cerebellar volumes, and demonstrated that brain growth is significantly impaired in PHVD. Ventricles were larger with a median volume of 48 cc (IQR: 27-145) than the ventricular volumes of both groups in our study (median volume: 18 cc and 24 cc in LT group and HT group, respectively). Brouwer et al. [4] reported data in a small group of infants and showed that PHVD was independently associated with decreased volumes of deep GM, cerebellum and extracerebral CSF, despite early intervention. They found a median ventricular volume of 18.3 cc (range: 8.6 - 64.5) in infants with PHVD, which is lower than we report in the HT group and overlaps considerably with values we found in the LT group. It has been shown that in infants with severe IVH who developed PHVD, ventricular size may be an important determinant of long-term neurodevelopmental outcome and infants with severe IVH who developed PHVD had worse neurodevelopmental scores compared with those who did not develop PHVD [122, 131, 132]. Recently, Leijser et al. [120] reported in their large cohort of preterm infants with PHVD that those who underwent intervention based on ventricular measurements, prior to the development of symptoms, even when eventually requiring a VP-shunt, had outcomes indistinguishable from those without intervention, all being within the normal range. Infants who first received intervention once clinical symptoms had occurred had worse outcomes. The volumes of the ventricles, and combined volume of the unmyelinated WM and GM regions were in favor of the LT therapy in the present study. We also measured the VI and AHW on TEA-MRI, which revealed smaller lateral ventricles in the LT group. Whether the smaller ventricular volumes, and preserved unmyelinated WM and GM volumes of infants who underwent LT therapy will be associated with improved neurodevelopmental outcomes in the ELVIS trial is being assessed. The higher Kidokoro scores in infants in the HT group are in line with the accumulating literature suggesting that progressive ventricular dilatation and prolonged pressure might be deleterious to the immature brain. A rapidly enlarging ventricular system could result in compression of adjacent brain parenchyma and this has been used as an explanation for the MRI signal abnormalities in various regions of the brain [133, 134]. Since infants in the HT group had larger ventricular volumes than those of the LT group, the Kidokoro scores of these infants, which increase directly with the presence of signal abnormalities could have increased. PHVD-induced microstructural white matter injury, as stated previously by Brouwer et al. [55], might serve as another explanation for the signal abnormalities on TEA-MRI in our cohort. It is also worth noting that FOH ratios showed good correlation with ventricular volumetric measurements, which can be used as a practical assessment tool for calculating the ventricular volumes in patients with PHVD. The present study has several limitations. First, as this was a multicenter study, MRI protocols were not the same across centers, which could have lead to varying image qualities. Second, a relatively large number of segmented MRIs could not be used for the volumetric analysis. This was due to the use of convolutional neural network technique, which was trained on seg-

Volumes	Low-threshold group (n = 21)	High-threshol dgroup (n = 26)	P value*
Ventricles			
Absolute	18 (11-30)	24 (17-37)	.07
Relative	0.05 (0.02-0.07)	0.06 (0.04-0.09)	.03
Absolute†	16 (12-27)	23 (15-31)	.4
Relative†	0.04 (0.03-0.06)	0.06 (0.04-0.07)	.1
Unmyelinated whitematter			
Absolute1	65 (136-186)	154 (128-184)	.5
Relative	0.3 (0.2-0.4)	0.3 (0.2-0.4)	.3
Absolute†	165 (140-186)	171 (129-185)	.7
Relative†	0.3 (0.2-0.4)	0.3 (0.2-0.4)	.6
Cortical gray matter			
Absolute1	28 (119-166)	131 (117-155)	.8
Relative	0.3 (0.2-0.4)	0.3 (0.1-0.4)	.5
Absolute†	129 (119-173)	126 (114-152)	.3
Relative†	0.3 (0.2-0.3)	0.2 (0.2-0.3)	.06
Combined white and gray matter			
Absolute	304 (238-378)	302 (239-316)	.3
Relative	0.6 (0.5-0.7)	0.5 (0.5-0.6)	.06
Absolute	†307 (295-349)	302 (244-315)	.1
Relative†	0.6 (0.5-0.7)	0.5 (0.5-0.6)	.03
Basal ganglia and thalami			
Absolute	25 (21-27)	23 (19-25)	.4
Relative	0.04 (0.03-0.05)	0.04 (0.04-0.05)	.9
Absolute†	25 (22-27)	24 (22-25)	.5
Relative†	0.05 (0.04-0.06)	0.05 (0.04-0.06)	.8
Cerebellum			
Absolute	31 (25-41)	31 (23-41)	.8
Relative	0.06 (0.05-0.07)	0.06 (0.03-0.18)	.6
Absolute†	31 (25-41)	32 (25-47)	.5
Relative†	0.06 (0.05-0.07)	0.05 (0.04-0.07)	.5
Extracerebral CSF			
Absolute	127 (98-140)	116 (102-137)	.9
Relative	0.2 (0.2-0.3)	0.2 (0.2-0.3)	.5
Absolute†	113 (98-145)	119 (105-158)	.4
Relative†	0.2 (0.2-0.3)	0.2 (0.2-0.3)	.3

Table 6.3: Distribution of absolute and relative brainand CSF volumes between the groups

Data are presented as median (interquartile range).

Absolute volumes are presented in milliliters.

Significant values are given in bold.

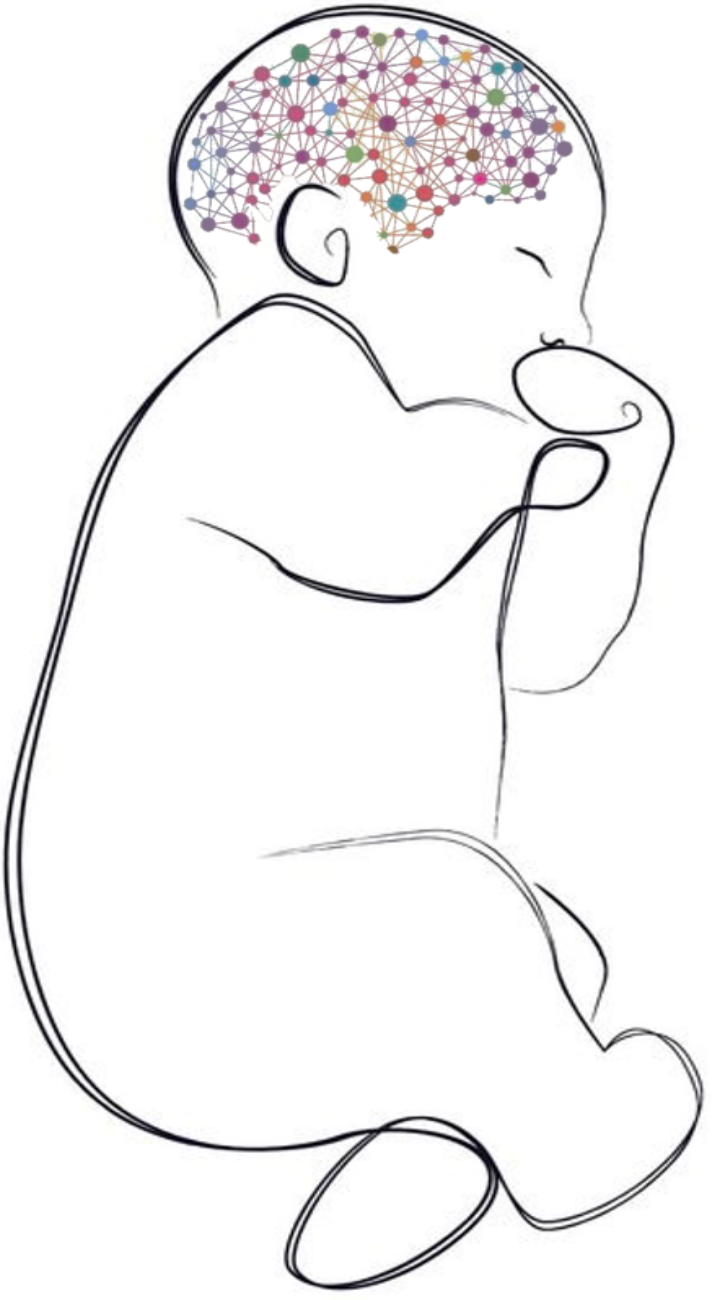
* *P* values are presented after logarithmic transformation of relative volumes.

†Volumes after excluding infants with periventricular hemorrhagic infarct (n = 17 in the low-threshold group and n = 15 in the high-threshold group).

mented images of preterm neonates without any pathology. As a consequence, the automatic segmentation had limitations in segmenting scans with PHVD. Furthermore, because the automatic segmentation method was trained with scans acquired in the main study site, it was not always able to provide optimal segmentation for images obtained with a different protocol. This technique also did not allow differentiation between the basal ganglia and thalamic volumes and could not demonstrate precise segmentation of myelinated white matter due to technical reasons. Third, because the present study was a nested substudy, it is possible that the lack of statistically significant differences in volumetric measurements was due to a limited sample size. Finally, there is a potential for selection bias arising from the excluded MRIs. However, we found that the excluded infants were similar with respect to demographic and clinical parameters. The main strength of our study was the use of an objective scoring system enabling assessment of the extent of brain injury and reliable quantification of the ventricular and brain volumes in 80% of the surviving infants. Today, PHVD remains a serious complication of IVH, and control of PHVD using LPs before or just after the VI crossed the p97 + 4mm line was associated with the lowest need for VP shunt reported in the literature, according to the ELVIS trial.5 This nested substudy of the ELVIS trial, designed to address parenchymal injury in different regions of the brain together with the quantification of the CSF and brain volumes, demonstrates beneficial effects of early intervention on the extent of brain injury and ventricular CSF volumes. Whether these findings translate into improved neurological development is being assessed and will be the subject of a later report.

6.5 Acknowledgement

The authors thank the MR technicians for their dedicated help to obtain the MR images.



Chapter 7

Brain tissue segmentation in MR images of infants with stroke

BASED ON: N. Khalili, N. van der Aa, F. Groenendaal, L. de Vries, J. Dudink, M. Zreik, N. Wagenaar, J. Breur, M. Viergever, and M. B. I. Išgum. "Brain tissue segmentation in neonatal MRI using multi-modal CNN", *in preparation* (2020).

Abstract

Perinatal arterial ischemic stroke (PAIS) causes significant morbidity and severe long-term neuropsychological impairment. Neurogenerative interventions might have promising effects on ischemic tissue. To evaluate the effect of the treatment and brain development after PAIS, brain MR images acquired after the acute phase following the stroke and three months later can be utilized. Such evaluation requires accurate quantitative measurements of the ischemic tissue and of the unaffected brain tissue volume. MR images made after symptoms of PAIS allow visualization of the stroke tissue on diffusion-weighted MRI (DWI), while different brain tissues are visible on T2-weighted MRI. Therefore, a multi-modal segmentation method analyzing both of these images is proposed to identify stroke and the tissue classes of the unaffected parts of the brain, i.e. cerebellum (CB), basal ganglia and thalami (BGT), ventricular cerebrospinal fluid (vCSF), white matter (WM), brain stem (BS), cortical gray matter (cGM) and the extracerebral cerebrospinal fluid (eCSF). The developed method uses a convolutional neural network with U-net-like architecture, encodes the features of DWI and T2-weighted scan in two separated encoding paths that are fused in the decoding path. At three months follow-up, CSF cysts appear in place the ischemic tissue. Hence, a single-modality network is applied to segment brain into the same seven tissue classes. To evaluate the segmentation quantitatively Dice coefficient (DC) and Mean surface distance (MSD) per tissue class are computed between automatic and manual expert annotations. In the neonatal scans, multi-modal network was tested on four scans and achieved the average DC of 0.89 and MSD of 0.44 mm over all tissue classes. In the three-months follow-up MR scans, segmentation method was tested on six scans and the average DC and MSD were 0.91 and 0.32. Furthermore, to evaluate the proposed networks on a larger set, the proposed methods were applied on MRI scans to 58 neonates. The quality of segmentation was scored using 5-point Likert score where 1 indicates the segmentation can not be used for further analysis, and 5 indicates accurate segmentation quality. The segmentation performance of multi-modal network and single-modality network achieved respectively 4.3 and 4.4.These results demonstrate that the proposed approach can achieve accurate segmentation of scans acquired in the acute phase following stroke and in scans acquired at three months follow-up. Thereby, the segmentation methods can potentially be used for quantification of brain tissue volume development in PAIS.

Stroke, Brain segmentation, Neonatal MRI, Deep learning, Multi-modal segmentation

7.1 Introduction

Perinatal arterial ischemic stroke (PAIS) is an important cause of long-lasting neurodevelopmental problems [135, 136]. PAIS is diagnosed with MR as soon as infants present with hemi-convulsions. Moreover, infants suffering from PAIS are imaged three months later to estimate the residual damage. Volumetric measurements obtained from segmentation of brain tissue classes between the two MR acquisitions quantitatively demonstrate the effect of ischemic brain tissue on brain development [137]. These analyses can be used to investigate the effect of neuroregenerative intervention such as recombinant human erythropoietin (rhEPO) on brain developments [56]. Given that obtaining extremely time-consuming manual annotations for these quantitative measurements is practically infeasible in the clinical routine or in large studies, user-independent automatic segmentation is required.

Several methods have been developed to segment brain tissue classes in neonatal MRI without visible pathology [12, 13, 71, 138–144]. However, stroke may cause large variations in unaffected brain tissues and ischemic tissue morphology which may substantially vary appearance and shape as illustrated in Figure 7.1. In large strokes, swelling may be observed, compressing the ventricles and in some cases it may lead to a midline shift [145]. Small strokes may occur in multiple brain regions and may cause local changes but they do not affect other anatomy. In the months following the acute phase following stroke, the brain deforms due to disintegration of the ischemic tissue that is replaced by CSF. Hence, the scans made about three months after the stroke show a wide variation in brain shapes. These variations, in both neonatal MR images and in the follow-up MRs, pose challenges for automatic segmentation methods.

Thus far, a few methods have been proposed for segmentation of ischemic lesions in neonatal MRI. Murphy et al. [146] proposed a machine learning method to segment hypoxic- ischemic (HI) brain tissue in encephalopathic asphyxiated neonates using DWI. The proposed method classifies ischemic pixels based on their spatial and intensity features using random forest. The method achieved moderate sensitivity and high specificity and demonstrated promising segmentation results for quantifying HI brain tissue. In another study, Išgum et al. [147] proposed a method for segmentation of ischemic stroke exploiting DWI and ADC maps. The method used spatial and texture features and performed classification using a linear classifier. The results demonstrate high sensitivity at the expense of limited false positive errors. Furthermore, Golsh et al. [148] compared three methods, namely symmetry-integrated region growing (SIRG), hierarchical region splitting (HRS) and modified watershed (MW) segmentation to detect HI using DWI. The segmentation method demonstrated moderate sensitivity and high specificity.

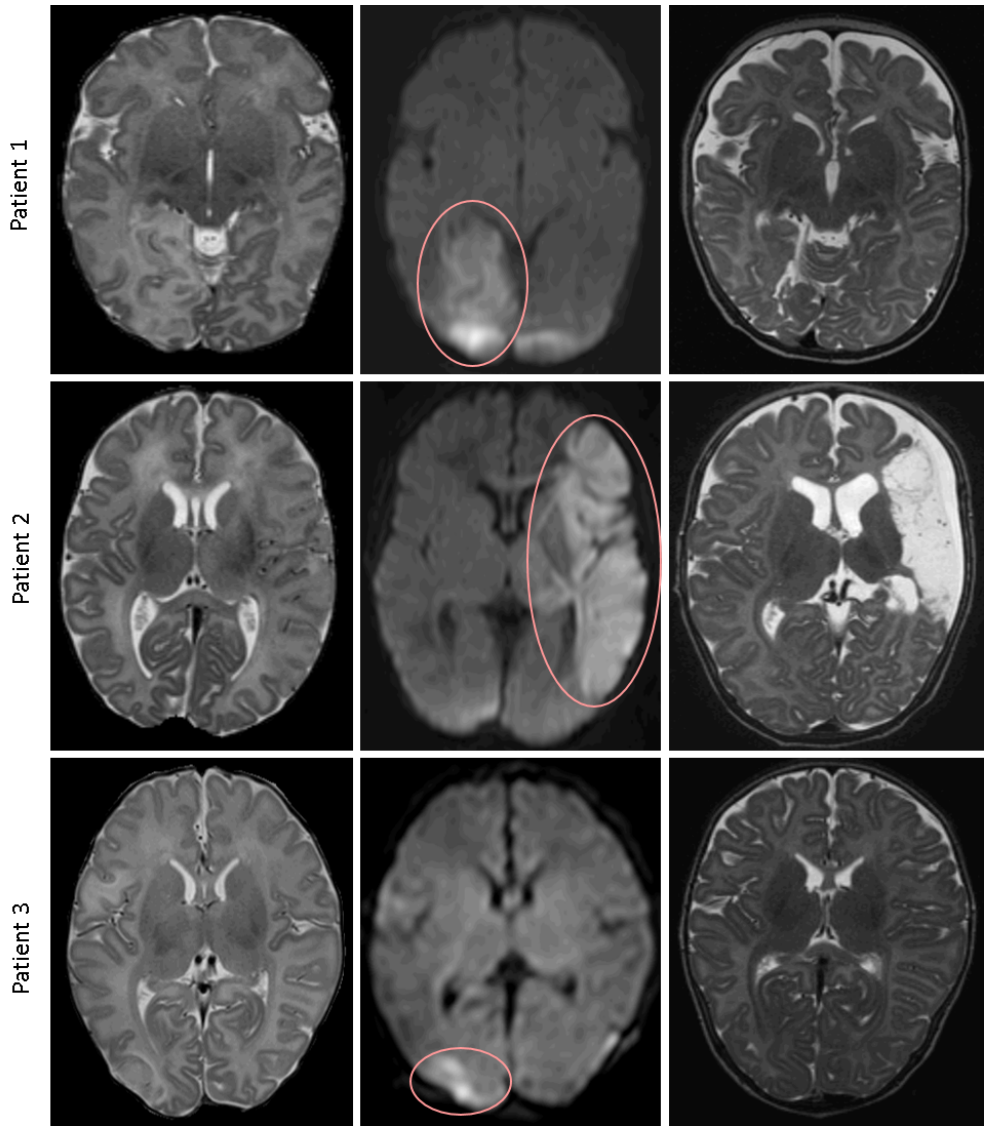


Figure 7.1: Examples illustrating slices from MR scans of three neonates with PAIS, showing the T2-weighted image (left) and diffusion weighted image (middle) in the acute phase. In T2-weighted image at three months after the stroke (right) some asymmetry can be observed (Patient 1), severe tissue loss, replaced by CSF (Patient 2), and limited tissue loss (Patient 3). The areas with stroke are indicated by a red arrays on DWI.

To the best of our knowledge, deep learning methods for ischemic stroke segmentation have only been applied to adult MR brain. Zhang et al. [149] proposed a 3D convolutional neural network (CNN) with a dense connection to segment ischemic stroke in adult diffusion-weighted brain MRI (DWI). The method was evaluated on local and public data from Ischemic stroke lesion segmentation (ISLES 2015) challenge. Praveen et al. [150] proposed an unsupervised feature learning approach for the same task. This method applied an autoencoder on four MR modalities comprising of T1, T2, DWI and FLAIR. The encoded features were classified into normal or stroke with a support vector machine classifier. The network was evaluated on the data from ISLS 2015 challenge and achieved an accurate segmentation performance.

These methods performed segmentation of the stroke lesions only and did not analyze unaffected brain tissues. However, investigating these tissues is important to e.g. quantify the effect of treatment or for prediction of stroke outcome. Hence, in this study, we propose automatic method for segmentation of both stroke lesions and brain tissues (CB, basal ganglia and thalami (BGT), ventricular cerebrospinal fluid (vCSF), WM, BS, cortical gray matter (cGM), eCSF) in infants suffering from PAIS in a fast developing neonatal brain. Brain tissues in neonates are best visible on T2-weighted MRI and ischemic stroke tissue is visible on DWI. Thus, to develop a method that segments brain tissues and stroke together in an end-to-end manner, a multi-modal network extracting features from both DWI and T2-weighted scan is needed. Given that only brain tissues are visible in the follow-up imaging, and no longer detectable ischemic lesion are present on DWI, only T2-weighted images need to be analyzed. Hence, to allow segmentation of stroke lesions in images acquired after the stroke onset and its consequence on residual brain tissues in images acquired in the three months follow-up, we developed two networks that can analyze multi-modal and single-modal input data. Inspired by recently proposed multi-modal segmentation network Fusenet [151], our multi-modal network has a U-net architecture where in the case of the multi-modal input two encoding paths - one for DWI and one for T2-weighted image - are connected to decoding path using skip connections [52]. Instead of fusing features in encoding paths, the skip connection of U-net is used to concatenate features in the decoding paths. In case of the single-modality input, network has a standard U-net like architecture. Both networks have the same number of layers and the same kernel sizes.

The remainder of this paper is organized as follows: In Section 7.2 the data set used for the method development and evaluation is described. In Section 7.3 the segmentation method is presented and in Section 7.4 the evaluation method is given. The performed experiments and their results are presented in Section 7.5, followed by a discussion in Section 7.6 and conclusion in Section 7.7.

7.2 Data

The study included 64 infants with suspected PAIS admitted to the Department of Neonatology, University Medical Center Utrecht (UMCU), the Netherlands. All infants had acute symptoms often in the first week after birth, most often (hemi) convulsions, but in a few infants their symptoms were less neurologically specific. All had PAIS confirmed on their neonatal MRI. Informed verbal parental consent was obtained to perform an MRI for clinical purposes. The Institutional Review Board of the UMCU approved the use of MRI data for anonymous data analysis and waived the requirement to obtain written informed consent.

7.2.1 Neonatal MRI

Infants suspected of PAIS underwent MR brain imaging 3 to 10 days after birth. 26 neonates were scanned with 1.5T Philips Achieva MR scanner and the remaining 40 neonates were scanned with 3T Philips Achieva MR scanner. The imaging parameters are listed in Table 7.1. Furthermore, clinical characteristics of the study population are listed in Table 7.2. In this study, T2-weighted MRI and DWI were used to segment stroke tissue and brain tissue classes. DWI has high sensitivity for detecting ischemic lesions in the acute phase, that is, during the first few days after the insult when signs on T2-weighted scan can be missed. On the other hand, brain tissue classes are not clearly visible on DWI, thus T2-weighted scan was used for this purpose. Both sequences were acquired in axial plane and B value in DWI was set to 1000 s/mm^2.

To define the reference standard, the brain was manually segmented into seven tissue types (CB, BGT, vCSF, WM, BS, cGM and eCSF). This was performed by an expert on 10 T2-weighted MRIs among which 5 were made with 1.5T MR and 5 with 3T MR scanner. The stroke tissue was manually segmented on 10 DWI acquired of the same infants by a different expert. Annotation was accomplished by manual pixel-wise painting of the brain tissues in each image slice using an in-house developed software. The labeling of each of the seven classes was indicated by a color overlay. The software allowed the user to zoom-in, zoom-out and scroll through the slices during the manual segmentation. The manual segmentation protocol was identical to the protocol described by Išgum et al. [22] for neonatal brain tissue segmentation. To merge the stroke segmentations with the brain tissue segmentation labels, DWI and T2 scans were registered using elastic transformation. Prior to registration, DWI scans were first resampled to match the resolution of T2-weighted scan using third-order B-spline interpolation. Thereafter, registration was performed using Elastix registration toolbox as a composition of a rigid Euler transform followed by a nonrigid B-spline transformation [152]. Both transformations were optimized with an adaptive stochastic gradient descent using mutual information as the similarity metric. A rigidity penalty [153] was applied to the entire volume during deformable registration. The

		Sequence	Nr	Matrix	Voxel Size	TE/TR
neonatal Scan	1.5T	T2	24	256 × 256 × 50	0.703 × 0.703 × 2	50/7670
		DWI	24	256 × 256 × 25	0.703 × 0.703 × 4	89/4000
	3T	T2	40	512 × 512 × 50	0.352 × 0.352 × 2	120/6629
		DWI	40	224 × 224 × 33	0.804 × 0.804 × 3	114/5453
Follow-up Scan	1.5T	T2	24	256 × 256 × 50	0.703 × 0.703 × 2	50/7670
	3T	T2	40	512 × 512 × 52	0.352 × 0.352 × 2	120/6629

Table 7.1: Parameters of MRI scans. For each set and MR sequence (Sequence) the table lists total number of scans (Nr), reconstruction matrix (Matrix), repetition time (TR) in ms, echo time (TE) in ms, and reconstructed voxel sizes expressed in mm.

	Middle cerebral artery					Posterior cerebral artery
	Main branch	Anterior trunk	Posterior trunk	Cortical branch	Lenticulostriatal branch	
Sex (male/female)	9/4	2/2	4/3	11/9	4/4	4/2
Gestational age (weeks)	39 [37 – 42]	40 [38-41]	39 [36-40]	40 [37-42]	40 [38-41]	40 [39-41]
Birth weight (grams)	3190 [2450-3572]	3285 [2535-3525]	3300 [2512-3515]	3530 [2150-4040]	3474 [2812-5005]	3143 [2145-4330]
Age at 1st scan (days)	4 [1-21]	4 [3-16]	3 [1-5]	4 [1-7]	4 [2-7]	4 [4-5]
Age at 2nd scan (days)	101 [80-127]	95 [84-115]	97 [75-118]	93 [74-111]	93 [78-108]	91 [89-97]
Seizures	13 (100%)	4 (100%)	6 (86%)	18 (90%)	6 (75%)	5 (83%)
Side of stroke (left/right/bilateral)	5/8/0	1/3/0	5/2/0	14/5/1	6/1/1	3/3/0

Table 7.2: Characteristics of the study population. Based on the arterial territory of the stroke, the table lists patient sex, gestational age, birth weight, age at neonatal scan acquisition, age at follow-up scan acquisition, presence of seizure symptoms and stroke side.

parameter files used for the registration are described by Florkow et al [154] [1]. Stroke labels were subsequently transformed from DWI to the T2 scan using the obtained transformation parameters.

7.2.2 Follow-up MRI

Same as scans acquired after the acute phase, 26 neonates were scanned with 1.5T Philips Achieva MR scanner and the remaining 40 neonates were scanned with 3T Philips Achieva MR scanner. Image acquisition parameters are listed in Table 7.1. Because in the follow-up imaging, only brain tissue classes are visible and ischemic area has disintegrated, we analyzed T2-weighted MRI.

Following the same procedure as in neonatal MR, the reference standard was defined on 12 scans following the same manual annotation protocol dividing the brain into seven tissue classes (CB, BGT, vCSF, WM, BS, cGM and eCSF).

[1] http://elastix.bigr.nl/wiki/index.php/Par0059

7.3 Method

To quantify volumetric brain development in infants with PAIS a multi-modal network to segment brain tissue classes and ischemic tissue using T2-weighted MRI and DWI is proposed. To align T2-weighted MRI showing brain tissue types of unaffected areas and DWI showing ischemic tissue into a common coordinate, they are resampled and aligned following the same procedure described in Section 7.2.1. Furthermore, a single-modality network with similar architecture was developed to segment brain tissue classes in T2-weighted scans acquired in three-months follow-up.

7.3.1 Network Architecture

To analyse multi-modal and single-modality input data, our networks are based on U-net architecture [52]. U-net architecture is widely used in medical image analysis particularly in segmentation. It consists of encoding and decoding paths. The encoding path comprises of repeated 3×3 convolutions followed by rectified linear units (ReLUs). The skip connections connect encoding path to decoding path in each layer after which a 2×2 max pooling downsamples the features. The number of feature channels doubles after every two convolutional layers. In the expansion path, an up-sampling is followed by a 2×2 convolution which halves the number of feature channels. The results are concatenated with encoding feature maps convolved by two 3×3 convolutional layers followed by a ReLU. In the final layer, one 1×1 convolutional layer maps each component of the feature vector to the desired class. Batch normalization [61] is applied after all convolutional layers to allow for faster convergence.

The multi-modal network architecture segments brain tissues and stroke from DWI and T2-weighted scan in the neonatal scans. For this, we add one extra encoding path to encode DWI in parallel with T2. The skip connections concatenate the feature maps from DWI and T2-weighted scan into decoding feature maps. Figure 7.2 illustrates the network architecture.

A single-modality network segments brain tissue classes in the three-months follow-up scans. For this, U-net architecture described above is employed. The network segments brain tissues into seven tissue classes.

Both networks were trained on three consecutive slices which are provided to the network as channels. Multi-modal input network was trained with 3D patches of $416 \times 416 \times 3$ voxels in each path. A single-modality network was trained with 3D patches of $512 \times 512 \times 3$ voxels. To compensate for the extreme imbalance among class distributions in the brain tissue segmentation, the average Dice coefficient over all classes was used as the loss function. In this way, the network was penalized equally for the mislabelling of classes with the least and most samples. The network was optimized using the Adam optimizer with Nesterov momentum [59, 60] using a fixed learning rate of 0.0001. Standard data augmentation techniques, namely random flip-

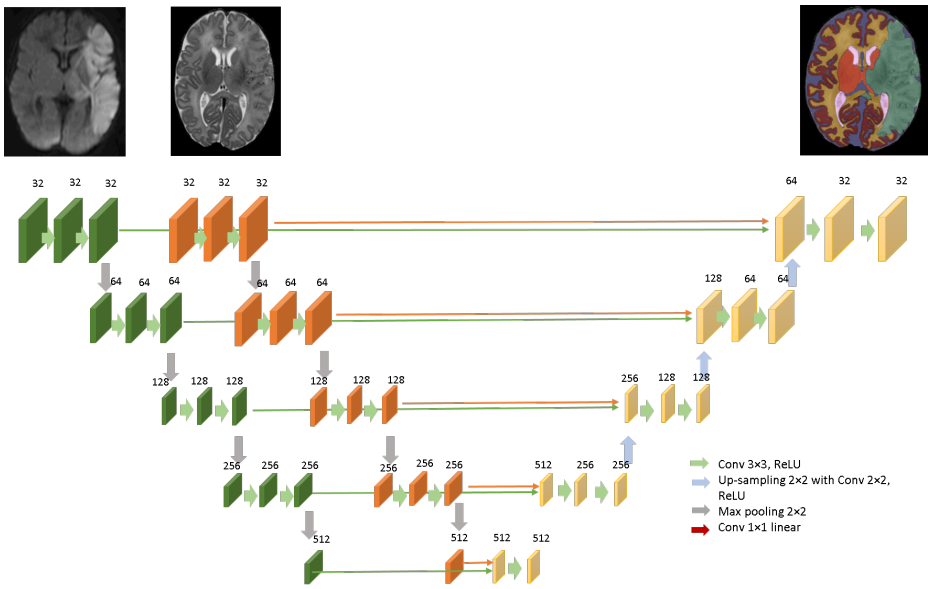


Figure 7.2: Multi-modal Unet-like architecture. The network consists of multiple contraction paths and an expansion path. Each contraction path consists of repeated convolution layers followed by max pooling. The expansion path concatenates two contracting paths with the upsampling feature maps, that are followed by convolution layers. The network inputs are three slices of DWI and T2-weighted images and the output is segmentation.

ping and rotation, were used during training to increase the variation in the training data. The slices were flipped in the horizontal and vertical direction with 20% probability and were rotated with a rotation angle randomly chosen between 0 and 20 degrees.

7.4 Evaluation

Defining manual reference segmentation for brain tissue classes and stroke requires highly specialized expert. In addition, the task is extremely time consuming. Therefore, in our study manual annotations were available in a limited subset of MR scans. In images with available manual annotation, the automatic brain tissue segmentation was evaluated by comparing manual reference segmentation and automatically obtained segmentation using the Dice coefficient (DC) as overlap metric, and the Mean surface distance (MSD) as a boundary measure. These metrics were calculated in 3D. In images without available manual reference annotation, qualitative evaluation was carried out by an expert using 5-points Likert scale [155]. The automatic segmentations of all brain tissue types and of ischemic tissue were scored separately. The score was 1 when the segmentation failed in all tissue types, 2 when the segmentation quality was poor in three or more tissues, 3 when the segmentation quality was poor in one or two tissues, 4 when the segmentation quality was good over all tissues, and 5 when the segmentation was accurate. Similarly, the ischemic tissue segmentation scored 1 when the automatic method failed, 2 when more than 50% of ischemic tissue was segmented incorrectly, 3 when about 25%-50% of ischemic tissue was segmented incorrectly, 4 when the ischemic tissue was well segmented but with errors along the boundary, and 5 when the segmentation was accurate.

7.5 Experiments and Results

Proposed multi-modal network for segmentation of ischemic stroke tissue and brain tissue types was applied to neonatal MRI. Proposed single-modality network was applied for segmentation in follow-up scans. In the ablation study performance of the proposed multi-modal network was compared with two networks.

7.5.1 Quantitative evaluation of the automatic segmentation in the neonatal MRI

The scans with available manual segmentation were divided equally into training and test set based on the strength of scanner (3T and 1.5T) and visually graded stroke size. Five scans were acquired with 3T scanner and remaining 5 scans were acquired with 1.5T scanner. Hence, the network was trained on DWI and T2-weighted scans of 6 patients and tested on the remaining 4 patients. The average performance of the multi-modal network per tissue over 4 patients is listed in Table 7.3 for each of the

		CB	eCSF	BGT	vCSF	WM	BS	cGM	Stroke	Mean
Multi-modal	DC	0.94	0.84	0.92	0.86	0.91	0.94	0.86	0.83	0.89
	MSD	0.97	0.21	0.40	0.18	0.16	0.19	0.13	1.33	0.44
Two separate networks	DC	0.82	0.79	0.81	0.81	0.88	0.88	0.84	0.77	0.82
	MSD	0.87	0.30	0.92	0.36	0.39	0.41	0.23	2.00	0.69
Multi-channel	DC	0.94	0.83	0.90	0.83	0.90	0.93	0.84	0.80	0.87
	MSD	1.04	0.25	0.91	0.23	0.19	0.77	0.17	4.56	1.01

Table 7.3: Segmentation performance for seven brain tissues and stroke. The evaluation of three experiments is shown 1) the proposed multi-modal network trained with DWI and T2-weighted MRI 2) multi-channel network trained with DWI and T2-weighted MRI as two input channels 3) two separate networks: one trained with DWI and the second one trained with T2-weighted MRI. The results are expressed as the mean Dice coefficient (DC) and the Mean surface distance (MSD) in mm over evaluated scans.

seven brain tissue classes and stroke. The average DC ranged from 0.83 for ischemic stroke to 0.94 for CB, and the average MSD ranged from 0.16 mm for WM, to 0.97 mm for CB. Figure 7.3 illustrates multi-modal segmentation performance compared with manual segmentation.

7.5.2 Quantitative evaluation of the automatic segmentation in three-months follow-up scans

The scans with available manual segmentation were divided equally into training and test set based on the strength of scanner (3T and 1.5T) and the stroke size replaced by CSF graded visually. Six scans were acquired on 1.5T scanner and the remaining six scans on 3T scanner. The network was trained with six T2-weighted neonatal MRIs and tested on the remaining six scans. The average segmentation performance over test set is listed in Table 7.4. The average DC ranged from 0.88 for eCSF to 0.96 for CB, and the average MSD ranged from 0.10 mm for cGM to 0.58 mm for vCSF. Figure 7.4 illustrates examples of automatic segmentation.

	CB	eCSF	BGT	vCSF	WM	BS	cGM	Mean
DC	0.96	0.88	0.93	0.89	0.89	0.92	0.91	0.91
MSD	0.45	0.25	0.34	0.58	0.14	0.37	0.10	0.32

Table 7.4: Performance of brain tissue segmentation into seven tissue classes in three-months follow-up scans. The results are expressed as the mean Dice coefficient (DC) and the Mean surface distance (MSD) in mm.

7.5.3 Qualitative Evaluation

To evaluate segmentation in a larger set of patients, the multi-modal network for analysis of neonatal images and the single-modal network for analysis of follow-up images

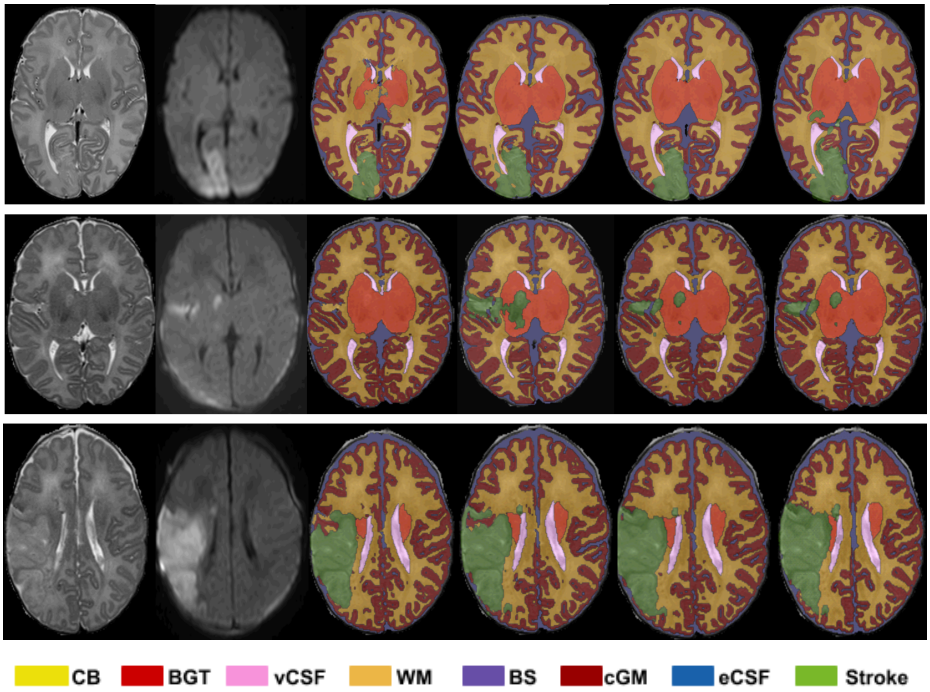


Figure 7.3: Examples of automatic ischemic stroke and brain tissue segmentation in neonatal MRI. A slice from T2-weighted MRI (first column); a slice from DWI (second column); segmentation obtained by two separate networks (third column); segmentation obtained by multi-channel network (fourth column); segmentation obtained by the proposed multi-modal segmentation network (fifth column); manual reference segmentation (last column).

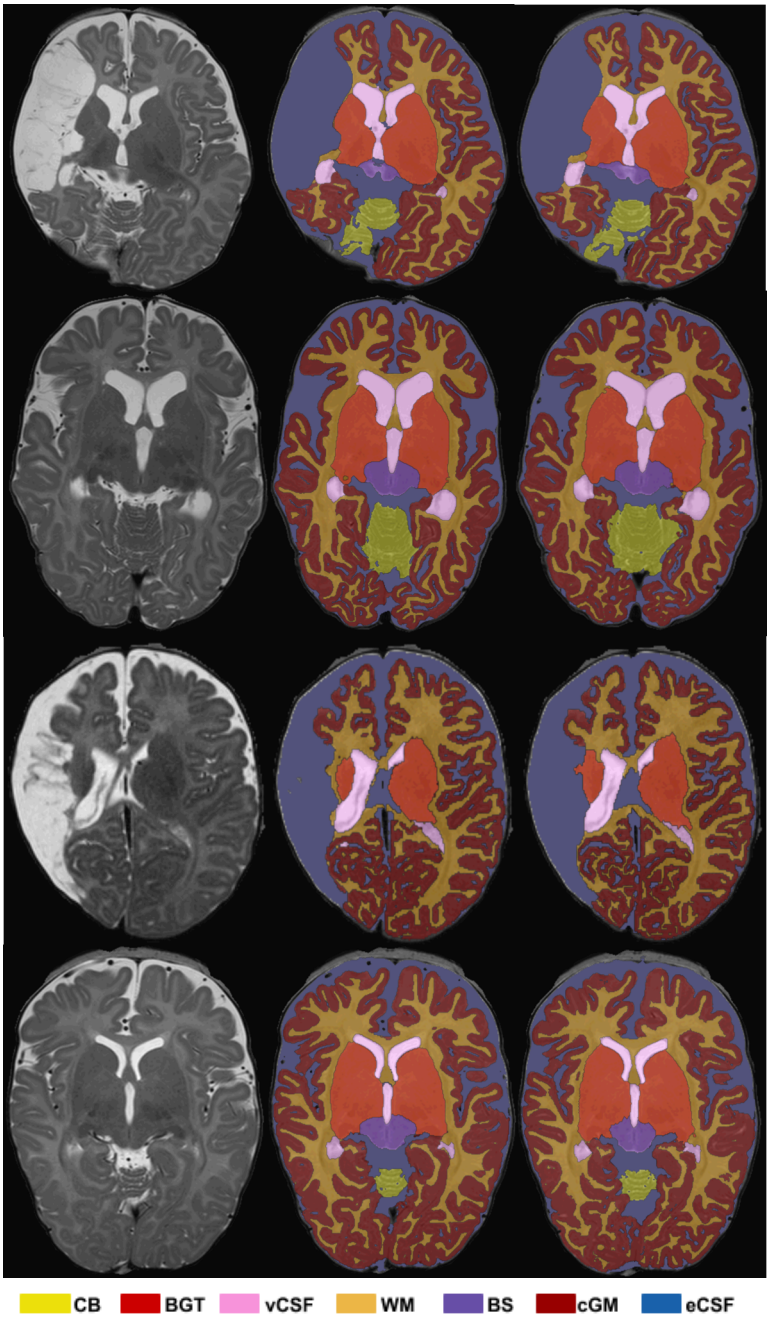


Figure 7.4: Examples of automatic brain tissue segmentations of three-months follow-up scans. A slice from T2-weighted neonatal MRI (first column); the obtained automatic segmentation (second column); manual reference segmentation (third column).

	All tissues	Stroke	All tissues ≤ 3-point	Stroke ≤ 3-point
Single-modal	4.4	-	11	-
Multi-modal	4.3	4.1	9	10

Table 7.5: Qualitative segmentation performance of neonatal and follow-up scans using 5-point Likert scale. The average scores for tissue types and stroke tissue are listed in first and second column, respectively. Number of scans in which tissue types overall and stroke tissue particularly are scored below 3 were listed in the third and forth column.

were applied to images of 58 infants. Given that the manual segmentation was not available for these, the results were evaluated visually as described in Section 7.4. The average scores on the 5-point Likert scale in neonatal scans and three-months follow-up are listed in Table 7.5. In neonatal scans, average assigned grade over all patients was 4.3 for brain tissues and 4.1 for stroke tissue. In total, 9/54 scans and 10/54 scans only were graded less than 4 for brain and stroke tissues, respectively. In the follow-up scans, average assigned grade for brain tissue segmentation over all patients was 4.4. In total 11/52 scans were graded less than 4. All of these scans had slices affected by motion artifacts that hampered segmentation. Figure 7.5 illustrates segmentation results obtained in scans acquired with 3T and 1.5T scanners.

7.5.4 Robustness with respect to imaging artifacts

Visual analysis revealed examples of images with artifacts that were accurately segmented and others that were not. Examples of multi-modal segmentation on images with susceptibility artifacts are illustrated in Figure 7.6. Illustrated DWI slices depict hyperintensity artifact in the occipital lobe of the brain which can mislead stroke diagnosis in areas with no infarct. This is a common artifact caused by an off-resonance field induced by differences in magnetic susceptibility at the air-tissue interface [156]. However, the proposed multi-modal network segmented these areas accurately as these artifacts were visible in the training set and this segmentation was graded 5.

Furthermore, examples of multi-modal segmentations in DWI with acutely low signal intensity are illustrated in Figure 7.7. In these examples, the multi-modal network under-segmented ischemic tissue due to low signal of intensity. This low intensity in DWI often appears in cases when ischemic lesions are small or when the MRI is acquired more than 7 days after birth, i.e. stroke onset.

Another common artifact in neonatal imaging is is caused by the movement of the infant during scanning. Figure 7.8 illustrates the segmentation result in T2-weighted MRI affected by motion artifact. The segmentation was challenging for tissue classes such as cGM and BGT, but stroke segmentation scored. Similar to neonatal MRI, motion artifacts in three-months follow-up scans are challenging for automatic segmentation. Examples of the single-modality network on scans with motion artifacts are

illustrated in Figure 7.9.

Figure 7.10 illustrates an automatic segmentation of eCSF and vCSF. These tissues have the same characteristics therefore, in a few examples where the vCSF was near the former ischemic tissue, the network segmented a few spots of vCSF in eCSF areas.

7.5.5 Multi-modal vs multi-channel vs two separated networks

To segment brain tissues and ischemic tissue from DWI and T2-weighted neonatal MRI, we proposed using multi-modal network. This network composes the features from both sequences and generates the segmentation. To evaluate advantages of the multi-modal network, we designed two experiments. In the first experiment, two separate networks with U-net architecture were trained. One network was trained on DWI to segment ischemic stroke tissue as a binary class and the other network was trained on T2-weighted scan to segment brain into seven tissue classes. The result of stroke segmentation masked the brain tissue segmentation i.e. in T2-weigted image brain voxels were replaced by ischemic tissue. The segmentation results are listed in Table 7.3. The results show that the average DC for all eight tissue classes decreased in comparison with the proposed multi-modal network. This demonstrates the benefit of combining these two segmentation tasks in one network.

In the second experiment, DWI and T2-weighted scans were given as two input channels to U-net architecture. Therefore, the network input comprised 2D slices of DWI and T2-weighted scan with $416 \times 416 \times 2$ input shape. The network has a single encoding path that encodes both DWI and T2-weighted scan together in the first layer. The segmentation results listed in Table 7.3 show that the segmentation performance decreased in comparison with multi-modal segmentation evaluated with both DC and MSD. However, the segmentation performance of this experiment outperforms the experiment using of two separate networks when evaluated using average DC over all tissue classes, but not using MSD. Figure 7.3 illustrates the automatic segmentation of three experiments compared with manual segmentation.

7.5.6 Comparison with state-of-the-art

We compared the performance of the proposed multi-modal segmentation method with the performance of published methods for brain tissue segmentation in neonates. Note that in the lack of stroke segmentation methods in infants, we can only compare segmentation results for tissue classes.(Table 7.6). The table shows that the performance of the proposed method is comparable to the performance of two top ranked methods in NeoBrainS12 challenge[2]. The multi-modal segmentation network generates more accurate segmentation over all tissue classes than two other methods. Given

[2]https://neobrains12.isi.uu.nl/

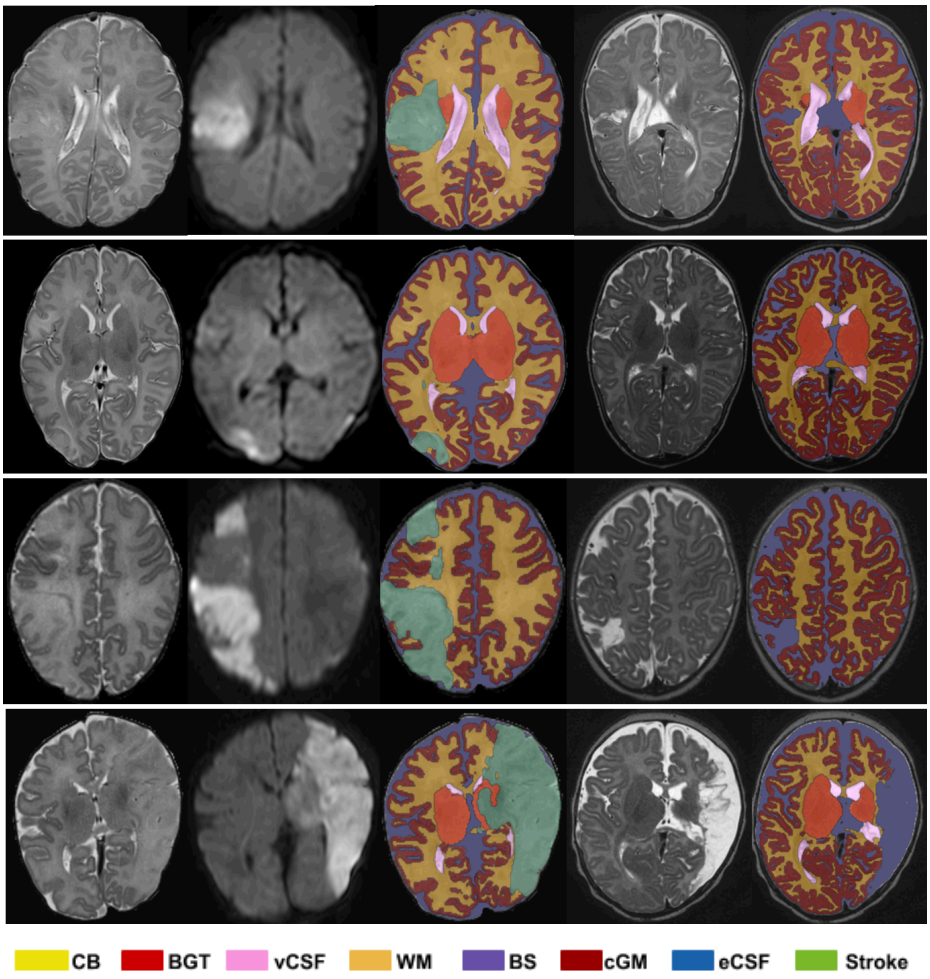


Figure 7.5: Examples of neonatal MRI with stroke in the perinatal scan and 3-months follow-up. A slice from T2-weighted perinatal scan (first column); the corresponding DWI slice (second column); automatic stroke and brain tissue segmentation obtained with the proposed multi-modal network (third column); T2-weighted follow-up scan in three months old neonate (fourth column); automatic brain tissue segmentation in the three months scan (last column). Two top rows are scanned with 3T strength field scanner and two bottom rows are scanned with 1.5 strength field scanner.

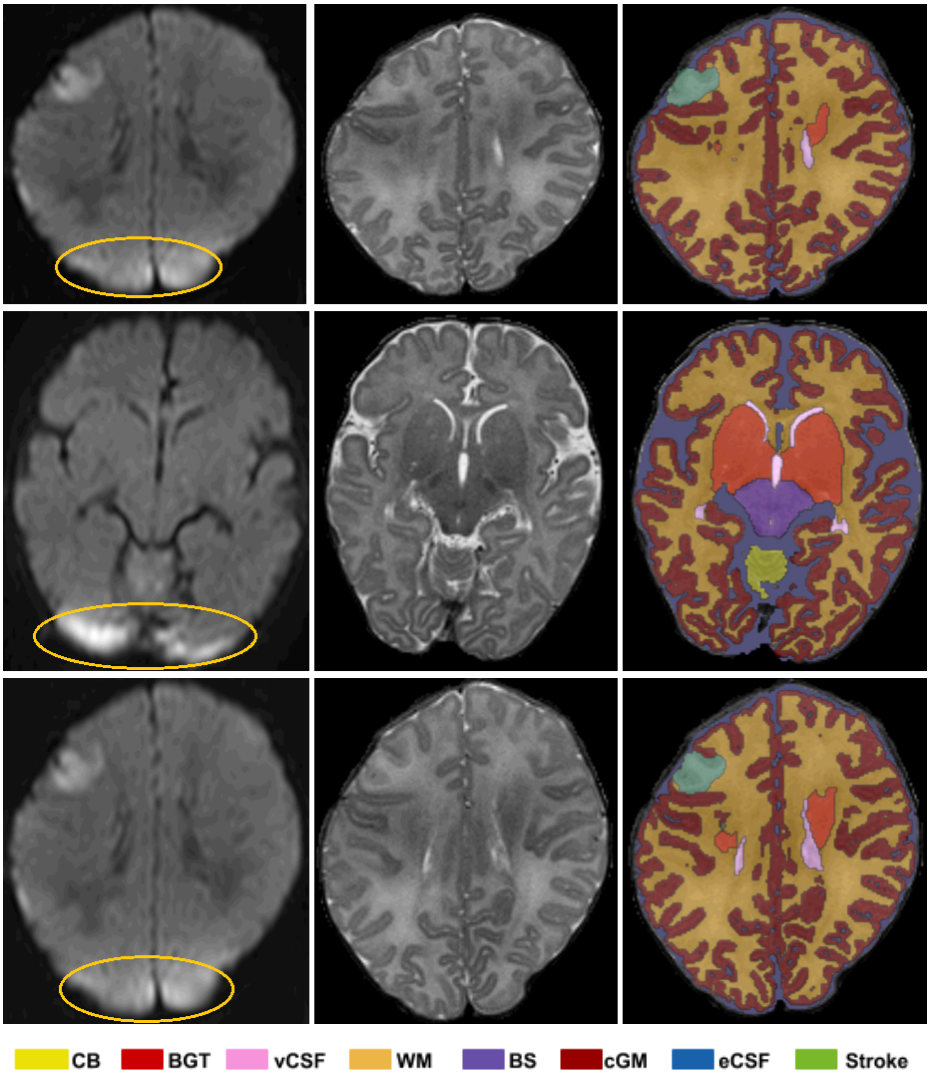


Figure 7.6: Examples of automatic segmentation in DWI slices with susceptibility artifacts. A slice from DWI with susceptibility artifacts in the occipital lobe (first column); the corresponding T2-weighted slice (second column); automatic stroke and brain tissue segmentation obtained with the proposed multi-modal network (third column). The artifacts are marked with yellow ellipse caused hyperintensity in DWI, however, these had no influence on the automatic segmentation result.

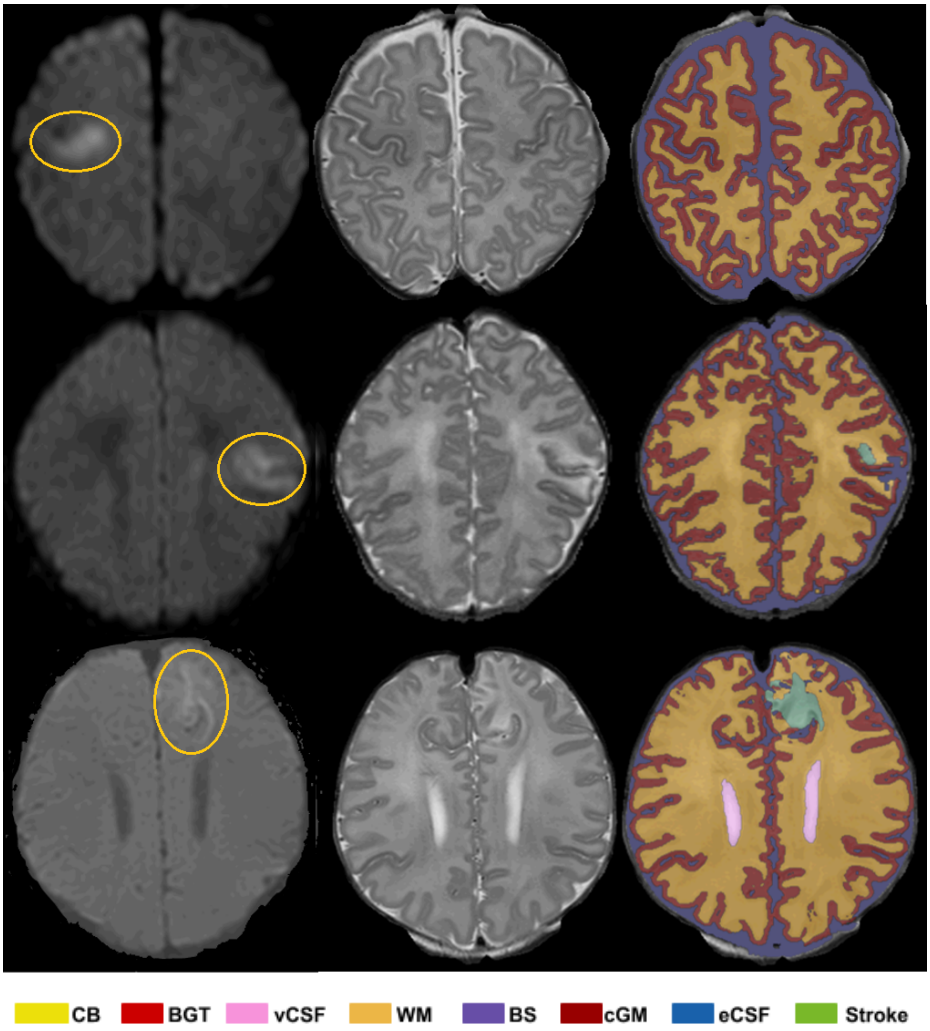


Figure 7.7: Examples of segmentation in DWI slices with low signal intensity. A slice from DWI with low signal intensity (first column); the corresponding T2-weighted MRI slice (second column); automatic stroke and brain tissue segmentation obtained with the proposed multi-modal network (third column). The ischemic tissue is marked with yellow ellipse. The automatic segmentation did not detect ischemic tissue (top row), and undersegmented ischemic tissue in the middle and bottom rows.

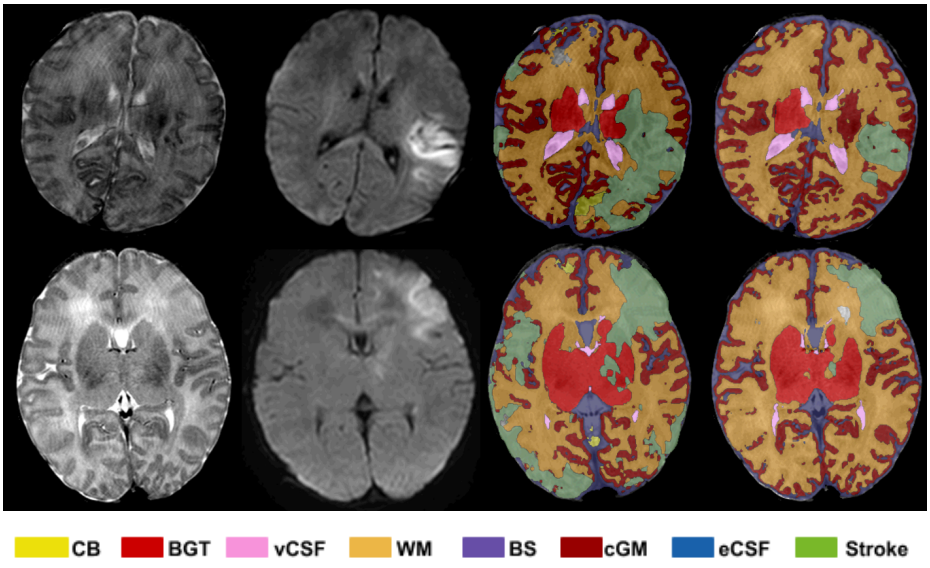


Figure 7.8: Examples of automatic segmentation in slices with motion artifacts. A slice from T2-weighted MRI with motion artifacts (first column); the corresponding DWI slice (second column); automatic brain tissue segmentation obtained with multi-channel network (third column); automatic brain tissue segmentation obtained with the proposed multi-modal network (forth column). Motion artifacts visualized only on T2 slices, the ischemic stroke segmentation generated by multi-modal network was affected less in compare with multi-channel network.

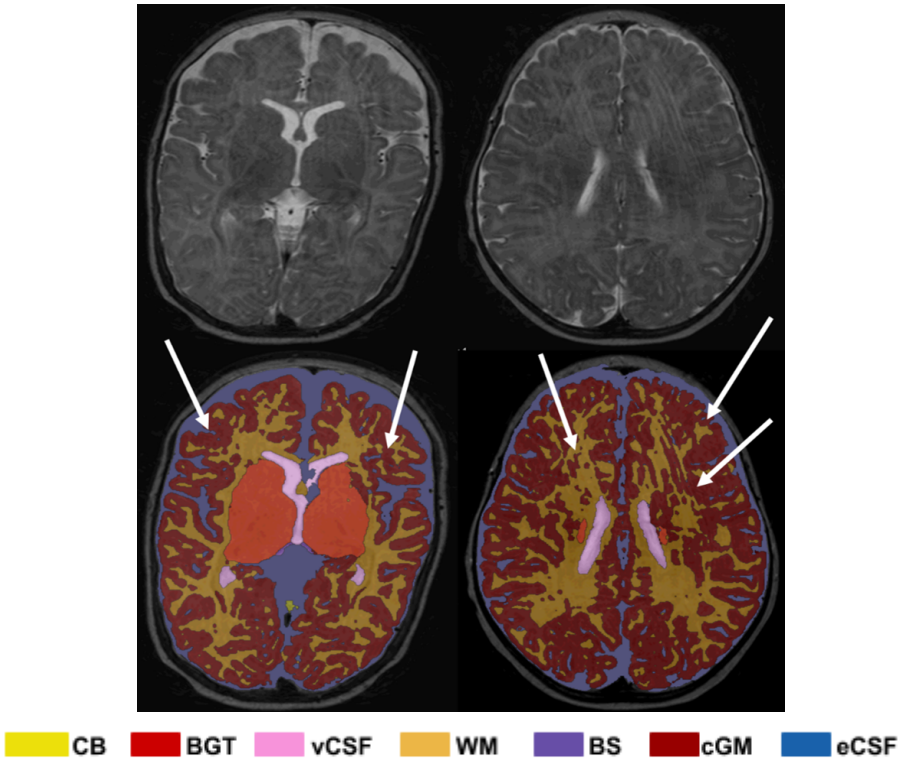


Figure 7.9: Examples of automatic segmentation in slices with motion artifacts. A slice from T2-weighted images with motion artifacts (top row); automatic brain tissue segmentation obtained with single-modal network (bottom row). The automatic method was challenged by motion artifacts, particularly in cGM. The segmentation errors in areas with motion are indicated by arrows.

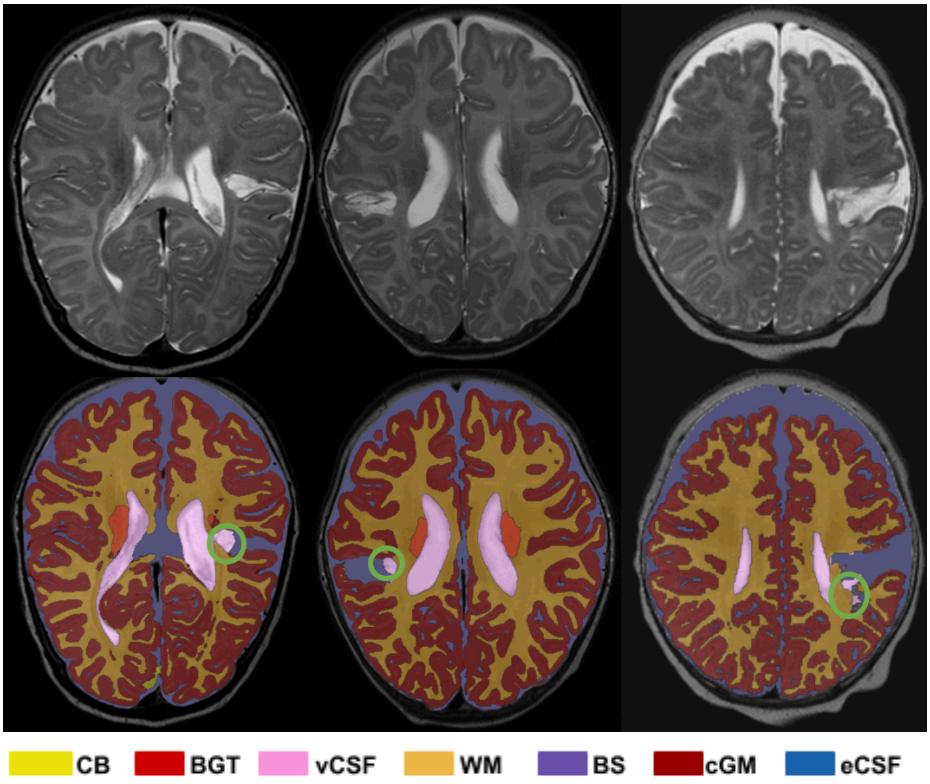


Figure 7.10: Examples of segmentation in the follow-up scans. A slice from T2-weighted image (top row); automatic brain tissue segmentation obtained with single-modal network (bottom row). The single-modal network segmented small part of eCSF as vCSF. The mistakes are indicated with green ellipses.

		CB	eCSF	BGT	vCSF	WM	BS	cGM	Stroke	Mean
Multi-modal	DC	0.94	0.84	0.92	0.86	0.91	0.94	0.86	0.83	0.89
	MSD	0.97	0.21	0.40	0.18	0.16	0.19	0.13	1.33	0.44
Moeskops et al. [13]	DC	0.93	0.83	0.91	0.81	0.93	0.85	0.87	–	0.87
	MSD	1.14	0.19	0.46	0.43	0.12	0.35	0.11	–	0.40
Wang et al. [142]	DC	0.92	0.79	0.92	–	0.92	0.83	0.86	–	0.87
	MSD	0.45	1.18	0.33	–	0.13	0.27	0.11	–	0.41

Table 7.6: Performance of the multi-modal network and two recent methods evaluated with the Dice coefficient and MSD. The previous methods were evaluated on NeoBrainS12 to challenge data in axial neonatal MRI without visible pathology. Hence, this comparison can be used as an indication only.

that the methods have been evaluated using different data sets this comparison provides an indication only.

7.6 Discussion

We presented automatic methods for the segmentation of ischemic and brain tissue classes in infants with PAIS. To the best of our knowledge, this is the first attempt to segment ischemic tissue as well as healthy brain tissue in MR scans of infants affected by stroke. The segmentation methods were applied to the neonatal MR images acquired after the acute phase following the stroke and about three months after the stroke. In the neonatal scans, ischemic tissue is visible in DWI and brain tissues are best shown in T2. Therefore, to segment brain tissues and stroke in neonatal MRI from DWI and T2-weighted scan, a multi-modal network was proposed. In three-months follow-up scans, CSF cysts replace ischemic tissue which is not visible in DWI anymore. Therefore, a single-modal network segmented brain tissues in three-months follow-up scans. The segmentation performance was evaluated quantitatively using DC and MSD in a data set of a limited size, and qualitatively using Likert scale in a large set. The results demonstrate that the proposed multi- and single-modal networks lead to accurate segmentations of the stroke lesions and brain tissue classes.

The population included in this study has a large variation in stroke volume. These variations are often an obstacle to automatic analysis methods. To overcome this, the multi-modal network was trained and tested on data which was visually balanced based on tissue volume distribution. Volumetric balancing was possibly important because of the small training set size but that availability of large set that would then necessarily contain variations in stroke lesion characteristics would circumvent the problem.

The ablation study shows that proposed multi-modal network outperforms segmentation using two separate networks and a multi-channel network. The multi-modal network encoded features from both modalities and therefore, generated more accurate segmentation in comparison with two separate networks. Moreover, having

two separate paths for each MRI modality specialized kernels to specific modality and therefore, the multi-modal segmentation method generated more accurate segmentation that the multi-channel network. The proposed multi-modal network can be easily adjusted for application in different image analysis tasks that require information from multi-modal MR and potentially other data. Furthermore, the network can be easily customized to accommodate more than two inputs straightforward by addition of extra encoding branches.

The proposed single-modal segmentation network accurately segments the brain into seven tissue classes. The large stroke volume variation in neonatal MRI results in a large tissue volume variation in three-month follow-up scans. This challenges the automatic segmentation of cGM, eCSF and WM therefore, the training set and evaluation set were roughly balanced based on tissue volume distribution by visual evaluation. Volumetric balancing was possibly important because of the small training set size but that availability of large set that would then necessarily contain a large variation caused by CSF cysts replaced of ischemic tissue. The variations in remaining tissue characteristics would circumvent the problem. However, in few cases where CSF cysts were near ventricle the network segmented cysts as vCSF (see Figure 7.10). The tissue characteristic of vCSF and eCSF are the same and we expect that the network distinguishes them based on the difference in their shape and location. However, in 3-months follow-up the scans the location and shape of eCSF and vCSF vary dramatically. This may be improved by increasing the number of training data with representative examples.

The population included in this study was scanned with two different MRI scanners, 24 infants were scanned with a scanner with 1.5T field strength and 40 infants with 3T field strength scanner. The proposed single-modality and multi-modal networks were trained on a well-balanced combination of 3T and 1.5T scans and the evaluation results show accurate segmentation performance in both set of scans.

Furthermore, as shown in Figure 7.7, multi-modal network was able to segment images in presence of susceptibility artifacts even though, these artifacts cause hyperintensity in DWI that can be confused with the ischemic tissue. This might be due to the presence of such artifacts in the training set, enabling, the network to distinguish between artifact hyperintensity and ischemic hyperintensity.

The segmentation performance in both single-modal and multi-modal networks was hampered by motion artifact particularly in cGM (see Figure 7.8). Motion artifacts can be removed using image-to-image-translation between scans affected by motion and scans without motion artifacts [144]. Joint correction of motion or other image artifacts and image segmentation may be an interesting direction for future work to improve segmentation performance.

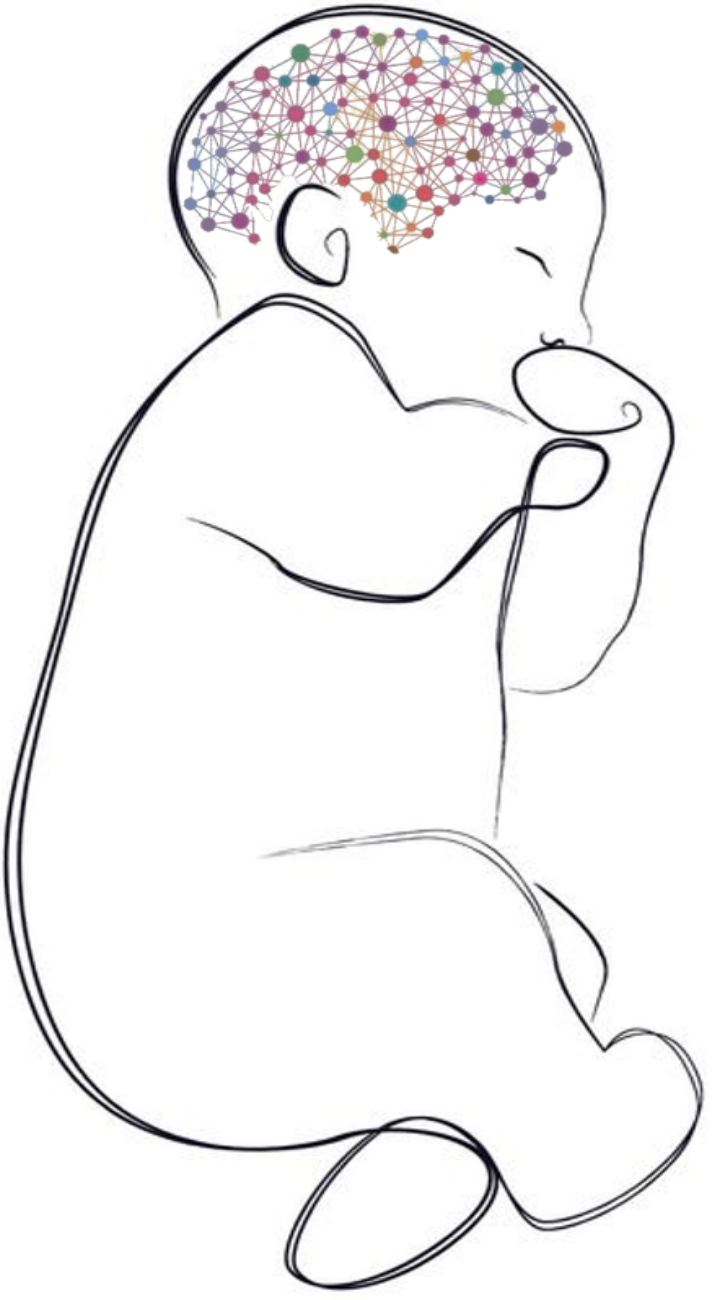
Our method was trained with only 6 and 5 scans for first and second scan segmentation tasks, respectively while previous deep learning methods was trained on 90 subjects to segment ischemic stroke from DWI [149]. Given that generating brain

tissues by manual annotations is extremely time-consuming, reducing the number of the needed annotated training scans substantially reduces the required annotation effort and associated costs. Using generative models to synthesize pathology on images from currently available manually segmented scans would increase training data size. Specifically, using conditional generative models to control pathology characteristics, such as size and shape of abnormality, could generate a large training set from a limited number of available manual segmentation. This may improve segmentation performance and can be a direction for future work.

Analyzing ischemic and brain tissue volumes in the neonatal scans and in three months follow-up scans allows investigating the effect of the neurodegenerative intervention on brain growth. Several studies previously showed contradictory results on the effect of neurodegenerative treatment namely rhEPO on ischemic tissue [56, 157, 158]. The proposed methods can potentially quantify the effect of rhEPO treatment in perinatal arterial ischemic stroke, which would be of great clinical interest. The volumetric measurements of brain tissue classes in neonatal MRI can be compared between infants treated with rhEPO and untreated infants. However, this requires further development of the an automatic method to split right and left hemispheres. Moreover, the analyses on effect of ehEPO can be extended using other imaging modalities such as diffusion tensor imaging and comparing fractional anisotropy index in neonates treated with rhEPO vs those who were not treated. Finally, functional analysis including baby's outcomes in older age would eventually enable relating brain tissue and stroke lesion characteristics in MRI and long term treatment effects.

7.7 Conclusion

We presented automatic methods for segmentation of ischemic stroke and brain tissue classes in MRI of infants diagnosed with PAIS. The methods are based on convolutional neural networks with U-net-like architecture that take single- or multi-modal MR scans as input. The networks obtained accurate segmentation and the results can be potentially used to evaluate the effect of neurogenerative interventions on perinatal arterial ischemic stroke.



Chapter 8

Summary and discussion

The previous chapters describe methods for automatic segmentation of the brain in neonatal and fetal MRI. This chapter summarizes the methods and findings and discusses implications and possible directions for future research.

8.1 Summary

CHAPTER 2 describes a method for automatic extraction of the intracranial volume in fetal and neonatal MRI using convolutional neural networks. A network using U-net architecture was trained with a combination of fetal scans and preterm born neonatal scans acquired in axial, coronal, and sagittal orientations. The method is evaluated on scans with different imaging parameters (field strength, image acquisition plane, image resolution) and different pathologies (posthemorrhagic ventricular dilatation (PHVD), stroke, asphyxia, Down syndrome). We showed that our method can reliably segment the intracranial volume (ICV) from a diverse set of scans including fetal and neonatal MRI with pathology.

CHAPTER 3 describes a method for the segmentation of brain tissues in fetal MRI using a convolutional neural network. The method first identifies the ICV from the fetal MRI using the method proposed in **CHAPTER 2**. Subsequently, the second network with an identical network architecture segments the brain into seven tissue classes, notably: cerebellum (CB), basal ganglia and thalami (BGT), ventricular cerebrospinal fluid (vCSF), white matter (WM), brain stem (BS), cortical gray matter (cGM) and extracerebral cerebrospinal fluid (eCSF). A common artifact in fetal MRI is the intensity inhomogeneity artifact caused by the fetal movements during the scanning, which complicates automatic segmentation. Additionally, generating manual segmentation on slices with artifacts is time consuming and in some scans infeasible. Hence, in this chapter, we proposed a data augmentation technique that synthesizes slices with intensity inhomogeneity artifacts to improve segmentation robustness. The images with synthesized artifacts were used during the training. The method was evaluated on fetal and neonatal MRI scans showing intensity inhomogeneity artifacts. We demonstrate that segmentation using the proposed data augmentation leads to accurate segmentation of the brain tissue classes.

Another common artifact in neonatal MRI is caused by the infant's motion during scanning, which makes brain tissue segmentation more difficult. **CHAPTER 4** introduces an automatic method for correction of such artifacts in reconstructed brain MR scans of preterm infants using generative adversarial network (GAN) with cycle consistency. The method was trained to transform slices affected by motion artifacts into slices without artifacts, and vice versa. To evaluate whether motion correction results in more accurate segmentation, the images were segmented into eight tissue classes using a convolution neural network. The segmentation network was augmented during training using the cycleGAN that synthesizes slices with artifacts from slices without artifacts. We demonstrate that the proposed correction for motion artifacts improves

image quality and allows accurate automatic segmentation of brain tissue classes in brain MRI of infants. In an ablation study, we also show that the proposed data augmentation improves segmentation results.

CHAPTER 5 analyzes longitudinal brain volume changes in infants with antenatal congenital heart disease (CHD). Infants with CHD undergo heart surgery within 30 days after birth. They are scanned with MRI in utero and neonatally (both preoperatively and postoperatively). To investigate the association between fetal and neonatal brain volumes, the brain tissue volumes were measured using automatic segmentation developed in **CHAPTER 2** and **CHAPTER 3**. Furthermore, the correlation between fetal brain volumes and neonatal brain volumes with ischemic injury was analyzed. This study reveals that fetal brain volumes positively correlated with neonatal brain volumes in critical CHD, with a negative correlation between fetal brain volumes and neonatal ischemic injury. Thus, fetal brain imaging has the potential to provide early neurological biomarkers.

CHAPTER 6 investigates the effects of early and late intervention for PHVD on additional brain injury and ventricular volume using term-equivalent age MRI. In this study, preterm infants with PHVD were randomized into two groups, early and late intervention. The Kidokoro Global Brain Abnormality Score and the frontal and occipital horn ratio were measured. Moreover, automatic segmentation using the method presented in **CHAPTER 3** was used for volumetric analysis. The findings demonstrate more brain injury and larger ventricular volumes in the late intervention group. These results support the positive effects of early intervention for PHVD.

CHAPTER 7 introduces automatic methods to segment brain tissues in infants with perinatal stroke. MR scans of infants suspected of having a stroke are acquired after the acute phase, and three months later as part of the clinical follow-up procedure. In the first scan, brain tissues are visible on T2-weighted MRI and ischemic stroke tissue is visible on DWI. To develop a method that segments brain tissues and stroke together, a multi-modal network extracting features from both T2 and DWI is needed. A network with two encoding branches that simultaneously encode DWI and T2 was developed to segment the brain into seven tissue classes and an ischemic stroke class. In the 3-months follow-up scan, ischemic tissue is not visible as it has been replaced by CSF. Accordingly, a neural network to segment the brain into seven tissue classes was applied on the T2-weighted MR images. The acute phase and 3-months networks are both convolutional neural networks with a U-net architecture. The multi-modal and single-modal networks were evaluated quantitatively in 4 and 6 scans respectively. Furthermore, these networks were evaluated qualitatively using Likert scale in 48 scans. The method could be used to evaluate the effects of ischemic tissue treatment such as administration of erythropoietin.

8.2 Discussion and future directions

This thesis presents automatic methods to segment brain MRI to perform volumetric measurements that provide the basis for evaluation of brain development in fetuses and neonates from the last trimester of pregnancy until 40 weeks of gestation. These automatic methods do not only accelerate the extremely time-consuming manual segmentation task in large-scale studies but may also be implemented in clinical practice directly to aid identifying infants at higher risk of neurodevelopmental impairment.

The methods proposed in this thesis are mainly based on supervised deep learning which are prone to overfitting on training data. Evaluating these method on external validation set would therefore, able to demonstrate the generalizability of these methods. In Chapter 1 the developed method was evaluated on diverse data sets including infants with various types of pathology and scans acquired in an external dataset, other chapters used data from a single hospital. A next step toward routine use of these methods would be validating on data with a wide variation in imaging and patient characteristics. Future work should, therefore, extend the evaluation to data from different centers acquired with different scanner types and imaging parameters.

However, the strength of supervised learning is its versatility to new data. In other words, the segmentation methods presented in this thesis may be adapted to different datasets by retraining with representative data without changing the method. Chapter 2 demonstrates that the network aimed at segmenting ICV from T2-weighted MRI is not only able to learn from each set separately but that it is also able to learn from a combination of fetal and preterm born neonatal scans acquired in axial, coronal, and sagittal orientations. This method is able to outperform the commonly used tools without need of developing data-specific method for each dataset. The results showed that despite the differences in image acquisition, image orientation, and brain morphology, fetal and neonatal scans share common features that improve the ability of the network to generalize, making it more robust and compensating for the lack of representative data.

Several segmentation tasks in this thesis are based on U-net architecture. U-net is a fully convolutional network well-known in medical image analysis for achieving fast and accurate segmentation [52, 159, 160]. Chapters 3 and 4 describe augmentation methods using U-net for brain tissue segmentation in neonatal and fetal MRI. Chapter 7 demonstrates a multi-modal variation of U-net to segment brain and ischemic stroke from T2-weighted MRI and DWI. These chapters demonstrate that the same architecture with necessary adjustments is able to perform different tasks and obtain good performance if data augmentation, training parameters and loss functions have been chosen well.

Even smaller deep learning network such as U-net are prone to overfitting largely for lack of representative training data. Augmentation is a well-known strategy addressing this issue by synthesizing data where variation can be introduced in the

training set. The standard augmentation techniques including random rotation, random rigid transformation, and random flipping in vertical direction were applied to images in this thesis. In addition to these, in Chapters 3 and 4 augmentation methods are proposed that were customized to scans with artifacts. In Chapter 3 intensity inhomogeneity artifacts were synthesized using mathematical models and in Chapter 4 motion artifacts were synthesized using the generative model on slices without artifacts. In this way the segmentation network could generate more accurate segmentations in the presence of artifacts. However, each of these methods is designed for one type of artifact only. Using a unified generative adversarial network for multi-domain image-to-image translation would allow to remove multiple types of artifacts [161]. Combining these artifacts in one network may enable the network perform more efficiently and an interesting direction for future research.

In Chapter 3 we presented an automatic method for segmentation of fetal MRI into seven brain tissue classes. Previous methods for segmentation of fetal MRI often reconstruct the brain from several MRI scans acquired in axial, coronal and sagittal planes [67, 68, 71]. Acquiring several scans from the fetus is time consuming and is not available in many cases. We proposed a method which can be applied to 2D slices of fetal MRI acquired in the coronal plane and can potentially be applied to other image planes. Combining MRIs acquired in different planes can enhance segmentation performance and compensate the information lost due to artifacts. Thus, generating a multi-task model to segment and reconstruct scans from MRIs acquired in different planes may improve the performance of both segmentation and image reconstruction and therefore, would be an interesting direction for future research.

The automatic segmentation methods introduced in Chapters 2, 3, 4, and 7 were used for clinical research. In Chapter 5, the brain development in infants with critical heart disease was quantified in a longitudinally scanned cohort. The automatic segmentation was applied to a large set of fetal MRI, and neonatal preoperative and postoperative MRI. In Chapter 6, the effects of early intervention in infants with PHVD were quantified using automatic measures of brain tissue volume. The method developed in Chapter 7 can be used to evaluate the effect neurogenerative interventions such as recombinant human erythropoietin (EPO) on brain development in infants with perinatal stroke. Brain tissue segmentation quantifies tissue volumes and thus is a requirement to measure cortical surface area, cortical folding, and cortical thickness that are commonly used for a quantitative description of the brain and its development. Extracting these quantitative features from a large set of (temporal) imaging data would model brain development in different stages. This thesis provides the foundation for such studies. Without reliable automatic segmentation and quantification methods, clinical evaluation in a large set of perinatal image data would hardly be feasible, if at all.

Recent achievements in machine learning and deep learning are expected to revolutionize modern health care in the coming years. Many methods perform comparably

to or better than human experts for a specific task. Repetitive tasks such as, but not limited to, brain tissue segmentation and brain injury detection in MRI scans could be replaced by fully automatic or semi-automatic approaches. The automatic segmentation techniques proposed in this thesis eventually can be directly implemented in the clinic and thereby clinicians are able to monitor brain development quantitatively. This may assist in making a more accurate, more reliable diagnosis and reduce human errors after long hour works and possibly at lower costs. However, complex medical tasks still benefit from human expertise. The current medical technology literature supports the concept of human-in-the-loop that outperforms the effort of human or machine alone [162]. This encourages using semi-automatic or automatic methods with human supervision. Recently, many methods were proposed to investigate the uncertainty of a trained network for a certain task. These techniques allow involving humans-in-the-loop to annotate correctly the uncertainty of the network. Furthermore, the network can retrain on adjusted labeled data to 'learn from mistakes'. This would be a successful example of human-machine collaboration which accelerates the adaptation to the new paradigm.

Nederlandse samenvatting

In het laatste trimester van de zwangerschap, d.w.z. tussen de 30 en 40 weken zwangerschap, vinden belangrijke neurologische ontwikkelingen plaats. Het volgen van de hersenontwikkeling in deze periode helpt clinici bij het detecteren, en indien nodig, behandelen van stoornissen. Magnetic resonance imaging (MRI) wordt steeds vaker gebruikt om zowel in utero als ex utero de hersenen van baby's te monitoren, om letsels die een langetermijneffect kunnen hebben op de neurocognitieve en motorische ontwikkeling op te sporen. Echter is men niet altijd in staat de ontwikkeling van de hersenen voldoende te kwantificeren door scans visueel te beoordelen. Dit maakt de evaluatie van langetermijneffecten niet eenvoudig. Een veelgebruikte methode om het hersenvolume te meten is segmentatie, oftewel het splitsen van het beeld in segmenten om verschillende anatomische structuren aan te geven. Niettemin is het buitengewoon tijdrovend en subjectief om een groot aantal MRI scans handmatig te segmenteren. Automatische segmentatie is daarom een essentieel hulpmiddel voor het kwantificeren van de ontwikkelingen van de hersenen en een vereiste voor andere metingen, zoals metingen van de morfologie. Dientengevolge kan automatische segmentatie meehelpen om op een betrouwbaardere wijze letsels vast te stellen en nauwkeuriger een verband te leggen met cognitieve uitkomsten.

HOOFDSTUK 2 beschrijft een methode voor het automatisch meten van het intracraniële volume op een foetale of neonatale MRI scan met behulp van een convolutioneel neuraal netwerk. Een netwerk met de architectuur van een U-net is getraind met een combinatie van foetale scans en scans van prematuur geboren baby's. Scans zijn gemaakt met axiale, coronale en sagittale oriëntaties. De methode is geëvalueerd met scans die gemaakt zijn met verschillende beeldvormingsparameters (veldsterkte, beeldvormingsvlak, resolutie) en bevatten verschillende ziektebeelden (posthemorrhagische ventriculaire dilatatie (PHVD), beroerte, zuurstoftekort, syndroom van Down). In dit hoofdstuk tonen wij aan dat onze methode het intracraniele volume (ICV) op een betrouwbare manier kan segmenteren in een diverse set van scans, bestaande uit foetale MRI scans en scans van pasgeboren baby's met een pathologie.

HOOFDSTUK 3 beschrijft een methode voor de segmentatie van hersenstructuren in foetale MRI doormiddel van een convolutioneel neuraal netwerk. Eerst identificeert de methode het ICV in de scan door gebruik te maken van de methode voorgesteld in **HOOFDSTUK 2**. Vervolgens segmenteert een tweede netwerk, wat dezelfde architectuur heeft, zeven verschillende weefselklassen in de hersenen. Dit zijn het cerebellum (CB), de basale ganglia en thalami (BGT), ventriculair hersenvocht (vCSF), witte stof (WM),

de hersenstam (BS), corticale grijze stof (cGM) en extracerebraal hersenvocht (eCSF). Een veelvoorkomend artefact bij foetale MRI is intensiteits inhomogeniteit, dat wordt veroorzaakt door beweging van de foetus tijdens het scannen en de automatische segmentatie bemoeilijkt. Daarnaast is het handmatig segmenteren van scan coupes met artefacten tijdrovend en soms zelfs onmogelijk. Daarom, om de robustheid van de segmentatietechniek te verbeteren, stellen we in dit hoofdstuk een data augmentatietechniek voor die scan coupes met intensiteits inhomogeniteits artefacten synthetiseert. We laten zien dat segmentatie met de voorgestelde data augmentatie-techniek leidt tot in een nauwkeurige segmentatie van de hersenweefselklassen.

Een ander veelvoorkomend artefact bij MRI scans van pasgeboren baby's wordt veroorzaakt door beweging van het kind tijdens het scannen, waardoor de segmentatie van hersenweefsel in de scan wordt bemoeilijkt. **HOOFDSTUK 4** introduceert een automatische methode voor het corrigeren van dergelijke artefacten in MRI scans van de hersenen van premature baby's, en maakt gebruik van een generative adversarial network met cyclusconsistentie (CycleGAN). De methode is getraind om scan coupes met bewegingsartefacten te veranderen in scan coupes zonder artefacten, en omgekeerd. Om te evalueren of deze bewegingscorrectie leidt tot een nauwkeuriger segmentatie resultaat, segmenteren we met een convolutioneel neuraal netwerk acht weefselklassen in scans. Tijdens het trainen is het segmentatie netwerk geaugmenteerd met behulp van de CycleGAN die scan coupes met artefacten synthetiseert uit scan coupes zonder artefacten. In dit hoofdstuk laten we zien dat de voorgestelde correctie voor bewegingsartefacten de beeldkwaliteit verbetert en leidt tot een nauwkeurige segmentatie van hersenweefselklassen in MRI scans van de hersenen van baby's. Daarnaast laten we in een ablatie-studie zien dat de beschreven data augmentatie techniek de segmentaties verbetert.

In **HOOFDSTUK 5** analyseren we veranderingen in het hersenvolume van baby's met prenatale aangeboren hartziekte (CHD) over de tijd. Baby's met CHD ondergaan binnen 30 dagen na de geboorte een hartoperatie. Ze krijgen een MRI scan als ze nog in de baarmoeder zijn en na de bevalling (zowel preoperatief als postoperatief). Om de associatie tussen foetaal en neonataal hersenvolume te onderzoeken, zijn de volumes van hersenweefsel gemeten met behulp van de automatische segmentatie methode, beschreven in **HOOFDSTUK 2** en **HOOFDSTUK 3**. Daarnaast is de correlatie geanalyseerd tussen foetale hersenvolumes en neonatale hersenvolumes met ischemisch letsel. Deze studie laat zien dat foetale hersenvolumes positief gecorreleerd zijn met neonatale hersenvolumes in het geval van kritieke CHD, met een negatieve correlatie tussen foetale hersenvolumes en neonataal ischemisch letsel. Beeldvorming van de feutale hersenen laat dus potentieel zien om neurologische biomarkers te verschaffen voor vroegtijdige diagnose.

In **HOOFDSTUK 6** onderzoeken we het effect van vroegtijdige en late interventie voor PHVD op bijkomend hersen letsel en het ventriculair volume, gebruikmakend van MRI scans die gemaakt zijn op de geboorte-equivalente leeftijd. In deze studie zijn

vroeggeboren baby's met PHVD in twee groepen gerandomiseerd: de vroegtijdige en de late interventie groep. De Kidokoro Global Brain Abnormality Score is gemeten en de ratio tussen de frontale en occipitale hoorn is bepaald. Verder is de automatische segmentatie methode, beschreven in **HOOFDSTUK 3**, gebruikt voor analyse van het volume. De bevindingen laten meer hersenletsel en grotere ventriculaire volumes zien in de late interventie groep. Deze resultaten ondersteunen de positive effecten van vroegtijdige interventie voor PHVD.

In **HOOFDSTUK 7** introduceren we een automatische methode voor het segmenteren van hersenstructuren in baby's met een perinatale beroerte. Als onderdeel van de klinische follow-up procedure, wordt bij verdenking van perinatale beroerte een MRI scan gemaakt na de acute fase en drie maanden later. In de eerste MRI scan is het hersenweefsel zichtbaar op het T2-gewogen beeld en de beroerte op het diffusie gewogen beeld (DWI). Om een methode te ontwikkelen die de hersenstructuren en de beroerte tegelijkertijd kan segmenteren, is een multimodaal netwerk nodig dat eigenschappen kan extraheren uit zowel het T2- als het DWI-beeld. Hiervoor is een netwerk ontworpen met twee coderende takken die tegelijkertijd het T2- en het DWI-beeld coderen om zo de hersenen in zeven hersenstructuur klassen en een beroerte klasse te segmenteren. In de follow-up scan is de beroerte niet meer zichtbaar omdat deze vervangen is door CSF. Daarom is het T2-gewogen beeld in zeven hersenstructuurklassen gesegmenteerd met een neuraal netwerk. Zowel het acute-fase-netwerk als het follow-up-netwerk zijn convolutionele neurale netwerken met de architectuur van een U-net. Het multi-modale netwerk en het single-modale netwerk zijn doormiddel van 4 en 6 scans kwantitafief geëvalueerd. Daarnaast zijn de netwerken met behulp van 48 scans geëvalueerd door gebruik te maken van de Likert schaal. De resultaten laten zien dat deze methode gebruikt zou kunnen worden om de effecten van behandeling van ischemische weefsel, zoals het toedienen van erytropoëtine, te evalueren.

Bibliography

1. I. Kostović and N. Jovanov-Milošević. "The development of cerebral connections during the first 20–45 weeks' gestation", *Seminars in Fetal and Neonatal Medicine*, vol. 11 (2006), pp. 415–422 (cited on pp. 10, 69).

2. K. Zilles, N. Palomero-Gallagher, and K. Amunts. "Development of cortical folding during evolution and ontogeny", *Trends in Neurosciences*, vol. 36 (2013), pp. 275–284 (cited on p. 10).

3. M. Schaer, M. B. Cuadra, N. Schmansky, B. Fischl, J.-P. Thiran, and S. Eliez. "How to measure cortical folding from MR images: A step-by-step tutorial to compute local gyrification index", *Journal of Visualized Experiments* (2012) (cited on p. 10).

4. I. Adams-Chapman, N. I. Hansen, B. J. Stoll, R. Higgins, et al. "Neurodevelopmental outcome of extremely low birth weight infants with posthemorrhagic hydrocephalus requiring shunt insertion", *Pediatrics*, vol. 121 (2008), pp. e1167–e1177 (cited on pp. 10, 93, 103).

5. J. Lee, L. A. Croen, C. Lindan, K. B. Nash, C. K. Yoshida, D. M. Ferriero, A. Barkovich, and Y. W. Wu. "Predictors of outcome in perinatal arterial stroke: a population-based study", *Annals of neurology*, vol. 58 (2005), pp. 303–308 (cited on p. 10).

6. B. S. Peterson, B. Vohr, L. H. Staib, C. J. Cannistraci, A. Dolberg, K. C. Schneider, K. H. Katz, M. Westerveld, S. Sparrow, A. W. Anderson, et al. "Regional brain volume abnormalities and long-term cognitive outcome in preterm infants", *Jama*, vol. 284 (2000), pp. 1939–1947 (cited on p. 10).

7. B. S. Peterson, A. W. Anderson, R. Ehrenkranz, L. H. Staib, M. Tageldin, E. Colson, J. C. Gore, C. C. Duncan, R. Makuch, and L. R. Ment. "Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants", *Pediatrics*, vol. 111 (2003), pp. 939–948 (cited on p. 10).

8. K. Keunen, K. Kersbergen, F. Groenendaal, I. Išgum, L. De Vries, and M. Benders. "Brain tissue volumes in preterm infants: prematurity, perinatal risk factors and neurodevelopmental outcome: a systematic review", *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 25 (2012), pp. 89–100 (cited on p. 10).

9. "World health organization(WHO)", *https://www.who.int/news-room/fact-sheets/detail/preterm-birth* () (cited on p. 10).

10. J. J. Volpe. "Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances", *The Lancet Neurology*, vol. 8 (2009), pp. 110–124 (cited on pp. 10, 102).

11. H. Kidokoro, J. Neil, and T. E. Inder. "New MR imaging assessment tool to define brain abnormalities in very preterm infants at term", *American Journal of Neuroradiology*, vol. 34 (2013), pp. 2208–2214 (cited on pp. 10, 95).

12. P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Išgum. "Automatic segmentation of MR brain images with a convolutional neural network", *IEEE Transactions on Medical Imaging*, vol. 35 (2016), pp. 1252–1261 (cited on pp. 11, 17, 34, 38, 46, 53, 58, 60, 109).

13. P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. Benders, and I. Išgum. "Automatic segmentation of MR brain images with a convolutional neural network", *IEEE Transactions on Medical Imaging*, vol. 35 (2016), pp. 1252–1261 (cited on pp. 11, 80, 95, 109, 128).

14. W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen. "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation", *NeuroImage*, vol. 108 (2015), pp. 214–224 (cited on p. 11).

15. T. E. Inder, P. S. Huppi, S. Warfield, R. Kikinis, G. P. Zientara, P. D. Barnes, F. Jolesz, and J. J. Volpe. "Periventricular white matter injury in the premature infant is followed by reduced cerebral cortical gray matter volume at term", *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 46 (1999), pp. 755–760 (cited on p. 17).

16. J. Dubois, M. Benders, A. Cachia, F. Lazeyras, R. Ha-Vinh Leuchter, S. Sizonenko, C. Borradori-Tolsa, J. Mangin, and P. S. Hüppi. "Mapping the early cortical folding process in the preterm newborn brain", *Cerebral Cortex*, vol. 18 (2007), pp. 1444–1454 (cited on p. 17).

17. P. Moeskops, M. J. Benders, K. J. Kersbergen, F. Groenendaal, L. S. de Vries, M. A. Viergever, and I. Išgum. "Development of cortical morphology evaluated with longitudinal MR brain images of preterm infants", *PloS one*, vol. 10 (2015), p. e0131552 (cited on p. 17).

18. P. Moeskops, I. Išgum, K. Keunen, N. H. Claessens, I. C. Haastert, F. Groenendaal, L. S. Vries, M. A. Viergever, and M. J. Benders. "Prediction of cognitive and motor outcome of preterm infants based on automatic quantitative descriptors from neonatal brain images", *Scientific Reports*, vol. 7 (2017), p. 2163 (cited on p. 17).

19. N. H. Claessens, P. Moeskops, A. Buchmann, B. Latal, W. Knirsch, I. Scheer, I. Išgum, L. S. De Vries, M. J. Benders, and M. Von Rhein. "Delayed cortical gray matter development in neonates with severe congenital heart disease", *Pediatric research*, vol. 80 (2016) (cited on pp. 17, 87, 88).

20. K. J. Kersbergen, F. Leroy, I. Išgum, F. Groenendaal, L. S. de Vries, N. H. Claessens, I. C. van Haastert, P. Moeskops, C. Fischer, J.-F. Mangin, et al. "Relation between clinical risk factors, early cortical changes, and neurodevelopmental outcome in preterm infants", *NeuroImage*, vol. 142 (2016), pp. 301–310 (cited on p. 17).

21. F. J. Drost, K. Keunen, P. Moeskops, N. H. Claessens, F. van Kalken, I. Išgum, E. S. Voskuil-Kerkhof, F. Groenendaal, L. S. de Vries, M. J. Benders, et al. "Severe retinopathy of prematurity is associated with reduced cerebellar and brainstem volumes at term and neurodevelopmental deficits at two years", *Pediatric research* (2018) (cited on p. 17).

22. I. Išgum, M. J. Benders, B. Avants, M. J. Cardoso, S. J. Counsell, E. F. Gomez, L. Gui, P. S. Hűppi, K. J. Kersbergen, A. Makropoulos, et al. "Evaluation of automatic neonatal brain segmentation algorithms: the NeoBrainS12 challenge", *Medical Image Analysis*, vol. 20 (2015), pp. 135–151 (cited on pp. 17, 21, 23, 46, 47, 112).

23. S. M. Smith. "Fast robust automated brain extraction", *Human brain mapping*, vol. 17 (2002), pp. 143–155 (cited on pp. 17, 58, 72).

24. J. E. Iglesias, C.-Y. Liu, P. M. Thompson, and Z. Tu. "Robust brain extraction across datasets and comparison with publicly available methods", *IEEE Transactions on Medical Imaging*, vol. 30 (2011), pp. 1617–1634 (cited on p. 17).

25. S. F. Eskildsen, P. Coupé, V. Fonov, J. V. Manjón, K. K. Leung, N. Guizard, S. N. Wassef, L. R. Østergaard, D. L. Collins, and The Alzheimer's Disease Neuroimaging Initiative. "Beast: brain extraction based on nonlocal segmentation technique", *NeuroImage*, vol. 59 (2012), pp. 2362–2373 (cited on p. 17).

26. F. Ségonne, A. M. Dale, E. Busa, M. Glessner, D. Salat, H. K. Hahn, and B. Fischl. "A hybrid approach to the skull stripping problem in MRI", *NeuroImage*, vol. 22 (2004), pp. 1060–1075 (cited on p. 17).

27. K. Yamaguchi, Y. Fujimoto, S. Kobashi, Y. Wakata, R. Ishikura, K. Kuramoto, S. Imawaki, S. Hirota, and Y. Hata. "Automated fuzzy logic based skull stripping in neonatal and infantile MR images", *Fuzzy Systems (FUZZ), 2010 IEEE International Conference*, 2010, pp. 1–7 (cited on p. 17).

28. A. Serag, M. Blesa, E. J. Moore, R. Pataky, S. A. Sparrow, A. Wilkinson, G. Macnaught, S. I. Semple, and J. P. Boardman. "Accurate learning with few atlases (alfa): an algorithm for MRI neonatal brain extraction and comparison with 11 publicly available methods", *Scientific reports*, vol. 6 (2016), p. 23470 (cited on pp. 17, 59).

29. D. Mahapatra. "Skull stripping of neonatal brain MRI: using prior shape information with graph cuts", *Journal of digital imaging. LNCS* vol. 25 (2012), pp. 802–814 (cited on p. 17).

30. J. Anquez, E. D. Angelini, and I. Bloch. "Automatic segmentation of head structures on fetal MRI", *The IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2009, pp. 109–112 (cited on pp. 18, 38, 46).

31. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez. "A survey on deep learning in medical image analysis", *Medical Image Analysis*, vol. 42 (2017), pp. 60–88 (cited on pp. 18, 25, 46).

32. H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng. "Voxresnet: deep voxelwise residual networks for brain segmentation from 3d MR images", *NeuroImage*, vol. 170 (2018), pp. 446–455 (cited on p. 18).

33. J. Dolz, C. Desrosiers, and I. B. Ayed. "3d fully convolutional networks for subcortical segmentation in MRI: a large-scale study", *NeuroImage*, vol. 170 (2018), pp. 456–470 (cited on p. 18).

34. Z. Tu and X. Bai. "Auto-context and its application to high-level vision tasks and 3d brain image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32 (2010), pp. 1744–1757 (cited on p. 18).

35. Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson. "Deep learning for brain MRI segmentation: state of the art and future directions", *Journal of digital imaging*, vol. 30 (2017), pp. 449–459 (cited on p. 18).

36. A. Makropoulos, S. J. Counsell, and D. Rueckert. "A review on automatic fetal and neonatal brain MRI segmentation", *NeuroImage*, vol. 170 (2018), pp. 231–248 (cited on pp. 18, 46).

37. R. Dey and Y. Hong. "Compnet: complementary segmentation network for brain MRI extraction", *arXiv preprint arXiv:1804.00521* (2018) (cited on p. 18).

38. J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller. "Deep MRI brain extraction: a 3D convolutional neural network for skull stripping", *NeuroImage*, vol. 129 (2016), pp. 460–469 (cited on p. 18).

39. J. Dolz, C. Desrosiers, L. Wang, J. Yuan, D. Shen, and I. B. Ayed. "Deep cnn ensembles and suggestive annotations for infant brain MRI segmentation", *arXiv preprint arXiv:1712.05319* (2017) (cited on p. 18).

40. C. Wachinger, M. Reuter, and T. Klein. "Deepnat: deep convolutional neural network for segmenting neuroanatomy", *NeuroImage*, vol. 170 (2018), pp. 434–445 (cited on p. 18).

41. M. Rajchl, M. C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz, et al. "Deepcut: object segmentation from bounding box annotations using convolutional neural networks", *IEEE Transactions on Medical Imaging*, vol. 36 (2017), pp. 674–683 (cited on pp. 18, 46).

42. M. Rajchl, M. C. Lee, F. Schrans, A. Davidson, J. Passerat-Palmbach, G. Tarroni, A. Alansary, O. Oktay, B. Kainz, and D. Rueckert. "Learning under distributed weak supervision", *arXiv preprint arXiv:1606.01100* (2016) (cited on p. 18).

43. S. S. M. Salehi, D. Erdogmus, and A. Gholipour. "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging", *IEEE Transactions on Medical Imaging*, vol. 36 (2017), pp. 2319–2330 (cited on p. 18).

44. P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig. "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability", *Neuroimage*, vol. 31 (2006), pp. 1116–1128 (cited on p. 18).

45. S. S. M. Salehi, S. R. Hashemi, C. Velasco-Annis, A. Ouaalam, J. A. Estroff, D. Erdogmus, S. K. Warfield, and A. Gholipour. "Real-time automatic fetal brain extraction in fetal MRI by deep learning", *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, IEEE. 2018, pp. 720–724 (cited on pp. 18, 20, 34, 39).

46. N. Khalili, P. Moeskops, N. H. P. Claessens, S. Scherpenzeel, E. Turk, R. de Heus, M. J. N. L. Benders, M. A. Viergever, J. P. W. Pluim, and I. Išgum. "Automatic segmentation of the intracranial volume in fetal MR images", *Fetal, Infant and Ophthalmic Medical Image Analysis*, Cham: Springer International Publishing, 2017, pp. 42–51 (cited on pp. 18, 19, 46).

47. M. Ison, R. Donner, E. Dittrich, G. Kasprian, D. Prayer, and G. Langs. "Fully automated brain extraction and orientation in raw fetal MRI", *Workshop on Paediatric and Perinatal Imaging, MICCAI, LNCS* Cham: Springer International Publishing, 2012, pp. 17–24 (cited on p. 18).

48. K. Keraudren, V. Kyriakopoulou, M. Rutherford, J. V. Hajnal, and D. Rueckert. "Localisation of the brain in fetal MRI using bundled sift features", *MICCAI, LNCS* Cham: Springer International Publishing, 2013, pp. 582–589 (cited on pp. 18, 38).

49. K. Keraudren, M. Kuklisova-Murgasova, V. Kyriakopoulou, C. Malamateniou, M. A. Rutherford, B. Kainz, J. V. Hajnal, and D. Rueckert. "Automated fetal brain segmentation from 2d MRI slices for motion correction", *NeuroImage*, vol. 101 (2014), pp. 633–643 (cited on p. 18).

50. V. Taimouri, A. Gholipour, C. Velasco-Annis, J. A. Estroff, and S. K. Warfield. "A template-to-slice block matching approach for automatic localization of brain in fetal MRI", (2015), pp. 144–147 (cited on pp. 18, 38).

51. S. Tourbier, C. Velasco-Annis, V. Taimouri, P. Hagmann, R. Meuli, S. K. Warfield, M. B. Cuadra, and A. Gholipour. "Automated template-based brain localization and extraction for fetal brain MRI reconstruction", *NeuroImage*, vol. 155 (2017), pp. 460–472 (cited on p. 18).

52. O. Ronneberger, P. Fischer, and T. Brox. "U-net: convolutional networks for biomedical image segmentation", *MICCAI, LNCS* Cham: Springer International Publishing, 2015, pp. 234–241 (cited on pp. 19, 25, 48, 111, 114, 136).

53. K. Keunen. "The neonatal brain: early connectome development and childhood cognition", PhD thesis. Utrecht University, 2017 (cited on p. 21).

54. N. H. Claessens, S. O. Algra, N. J. Jansen, F. Groenendaal, E. de Wit, A. A. Wilbrink, F. Haas, A. N. Schouten, R. A. Nievelstein, M. J. Benders, et al. "Clinical and neuroimaging characteristics of cerebral sinovenous thrombosis in neonates undergoing cardiac surgery", *The Journal of thoracic and cardiovascular surgery* (2017) (cited on p. 21).

55. M. J. Brouwer, L. S. De Vries, K. J. Kersbergen, N. E. Van Der Aa, A. J. Brouwer, M. A. Viergever, I. Išgum, K. S. Han, F. Groenendaal, and M. J. Benders. "Effects of posthemorrhagic ventricular dilatation in the preterm infant on brain volumes and white matter diffusion variables at term-equivalent age", *The Journal of pediatrics*, vol. 168 (2016), pp. 41–49 (cited on pp. 21, 93, 103).

56. M. J. Benders, N. E. van der Aa, M. Roks, H. L. van Straaten, I. Išgum, M. A. Viergever, F. Groenendaal, L. S. de Vries, and F. van Bel. "Feasibility and safety of erythropoietin for neuroprotection after perinatal arterial ischemic stroke", *The Journal of pediatrics*, vol. 164 (2014), pp. 481–486 (cited on pp. 22, 109, 130).

57. T. Alderliesten, L. S. de Vries, L. Staats, I. C. van Haastert, L. Weeke, M. J. Benders, C. Koopman-Esseboom, and F. Groenendaal. "MRI and spectroscopy in (near) term neonates with perinatal asphyxia and therapeutic hypothermia", *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 102 (2017), F147–F152 (cited on p. 22).

58. J. T. Coyle, M. Oster-Granite, and J. D. Gearhart. "The neurobiologie consequences of down syndrome", *Brain research bulletin*, vol. 16 (1986), pp. 773–787 (cited on p. 22).

59. D. Kingma and J. B. Adam. "A method for stochastic optimisation", *International Conference on Learning Representations*, 2015 (cited on pp. 25, 50, 72, 114).

60. T. Dozat. "Incorporating nesterov momentum into Adam", *4th International Conference on Learning Representations, Workshop*, 2016 (cited on pp. 25, 50, 114).

61. S. Ioffe and C. Szegedy. "Batch normalization: accelerating deep network training by reducing internal covariate shift", *The 32nd International Conference on Machine Learning (ICML-15)*, 2015, pp. 448–456 (cited on pp. 25, 50, 114).

62. F. Chollet et al. "Keras", `https://keras.io`, 2015 (cited on pp. 25, 51).

63. A. A. Taha and A. Hanbury. "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool", *BMC medical imaging*, vol. 15 (2015), p. 29 (cited on p. 25).

64. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. "Brain tumor segmentation with deep neural networks", *Medical Image Analysis*, vol. 35 (2017), pp. 18–31 (cited on p. 38).

65. J. M. Wolterink, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, and I. Išgum. "Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks", *Medical Image Analysis*, vol. 34 (2016), pp. 123–136 (cited on p. 38).

66. K. Simonyan, A. Vedaldi, and A. Zisserman. "Deep inside convolutional networks: visualising image classification models and saliency maps", *arXiv preprint arXiv:1312.6034* (2013) (cited on p. 40).

67. P. A. Habas, K. Kim, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme. "Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses", *Human Brain Mapping*, vol. 31 (2010), pp. 1348–1358 (cited on pp. 45, 54, 58, 137).

68. A. Gholipour, A. Akhondi-Asl, J. A. Estroff, and S. K. Warfield. "Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly", *NeuroImage*, vol. 60 (2012), pp. 1819–1831 (cited on pp. 45, 59, 137).

69. A. Gholipour, J. A. Estroff, C. E. Barnewolt, S. A. Connolly, and S. K. Warfield. "Fetal brain volumetry through MRI volumetric reconstruction and segmentation", *International Journal of Computer Assisted Radiology and Surgery. LNCS* vol. 6 (2011), pp. 329–339 (cited on p. 45).

70. S. K. Warfield, K. H. Zou, and W. M. Wells. "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation", *IEEE Transactions on Medical Imaging*, vol. 23 (2004), pp. 903–921 (cited on p. 45).

71. A. Serag, V. Kyriakopoulou, M. Rutherford, A. Edwards, J. Hajnal, P. Aljabar, S. Counsell, J. Boardman, and D. Rueckert. "A multi-channel 4D probabilistic atlas of the developing brain: application to fetuses and neonates", *Annals of the British Machine Vision Association*, vol. 2012 (2012), pp. 1–14 (cited on pp. 45, 54, 58, 109, 137).

72. S. Jiang, H. Xue, A. Glover, M. Rutherford, D. Rueckert, and J. V. Hajnal. "MRI of moving subjects using multislice snapshot images with volume reconstruction (SVR): application to fetal, neonatal, and adult brain studies", *IEEE Transactions on Medical Imaging*, vol. 26 (2007), pp. 967–980 (cited on p. 45).

73. N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. "N4ITK: improved N3 bias correction", *IEEE Transactions on Medical Imaging*, vol. 29 (2010), pp. 1310–1320 (cited on p. 46).

74. Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson. "Deep learning for brain MRI segmentation: state of the art and future directions", *Journal of Digital Imaging*, vol. 30 (2017), pp. 449–459 (cited on p. 46).

75. Y.-D. Zhang, C. Pan, X. Chen, and F. Wang. "Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling", *Journal of Computational Science*, vol. 27 (2018), pp. 57–68 (cited on p. 46).

76. S.-H. Wang, Y.-D. Lv, Y. Sui, S. Liu, S.-J. Wang, and Y.-D. Zhang. "Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling", *Journal of Medical Systems*, vol. 42 (2018) (cited on p. 46).

77. M. Bouyssi-Kobar, A. J. du Plessis, R. McCarter, M. Brossard-Racine, J. Murnick, L. Tinkleman, R. L. Robertson, and C. Limperopoulos. "Third trimester brain growth in preterm infants compared with in utero healthy fetuses", *Pediatrics*, vol. 138 (2016) (cited on p. 46).

78. J. Dobbing and J. Sands. "Quantitative growth and development of human brain", *Archives of Disease in Childhood*, vol. 48 (1973), pp. 757–767 (cited on p. 46).

79. S. S. M. Salehi, S. R. Hashemi, C. Velasco-Annis, A. Ouaalam, J. A. Estroff, D. Erdogmus, S. K. Warfield, and A. Gholipour. "Real-time automatic fetal brain extraction in fetal MRI by deep learning", *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, 2018, pp. 720–724 (cited on p. 46).

80. "NeoBrainS2", `http://neobrains12.isi.uu.nl/reference.php`, (cited on p. 48).

81. D. Atkinson, D. L. Hill, P. N. Stoyle, P. E. Summers, and S. F. Keevil. "Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion", *IEEE Transactions on Medical imaging*, vol. 16 (1997), pp. 903–910 (cited on p. 69).

82. F. Godenschweger, U. Kägebein, D. Stucht, U. Yarach, A. Sciarra, R. Yakupov, F. Lüsebrink, P. Schulze, and O. Speck. "Motion correction in MRI of the brain", *Physics in Medicine & Biology*, vol. 61 (2016), R32 (cited on p. 69).

83. B. A. Duffy, W. Zhang, H. Tang, L. Zhao, M. Law, A. W. Toga, and H. Kim. "Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion", (2018) (cited on p. 69).

84. K. Pawar, Z. Chen, N. J. Shah, and G. F. Egan. "Moconet: motion correction in 3D MPRAGE images using a convolutional neural network approach", *arXiv preprint arXiv:1807.10831* (2018) (cited on p. 69).

85. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks", *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232 (cited on p. 69).

86. K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition", *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778 (cited on p. 71).

87. J. Johnson, A. Alahi, and L. Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution", *European conference on computer vision*, Springer. 2016, pp. 694–711 (cited on p. 71).

88. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. "Image-to-image translation with conditional adversarial networks", *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134 (cited on p. 71).

89. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative adversarial nets", *Advances in neural information processing systems*, 2014, pp. 2672–2680 (cited on p. 71).

90. N. H. Claessens, C. J. Kelly, S. J. Counsell, and M. J. Benders. "Neuroimaging, cardiovascular physiology, and functional outcomes in infants with congenital heart disease", *Developmental Medicine & Child Neurology*, vol. 59 (2017), pp. 894–902 (cited on pp. 79, 87).

91. S. Peyvandi, V. Chau, T. Guo, D. Xu, H. C. Glass, A. Synnes, K. Poskitt, A. J. Barkovich, S. P. Miller, and P. S. McQuillen. "Neonatal brain injury and timing of neurodevelopmental assessment in patients with congenital heart disease", *Journal of the American College of Cardiology*, vol. 71 (2018), pp. 1986–1996 (cited on pp. 79, 87).

92. N. H. Claessens, S. O. Algra, T. L. Ouwehand, N. J. Jansen, R. Schappin, F. Haas, M. J. Eijsermans, L. S. de Vries, M. J. Benders, C. L. S. G. Utrecht, et al. "Perioperative neonatal brain injury is associated with worse school-age neurodevelopment in children with critical congenital heart disease", *Developmental Medicine & Child Neurology*, vol. 60 (2018), pp. 1052–1058 (cited on pp. 79, 87).

93. M. von Rhein, A. Buchmann, C. Hagmann, H. Dave, V. Bernet, I. Scheer, W. Knirsch, B. Latal, C. Bürki, R. Prêtre, et al. "Severe congenital heart defects are associated with global reduction of neonatal brain volumes", *The Journal of Pediatrics*, vol. 167 (2015), pp. 1259–1263 (cited on pp. 79, 87, 88).

94. C. J. Kelly, A. Makropoulos, L. Cordero-Grande, J. Hutter, A. Price, E. Hughes, M. Murgasova, R. P. A. Teixeira, J. K. Steinweg, S. Kulkarni, et al. "Impaired development of the cerebral cortex in infants with congenital heart disease is correlated to reduced cerebral oxygen delivery", *Scientific reports*, vol. 7 (2017), pp. 1–10 (cited on pp. 79, 88).

95. C. Limperopoulos, W. Tworetzky, D. B. McElhinney, J. W. Newburger, D. W. Brown, R. L. Robertson Jr, C. Dunbar-Masterson, B. Trainor, P. C. Laussen, and A. J. du Plessis. "Brain volume and metabolism in fetuses with congenital heart disease", *Evaluation with quantitative magnetic resonance imaging and spectroscopy*, vol. 2010 (2009), p. 121 (cited on pp. 79, 87, 88).

96. L. Sun, C. K. Macgowan, J. G. Sled, S.-J. Yoo, C. Manlhiot, P. Porayette, L. Grosse-Wortmann, E. Jaeggi, B. W. McCrindle, J. Kingdom, et al. "Reduced fetal cerebral oxygen consumption is associated with smaller brain size in fetuses with congenital heart disease", *Circulation*, vol. 131 (2015), pp. 1313–1323 (cited on pp. 79, 88).

97. J. M. Lim, T. Kingdom, B. Saini, V. Chau, M. Post, S. Blaser, C. Macgowan, S. P. Miller, and M. Seed. "Cerebral oxygen delivery is reduced in newborns with congenital heart disease", *The Journal of thoracic and cardiovascular surgery*, vol. 152 (2016), pp. 1095–1103 (cited on p. 79).

98. C. Garel and C. Alberti. "Coronal measurement of the fetal lateral ventricles: comparison between ultrasonography and magnetic resonance imaging", *Ultrasound in Obstetrics and Gynecology*, vol. 27 (2006), pp. 23–27 (cited on p. 80).

99. L.-A. Papile, J. Burstein, R. Burstein, and H. Koffler. "Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm", *The Journal of pediatrics*, vol. 92 (1978), pp. 529–534 (cited on p. 80).

100. K. J. Kersbergen, A. Makropoulos, P. Aljabar, F. Groenendaal, L. S. De Vries, S. J. Counsell, and M. J. Benders. "Longitudinal regional brain development and clinical risk factors in extremely preterm infants", *The Journal of pediatrics*, vol. 178 (2016), pp. 93–100 (cited on p. 87).

101. C. Schellen, S. Ernst, G. M. Gruber, E. Mlczoch, M. Weber, P. C. Brugger, B. Ulm, G. Langs, U. Salzer-Muhar, D. Prayer, et al. "Fetal MRI detects early alterations of brain development in tetralogy of fallot", *American journal of obstetrics and gynecology*, vol. 213 (2015), pp. 392–e1 (cited on p. 87).

102. N. Andescavage, A. Yarish, M. Donofrio, D. Bulas, I. Evangelou, G. Vezina, R. McCarter, C. Limperopoulos, et al. "3-d volumetric MRI evaluation of the placenta in fetuses with complex congenital heart disease", *Placenta*, vol. 36 (2015), pp. 1024–1030 (cited on p. 87).

103. M. Brossard-Racine, A. J. Du Plessis, G. Vezina, R. Robertson, D. Bulas, I. E. Evangelou, M. Donofrio, D. Freeman, and C. Limperopoulos. "Prevalence and spectrum of in utero structural brain abnormalities in fetuses with complex congenital heart disease", *American Journal of Neuroradiology*, vol. 35 (2014), pp. 1593–1599 (cited on p. 87).

104. N. N. Andescavage, A. DuPlessis, R. McCarter, G. Vezina, R. Robertson, and C. Limperopoulos. "Cerebrospinal fluid and parenchymal brain development and growth in the healthy fetus", *Developmental neuroscience*, vol. 38 (2016), pp. 420–429 (cited on p. 87).

105. K. N. Heye, W. Knirsch, B. Latal, I. Scheer, K. Wetterling, A. Hahn, H. Akintürk, D. Schranz, I. Beck, B. Reich, et al. "Reduction of brain volumes after neonatal cardiopulmonary bypass surgery in single-ventricle congenital heart disease before fontan completion", *Pediatric research*, vol. 83 (2018), p. 63 (cited on p. 87).

106. C. Clouchoux, A. Du Plessis, M. Bouyssi-Kobar, W. Tworetzky, D. McElhinney, D. Brown, A. Gholipour, D. Kudelski, S. Warfield, R. McCarter, et al. "Delayed cortical development in fetuses with complex congenital heart disease", *Cerebral cortex*, vol. 23 (2013), pp. 2932–2943 (cited on p. 87).

107. M. Owen, M. Shevell, M. Donofrio, A. Majnemer, R. McCarter, G. Vezina, M. Bouyssi-Kobar, I. Evangelou, D. Freeman, N. Weisenfeld, et al. "Brain volume and neurobehavior in newborns with complex congenital heart defects", *The Journal of pediatrics*, vol. 164 (2014), pp. 1121–1127 (cited on p. 87).

108. M. von Rhein, A. Buchmann, C. Hagmann, R. Huber, P. Klaver, W. Knirsch, and B. Latal. "Brain volumes predict neurodevelopment in adolescents after surgery for congenital heart disease", *Brain*, vol. 137 (2014), pp. 268–276 (cited on p. 87).

109. A. Dimitropoulos, P. S. McQuillen, V. Sethi, A. Moosa, V. Chau, D. Xu, R. Brant, A. Azakie, A. Campbell, A. J. Barkovich, et al. "Brain injury and development in newborns with critical congenital heart disease", *Neurology*, vol. 81 (2013), pp. 241–248 (cited on p. 88).

110. D. B. Andropoulos, J. V. Hunter, D. P. Nelson, S. A. Stayer, A. R. Stark, E. D. McKenzie, J. S. Heinle, D. E. Graves, and C. D. Fraser Jr. "Brain immaturity is associated with brain injury before and after neonatal cardiac surgery with high-flow bypass and cerebral oxygenation monitoring", *The Journal of thoracic and cardiovascular surgery*, vol. 139 (2010), pp. 543–556 (cited on p. 88).

111. M. Donofrio, Y. Bremer, R. Schieken, C. Gennings, L. Morton, B. W. Eidem, F. Cetta, C. Falkensammer, J. Huhta, and C. Kleinman. "Autoregulation of cerebral blood flow in fetuses with congenital heart disease: the brain sparing effect", *Pediatric cardiology*, vol. 24 (2003), pp. 436–443 (cited on p. 88).

112. S. Peyvandi, H. Kim, J. Lau, A. J. Barkovich, A. Campbell, S. Miller, D. Xu, and P. McQuillen. "The association between cardiac physiology, acquired brain injury, and postnatal brain growth in critical congenital heart disease", *The Journal of thoracic and cardiovascular surgery*, vol. 155 (2018), pp. 291–300 (cited on p. 88).

113. B. J. Stoll, N. I. Hansen, E. F. Bell, M. C. Walsh, W. A. Carlo, S. Shankaran, A. R. Laptook, P. J. Sánchez, K. P. Van Meurs, M. Wyckoff, et al. "Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012", *Jama*, vol. 314 (2015), pp. 1039–1051 (cited on p. 93).

114. J. R. Ellenbogen, M. Waqar, and B. Pettorini. "Management of post-haemorrhagic hydrocephalus in premature infants", *Journal of Clinical Neuroscience*, vol. 31 (2016), pp. 30–34 (cited on p. 93).

115. L. S. de Vries, F. Groenendaal, K. D. Liem, A. Heep, A. J. Brouwer, E. van't Verlaat, I. Benavente-Fernández, H. L. van Straaten, G. van Wezel-Meijler, B. J. Smit, et al. "Treatment thresholds for intervention in posthaemorrhagic ventricular dilation: a randomised controlled trial", *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 104 (2019), F70–F75 (cited on p. 93).

116. J. C. Wellons, C. N. Shannon, R. Holubkov, J. Riva-Cambrin, A. V. Kulkarni, D. D. Limbrick, W. Whitehead, S. Browd, C. Rozzelle, T. D. Simon, et al. "Shunting outcomes in posthemorrhagic hydrocephalus: results of a hydrocephalus clinical research network prospective cohort study", *Journal of Neurosurgery: Pediatrics*, vol. 20 (2017), pp. 19–29 (cited on p. 93).

117. J. Hislop, L. Dubowitz, A. Kaiser, M. Singh, and A. Whitelaw. "Outcome of infants shunted for post-haemorrhagic ventricular dilatation", *Developmental Medicine & Child Neurology*, vol. 30 (1988), pp. 451–456 (cited on p. 93).

118. S. Cherian, A. Whitelaw, M. Thoresen, and S. Love. "The pathogenesis of neonatal post-hemorrhagic hydrocephalus", *Brain pathology*, vol. 14 (2004), pp. 305–311 (cited on p. 93).

119. H. Bassan, R. Eshel, I. Golan, D. Kohelet, L. B. Sira, D. Mandel, L. Levi, S. Constantini, L. Beni-Adani, E. V. D. S. Investigators, et al. "Timing of external ventricular drainage and neurodevelopmental outcome in preterm infants with posthemorrhagic hydrocephalus", *european journal of paediatric neurology*, vol. 16 (2012), pp. 662–670 (cited on p. 93).

120. L. M. Leijser, S. P. Miller, G. van Wezel-Meijler, A. J. Brouwer, J. Traubici, I. C. van Haastert, H. E. Whyte, F. Groenendaal, A. V. Kulkarni, K. S. Han, et al. "Posthemorrhagic ventricular dilatation in preterm infants: when best to intervene?", *Neurology*, vol. 90 (2018), pp. e698–e706 (cited on pp. 93, 103).

121. M. I. Levene. "Measurement of the growth of the lateral ventricles in preterm infants with real-time ultrasound." *Archives of Disease in Childhood*, vol. 56 (1981), pp. 900–904 (cited on pp. 93, 94).

122. T. E. Inder, J. M. Perlman, and J. J. Volpe. "Preterm intraventricular hemorrhage/posthemorrhagic hydrocephalus", *Volpe's Neurology of the Newborn*, Elsevier, 2018, pp. 637–698 (cited on pp. 93, 103).

123. M. Davies, M. Swaminathan, S. Chuang, and F. Betheras. "Reference ranges for the linear dimensions of the intracranial ventricles in preterm neonates", *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 82 (2000), F218–F223 (cited on p. 94).

124. A. V. Kulkarni, J. M. Drake, D. C. Armstrong, and P. B. Dirks. "Measurement of ventricular size: reliability of the frontal and occipital horn ratio compared to subjective assessment", *Pediatric neurosurgery*, vol. 31 (1999), pp. 65–70 (cited on p. 94).

125. K. Sävman, M. Blennow, H. Hagberg, E. Tarkowski, M. Thoresen, and A. Whitelaw. "Cytokine response in cerebrospinal fluid from preterm infants with posthaemorrhagic ventricular dilatation", *Acta paediatrica*, vol. 91 (2002), pp. 1357–1363 (cited on p. 102).

126. A. Whitelaw, S. Cherian, M. Thoresen, and I. Pople. "Posthaemorrhagic ventricular dilatation: new mechanisms and new treatment", *Acta paediatrica*, vol. 93 (2004), pp. 11–14 (cited on p. 102).

127. P. Srinivasakumar, D. Limbrick, R. Munro, D. Mercer, R. Rao, T. Inder, and A. Mathur. "Posthemorrhagic ventricular dilatation—impact on early neurodevelopmental outcome", *American journal of perinatology*, vol. 30 (2013), pp. 207–214 (cited on p. 102).

128. E. W. Tam, S. P. Miller, C. Studholme, V. Chau, D. Glidden, K. J. Poskitt, D. M. Ferriero, and A. J. Barkovich. "Differential effects of intraventricular hemorrhage and white matter injury on preterm cerebellar growth", *The Journal of pediatrics*, vol. 158 (2011), pp. 366–371 (cited on p. 102).

129. S. Jary, A. De Carli, L. A. Ramenghi, and A. Whitelaw. "Impaired brain growth and neurodevelopment in preterm infants with posthaemorrhagic ventricular dilatation", *Acta Paediatrica*, vol. 101 (2012), pp. 743–748 (cited on pp. 102, 103).

130. J. S. Soul, G. A. Taylor, D. Wypij, A. J. Duplessis, and J. J. Volpe. "Noninvasive detection of changes in cerebral blood flow by near-infrared spectroscopy in a piglet model of hydrocephalus", *Pediatric research*, vol. 48 (2000), pp. 445–449 (cited on p. 102).

131. B. Resch, A. Gerdermann, U. Maurer, E. Ritschl, and W. Müller. "Neurodevelopmental outcome of hydrocephalus following intra-/periventricular hemorrhage in preterm infants: short-and long-term results", *Child's Nervous System*, vol. 12 (1996), pp. 27–33 (cited on p. 103).

132. S. A. van Zanten, T. R. de Haan, J. Ursum, and L. van Sonderen. "Neurodevelopmental outcome of post-hemorrhagic ventricular dilatation at 12 and 24 months corrected age with high-threshold therapy", *european journal of paediatric neurology*, vol. 15 (2011), pp. 487–492 (cited on p. 103).

133. B. F. Matta, D. K. Menon, and M. Smith. "Core topics in neuroanaesthesia and neurointensive care", Cambridge University Press, 2011 (cited on p. 103).

134. B. Mokri. "The monro–kellie hypothesis: applications in csf volume depletion", *Neurology*, vol. 56 (2001), pp. 1746–1748 (cited on p. 103).

135. D. Fernández-López, N. Natarajan, S. Ashwal, and Z. S. Vexler. "Mechanisms of perinatal arterial ischemic stroke", *Journal of Cerebral Blood Flow & Metabolism*, vol. 34 (2014), pp. 921–932 (cited on p. 109).

136. E. Mercuri, M. Rutherford, F. Cowan, J. Pennock, S. Counsell, M. Papadimitriou, D. Azzopardi, G. Bydder, and L. Dubowitz. "Early prognostic indicators of outcome in infants with neonatal cerebral infarction: a clinical, electroencephalogram, and magnetic resonance imaging study", *Pediatrics*, vol. 103 (1999), pp. 39–46 (cited on p. 109).

137. S. Bava, S. L. Archibald, and D. A. Trauner. "Brain structure in prenatal stroke: quantitative magnetic resonance imaging (mri) analysis", *Journal of child neurology*, vol. 22 (2007), pp. 841–847 (cited on p. 109).

138. P. Anbeek, K. L. Vincken, F. Groenendaal, A. Koeman, M. J. Van Osch, and J. Van der Grond. "Probabilistic brain tissue segmentation in neonatal magnetic resonance imaging", *Pediatric research*, vol. 63 (2008), pp. 158–163 (cited on p. 109).

139. A. Makropoulos, C. Ledig, P. Aljabar, A. Serag, J. V. Hajnal, A. D. Edwards, S. J. Counsell, and D. Rueckert. "Automatic tissue and structural segmentation of neonatal brain MRI using expectation-maximization", *MICCAI Grand Challenge on Neonatal Brain Segmentation*, vol. 2012 (2012), pp. 9–15 (cited on p. 109).

140. C. N. Devi, A. Chandrasekharan, V. Sundararaman, and Z. C. Alex. "Neonatal brain MRI segmentation: a review", *Computers in biology and medicine*, vol. 64 (2015), pp. 163–178 (cited on p. 109).

141. G. Sanroma, O. M. Benkarim, G. Piella, and M. Á. G. Ballester. "Building an ensemble of complementary segmentation methods by exploiting probabilistic estimates", *Machine Learning in Medical Imaging*, edited by L. Wang, E. Adeli, Q. Wang, Y. Shi, and H.-I. Suk. Cham: Springer International Publishing, 2016, pp. 27–35 (cited on p. 109).

142. L. Wang, Y. Gao, F. Shi, G. Li, J. H. Gilmore, W. Lin, and D. Shen. "Links: learning-based multi-source integration framework for segmentation of infant brain images", *NeuroImage*, vol. 108 (2015), pp. 160–172 (cited on pp. 109, 128).

143. N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. de Heus, T. Kolk, M. A. Viergever, M. J. Benders, and I. Išgum. "Automatic brain tissue segmentation in fetal MRI using convolutional neural networks", *Magnetic Resonance Imaging*, vol. 64 (2019), pp. 77–89 (cited on p. 109).

144. N. Khalili, E. Turk, M. Zreik, M. A. Viergever, M. J. N. L. Benders, and I. Išgum. "Generative adversarial network for segmentation of motion affected neonatal brain MRI", *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Cham: Springer International Publishing, 2019, pp. 320–328 (cited on pp. 109, 129).

145. N. Van der Aa, M. Benders, F. Groenendaal, and L. De Vries. "Neonatal stroke: a review of the current evidence on epidemiology, pathogenesis, diagnostics and therapeutic options", *Acta Paediatrica*, vol. 103 (2014), pp. 356–364 (cited on p. 109).

146. K. Murphy, N. E. van der Aa, S. Negro, F. Groenendaal, L. S. de Vries, M. A. Viergever, G. B. Boylan, M. J. Benders, and I. Išgum. "Automatic quantification of ischemic injury on diffusion-weighted MRI of neonatal hypoxic ischemic encephalopathy", *NeuroImage: Clinical*, vol. 14 (2017), pp. 222–232 (cited on p. 109).

147. I. Išgum, N. van der Aa, F. Groenendaal, L. de Vries, M. Benders, and M. Viergever. "MRI-based delineation of perinatal arterial ischemic stroke", *Image Analysis of Human Brain Development workshop, 14th International Conference on Medical Image Computing and Computer Assisted Intervention* () (cited on p. 109).

148. N. Ghosh, Y. Sun, B. Bhanu, S. Ashwal, and A. Obenaus. "Automated detection of brain abnormalities in neonatal hypoxia ischemic injury from MR images", *Medical Image Analysis*, vol. 18 (2014), pp. 1059–1069 (cited on p. 109).

149. R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. Mok, W. C. Chu, D. Wang, and L. Shi. "Automatic segmentation of acute ischemic stroke from dwi using 3-d fully convolutional densenets", *IEEE Transactions on Medical Imaging*, vol. 37 (2018), pp. 2149–2160 (cited on pp. 111, 129).

150. G. Praveen, A. Agrawal, P. Sundaram, and S. Sardesai. "Ischemic stroke lesion segmentation using stacked sparse autoencoder", *Computers in biology and medicine*, vol. 99 (2018), pp. 38–52 (cited on p. 111).

151. C. Hazirbas, L. Ma, C. Domokos, and D. Cremers. "Fusenet: incorporating depth into semantic segmentation via fusion-based cnn architecture", *Asian conference on computer vision*, Springer. 2016, pp. 213–228 (cited on p. 111).

152. S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim. "Elastix: a toolbox for intensity-based medical image registration", *IEEE Transactions on Medical Imaging*, vol. 29 (2009), pp. 196–205 (cited on p. 112).

153. M. Staring, S. Klein, and J. P. Pluim. "A rigidity penalty term for nonrigid registration", *Medical physics*, vol. 34 (2007), pp. 4098–4108 (cited on p. 112).

154. M. C. Florkow, F. Zijlstra, K. Willemsen, M. Maspero, C. A. van den Berg, L. G. Kerkmeijer, R. M. Castelein, H. Weinans, M. A. Viergever, M. van Stralen, et al. "Deep learning–based mr-to-ct synthesis: the influence of varying gradient echo–based MR images as input channels", *Magnetic resonance in medicine*, vol. 83 (2020), pp. 1429–1441 (cited on p. 113).

155. R. Likert. "A technique for the measurement of attitudes." *Archives of psychology* (1932) (cited on p. 116).

156. M. S. Graham, I. Drobnjak, M. Jenkinson, and H. Zhang. "Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI", *PloS one*, vol. 12 (2017) (cited on p. 120).

157. H. Ehrenreich, K. Weissenborn, H. Prange, D. Schneider, C. Weimar, K. Wartenberg, P. D. Schellinger, M. Bohn, H. Becker, M. Wegrzyn, et al. "Recombinant human erythropoietin in the treatment of acute ischemic stroke", *Stroke*, vol. 40 (2009), pp. e647–e656 (cited on p. 130).

158. H. Ehrenreich, A. Kästner, K. Weissenborn, J. Streeter, S. Sperling, K. K. Wang, H. Worthmann, R. L. Hayes, N. von Ahsen, A. Kastrup, et al. "Circulating damage marker profiles support a neuroprotective effect of erythropoietin in ischemic stroke patients", *Molecular medicine*, vol. 17 (2011), pp. 1306–1310 (cited on p. 130).

159. X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng. "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from CT volumes", *IEEE Transactions on Medical Imaging*, vol. 37 (2018), pp. 2663–2674 (cited on p. 136).

160. Y. Weng, T. Zhou, Y. Li, and X. Qiu. "Nas-unet: neural architecture search for medical image segmentation", *IEEE Access*, vol. 7 (2019), pp. 44247–44257 (cited on p. 136).

161. Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. "Stargan: unified generative adversarial networks for multi-domain image-to-image translation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018 (cited on p. 137).

162. D. D. Miller and E. W. Brown. "Artificial intelligence in medical practice: the question to the answer?", *The American Journal of Medicine*, vol. 131 (2018), pp. 129–133 (cited on p. 138).

Acknowledgments

Four years ago, when I took my first step in the UMC, I could not imagine the great adventure I would have, resulting in this thesis. The significant impact is beyond this thesis, and here, I would like to express my gratitude to all who directly and indirectly have contributed.

First of all, I would like to thank my supervisor Ivana. Thank you for your trust in me to start my Ph.D. in your group and for your advice and guidance. In particular, thank you for your patience during the writing process and detailed feedback. Without your hard work, sometimes over hours, and on vacation, finishing this thesis would have been impossible.

Dear Max, your encouraging words for each achievement, and your impartial words in hard times lifted my energy. Thank you for your constructive feedback on the publications and being present whenever I needed it.

Dear Manon, thank you for bringing the clinical perspective to the project. Thank you for connecting me with many people in your team. Without neonatology collaboration, this thesis would not have been possible.

This thesis is the result of many people's efforts and collaborations. I would like to thank all my collaborators and co-authors for many helpful feedback and discussions. I would like to thank former and current QIA members, Jelmer, Bob, Sanne, Nils, Louis, Jorg, Julia, Steffen, Majd, Niko, Jurica, and Matthijs for many meetings and feedback on my work, for all dinners and cakes. I would like to thank members of the reading committee, Prof. dr. Ir. M.J.C Eijkemans, Prof. dr. P.R Luijten, Prof. dr. N.F. Ramsey, Dr. M. de Bruijne, Prof. dr. R. van Elburg and Dr. M. Bekker for evaluating this thesis.

I was lucky to have wonderful colleagues, who had time for me for either conversation or a cup of coffee. Majd, thank you for advice, scientific discussions as well as uncountable jokes and laughter at the office. You made this journey more fun and gave me confidence in hard times, and thank you for being my paranymph. Niko, thank you for the unlimited support and contributions. I enjoyed my time sharing the office with you, and thank you for being my paranymph. Pim, thank you for supporting me at the start of my Ph.D., and any other time, having someone who had the same experience in the same field and group was a privilege. Sanne, thank you for being 'the lovely colleague'. I had a great time in MICCAI with you. Thank you for helping with the dutch summary together with Julia. Steffen, I enjoyed my time sharing the office with you, as well as many game nights together. Thank you for teaching me how to hide my sweets in the office properly. Dr. MD Max, thank you for all heart attacks, now I

know my heart is not easily failing. Mateusz, thank you for informal and formal discussions, some ideas of this thesis were popping up from our random conversations. And more than that for being a truthful friend. Kim and Bea, we bonded beyond colleagues; you both made this journey smoother and more delightful. I always knew which office to go at the end of tough Ph.D. days. Thank you for all the sisterhood supports. Ishaan, thank you for all coffee breaks and an excellent sense of movie and book suggestions, and everyone at OIO steeg for many refreshing lunch and coffee breaks. And of course, all ISI members, thank you for the enjoyable time at Christmas parties, Sinterklaas, many dinners, and drinks.

Many people were indirectly part of this journey. Prof. Philippe Cattin, you had an immense impact on my research skills during my master thesis, you encouraged me the most to join the ISI group and start my Ph.D. Thank you for that. Banafshe, you were an achievement of ISMRM and one of few friends I feel I knew for long. Parvaneh, even though we were miles away for years, our friendship stays the same, thank you for always having time for me. To my many Persian friends who form a family in Utrecht, special thanks to Marjane, Sohrab, Zeynab, Parisa for tons of gathering, dancing, and celebrating.

Shokofeh and Mojtaba, dear in-laws, thank you for your warm hearts and your constant supports. We are lucky to have you in our life.

Dear Dad and Mom, thank you for teaching me to explore and never stop questioning; this was the first tiny step towards being a researcher. You encourage me to step out of my comfort zone and push me to excel. Without you, this thesis would have ended differently. Hanieh, thank you for being always there for me with cheerful and good spirits.

Dear Mohsen, reuniting with you in the Netherlands was the happiest moment of this journey. Thank you for endless care and love, for believing in me in the moments I myself don't. Without your support, love, and present, this thesis would not have been possible.

Publications

Journal publications

N. Khalili, N. van der Aa, F. Groenendaal, L. de Vries, J. Dudink, M. Zreik, N. Wagenaar, J. Breur, M. Viergever, and M. B. I. Išgum. "Brain tissue segmentation in neonatal MRI using multi-modal CNN", *in preparation* (2020).

M. Zreik, N. Hampe, T. Leiner, **N. Khalili**, J. Wolterink, M. Voskuil, M. Viergever, and I. Išgum. "Combined analysis of coronary arteries and the left ventricular myocardium in cardiac CT angiography for detection of patients with functionally significant stenosis", *Submitted* (2020).

M. Zreik, R. W. van Hamersvelt, **N. Khalili**, J. M. Wolterink, M. Voskuil, M. A. Viergever, T. Leiner, and I. Išgum. "Deep learning analysis of coronary arteries in cardiac CT angiography for detection of patients requiring invasive coronary angiography", *IEEE Transactions on Medical Imaging*, vol. 39 (2020), pp. 1545–1557.

N. Khalili, E. Turk, M. Benders, P. Moeskops, N. Claessens, R. de Heus, A. Franx, N. Wagenaar, J. Breur, M. Viergever, and I. Išgum. "Automatic extraction of the intracranial volume in fetal and neonatal MR scans using convolutional neural networks", *NeuroImage: Clinical*, vol. 24 (2019), p. 102061.

N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. de Heus, T. Kolk, M. A. Viergever, M. J. Benders, and I. Išgum. "Automatic brain tissue segmentation in fetal MRI using convolutional neural networks", *Magnetic Resonance Imaging*, vol. 64 (2019), pp. 77–89.

N. Claessens, **N. Khalili**, I. Išgum, H. Ter Heide, T. Steenhuis, E. Turk, N. Jansen, L. de Vries, J. Breur, R. de Heus, and M. Benders. "Brain and CSF volumes in fetuses and neonates with antenatal diagnosis of critical congenital heart disease: a longitudinal MRI study", *American Journal of Neuroradiology*, vol. 40 (2019), pp. 885–891.

M. N. Cizmeci, **N. Khalili**, N. H. Claessens, F. Groenendaal, K. D. Liem, A. Heep, I. Benavente-Fernández, H. L. van Straaten, G. van Wezel-Meijler, S. J. Steggerda, J. Dudink, I. Išgum, A. Whitelaw, M. J. Benders, L. de Vries, and the ELVIS study group. "Assessment of brain injury and brain volumes after posthemorrhagic ventricular dilatation:

a nested substudy of the randomized controlled elvis trial", *The Journal of Pediatrics*, vol. 208 (2019), pp. 191–197.

Conference proceedings

N. Khalili, E. Turk, M. Zreik, M. A. Viergever, M. J. N. L. Benders, and I. Išgum. "Generative adversarial network for segmentation of motion affected neonatal brain MRI", *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Cham: Springer International Publishing, 2019, pp. 320–328.

J. Fernandes, V. Alves, **N. Khalili**, M. J. Benders, I. Išgum, J. Pluim, and P. Moeskops. "Convolutional neural network-based regression for quantification of brain characteristics using MRI", *World Conference on Information Systems and Technologies*, Springer. 2019, pp. 577–586.

N. Khalili, P. Moeskops, N. H. P. Claessens, S. Scherpenzeel, E. Turk, R. de Heus, M. J. N. L. Benders, M. A. Viergever, J. P. W. Pluim, and I. Išgum. "Automatic segmentation of the intracranial volume in fetal MR images", *Fetal, Infant and Ophthalmic Medical Image Analysis*, Cham: Springer International Publishing, 2017, pp. 42–51.

Conference abstracts

R. Stegemana, S. F. Hak, E. Turk, E. Overbeek, **N. Khalili**, F. Groenendaal, N. H. Claessens, F. Lammertink, M. R. A. M. Tataranno, H. ter Heide2, T. J. Steenhuis, J. M. Breur, N. J. Jansen, I. Išgum, R. de Heus, and M. Benders. "In utero brain development of infants with critical congenital heart disease compared to healthy controls", *Submitted to European Academy of Paediatric Societies*, 2019.

N. Khalili, N. Lessmann, E. Turk, N. Claessens, R. de Heus, T. Kolk, M. Viergever, M. Benders, and I. Išgum. "Brain tissue segmentation in fetal MRI using convolutional neural networks with simulated intensity inhomogeneities", *International Society for Magnetic Resonance in Medicine*, 2019.

Biography

Nadieh Khalili was born on July 12, 1989 in Tehran, Iran. She performed her bachelor's studies in Biomedical Engineering at the Science Research Branch of Azad University, Tehran, Iran. Nadieh moved to Switzerland for her master's studies in Biomedical Engineering at Bern University and obtained her MSc magna cum laude. Her master's thesis was on Multi-modal registration of 2D histology images on 3D CT. In 2016, she moved to the Netherlands and joined the Image Sciences Institute as a Ph.D. candidate. The results of her research, focusing on neonatal and fetal MRI analysis using machine and deep learning, are presented in this thesis.