Article

NeBSS: Semi-Automated Parcellation of Neonatal Structural Brain MRI

Rafael Ceschin [1,2,*], **Alexandria Zahner** [2], **Vanathi Gopalakrishnan** [1] **and Ashok Panigrahy** [1,2]

[1] Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA; vanathi@pitt.edu (V.G.); Ashok.Panigrahy@vchp.edu (A.P.)

[2] Department of Radiology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA 15224, USA; alexandria.zahner@chp.edu

* Correspondence: rcc10@pitt.edu; Tel.: +1-412-692-5510; Fax: +1-412-864-8622

Abstract: 1) Introduction: Brain parcellation is an important processing step in the analysis of structural brain MRI. Existing software implementations are optimized for fully developed adult brains, and provide inadequate results when applied to neonatal brain imaging. 2) Methods: We developed a semi-automated pipeline, NeBSS, for extracting 50 discrete brain structures from neonatal brain MRI, using an atlas registration method that leverages the existing ALBERT neonatal atlas 3) Results: We demonstrate a simple linear workflow for neonatal brain parcellation. NeBSS is robust to variation in imaging acquisition protocol and magnet field strength. 4) Conclusion: NeBSS is a robust pipeline capable of parcellating neonatal brain MRIs using a simple processing workflow. NeBSS fills a need in clinical translational research in neonatal imaging, where existing automated or semi-automated implementations are too rigid to be successfully applied to multi-center neuroprotection studies and clinically heterogeneous cohorts. The software is open source and freely available.

Keywords: neonatal MRI; brain structure segmentation; volume extraction

1. Introduction

1.1 Brain Parcellation in Clinical Research

Brain parcellation, sometimes referred to as brain segmentation, is the process of discretely labeling a brain image into anatomically or functionally congruent regions. Classical parcellation methods involved the manual delineation of each region of interest, one subject at a time, by a trained user usually following a brain atlas guided protocol. Naturally, this was prohibitively time consuming for analyzing large cohorts, with each brain taking in the order of days to weeks to fully segment, depending on the granularity and number of regions desired. More recently, many attempts at automatic brain segmentation have been created [1–3]. However, these methods are generated and optimized for the parcellation of fully developed brains, and break down when applied to neonatal datasets.

As a result, neonatal brain parcellation is still primarily done via manual segmentation of the specific structures of interest. New computational approaches have shown great promise in automated neonatal segmentation [4], but these methods are inflexible when processing multi-center neonatal neuroimaging data and clinically heterogeneous populations, and are also highly sensitive to the protocol used in image acquisition. This restricts their utility in clinical translational research, as they are unable to generalize to at-risk or injured populations, such as infants born with congenital heart disease (CHD). These populations often suffer from higher variability in both imaging protocols and image quality as a result of necessary trade-offs in clinically acquired imaging. Here we introduce Neonatal Brain Structure Segmentation (NeBSS), an open source semi-automated processing pipeline, where we have adapted existing atlas registration based methods to be more robust to both

structurally abnormal subjects and less than ideal image quality, with the tradeoff of requiring a final manual correction step.

1.2 Structural MRI Acquisition in Neonates

Conventional MRI takes advantage of the physical properties of protons (hydrogen nuclei in polar covalent bonds in fat or water molecules) when in the presence of a strong magnetic field. The magnetic field strength in a typical clinical MRI ranges from 1.5 - 3.0 Tesla, which is around thirty thousand times the strength of earth's magnetic field. In this environment, molecules with an odd number of nucleons such as ^{1}H, abundantly present in water and fat, will align their axis of precession with that of the local magnetic field. When we apply a radio pulse precisely at the resonating frequency of the molecule's precession, the molecules absorb a small amount of energy and the axis of precession flips vertically with respect to the orientation of the magnetic field, and will all initially synchronize their precession. MRI acquires an image by measuring the localized energy released when the orientation of these hydrogen nuclei snaps back into their original position; this is called relaxation. The contrast between tissue types observed across the various types of imaging is a result of the precise timing at which we measure this release of energy. This is called the Time to Echo (TE), and it is conventionally optimized to maximize the signal measured from either water or fat molecules (in structural imaging). Thus the image intensity at each voxel is the average signal measured at that specific TE, and is proportional to the composition of the underlying tissue [5].

When we talk about structural imaging, we generally refer to either T1 or T2 weighted images. T1 imaging directly measures the longitudinal relaxation time of the hydrogen nuclei, i.e. how quickly the nuclei snap back to their relaxed position. T2 imaging measures what is referred to as the transverse relaxation time, and is indicative of how quickly a proton's precession goes out of phase with respect to their neighboring nuclei. Defined tissue contrast in MRI is made possible because the underlying structure of molecules directly affects both their T1 and T2 relaxation times. In this context, fat is a more structured tissue type compared to water, as a proton lattice exists as a result of its rigid carbon chain, compared to more randomly distributed water molecules in fluid. In simplified terms, a proton lattice allows for a more rapid dispersion of the energy absorbed from the RF pulse, resulting in faster T1 and T2 relaxation times. It is important to emphasize that in practice we acquire T1 and T2 weighted images, where we measure an optimal combination of the tissue's T1 and T2 properties by varying both the interval at which we apply the RF pulse (called the repetition time - TR), and the interval at which we measure the signal (TE).

Heuristically, T1 weighted imaging will have much shorter TR and TE, and tissue with higher fat content will appear brighter in the image. With T2 weighted imaging, the TR and TE will be longer, so the signal from higher fat content decays before we measure the release of energy, and we are left with the signal from tissue with slower decaying relaxation, i.e. higher water content. This property is the driving force behind neuroimaging. The three main tissue types visible in brain MRI are white matter, grey matter, and cerebral spinal fluid (CSF). In fully developed brains, the axon portion of the neurons, collectively known as white matter, is wrapped in myelin, a fatty insulating sheath. This gives the axons a much shorter relaxation time, which makes white matter highly visible in T1 weighted imaging. Conversely, CSF is primarily water, and therefore has a longer relaxation time, showing up brightly on T2 weighted images. Grey matter conveniently falls in between white matter and CSF, allowing us to optimize the TR/TE combination in order to create the most contrast between these three tissue types.

Neonatal brain development, however, makes it difficult to use adult optimized imaging parameters in infants primarily due to two distinct properties. First, the neonatal brain has a much higher water content at birth, with rapid decrease in water content in both grey and white matter as the brain develops. This causes the overall tissue to have longer relaxation times, and provides a moving target for optimizing sequence parameters. But more importantly, neonatal white matter is largely unmyellinated at birth. Rapid myelination occurs through the first year of life, creating localized regions of higher T1 signal starting with the corticospinal tract, followed by visual and inter-hemispheric connections, and ending with cortical association connections. As a result, neonatal

development as observed by MRI can be roughly broken down into three distinct temporal stages [6]. The first stage is referred to as the infantile pattern (< 6 mo), and is approximately a reversal of the grey-white tissue contrast observed in the adult pattern (T1 and T2 are effectively switched). The isointense period (between 8-12 months of age), is characterized by very poor differentiation between white and grey matter, with regions of higher T1 signal where myellination is occurring. And finally, the early-adult pattern (> 12 months) is reached once the rate of myellination drops off, with tissue contrast more closely resembling a fully developed brain. All of this leads to sequence development challenges, requiring age specific optimization, as well as acquisition challenges, as smaller brains provide a lower signal to noise ratio (SNR) and neonates are more susceptible to motion in the loud MRI environment.

1.3 Structure Parcellation

By and large, Freesurfer (FS) is the most widely adopted automated cortical [7] and subcortical [8] segmentation software in the neuroimaging community. FS uses two independent, but similar in approach, probabilistic pipelines for parcellating the cortex and subcortical structures. One of the major challenges in automating brain segmentation is the inherent discord between geometrical structure and their conceptually generated labels. Structurally contiguous gyri and sulci may change labels as they cross cerebral lobes or perform different functions. Additionally, deep grey matter structures may not have well enough defined tissue contrast at their border, or may be susceptible to partial volume artifacts due to their proximity to white matter or CSF. To attenuate these intrinsic properties into the segmentation, FS and most other methods currently published use a probabilistic atlas as a prior in conjunction with the local image intensity to classify each voxel into discrete regions. The FS atlas was created by transforming manually pre-segmented subject brains into a common space and calculating the probability of each voxel belonging to each target class. This allows us to then transform our subject image into this probabilistic atlas space, and compute for each voxel the probability of it belong to a particular class given both its new position in this standardized space and the original voxel intensity. The final step in the FS segmentation pipeline uses a Markov random field to model local spatial relationships based on intensity and class, improving the segmentation at tissue boundaries.

A problem arises when applying these existing implementations, which have been designed and optimized for fully developed brains, to a neonatal cohort. As described in the previous section, the developing brain has drastically different tissue contrast early on, and it undergoes rather significant changes over a very short period of time. This renders both the spatial and intensity priors used in the posterior calculation inadequate. The change in tissue contrast is significant enough that ideally we would require a different set of priors for each gestational week until development reaches the early-adult pattern. This is both labor and cost prohibitive, and does not take into account any pathology that may delay or alter white matter development throughout the first year of life. Moving in the right direction, however, Gousias et. al have made publicly available a dataset, named ALBERT, of 20 manually segmented neonates ranging from 26 weeks gestation to full term [9]. The ALBERT subjects have been manually segmented into 50 cortical and subcortical structures, with both T1 and T2 contrast images made available. The benefit of having such a dataset available cannot be overstated, as generating such dataset in house would require an incredible amount of work. The authors state that for the 20 available subjects, a total of 18 person-months was required to generate all labels.

A promising implementation of automatic neonatal segmentation utilizing the existing ALBERT template has been proposed [10]. This work provides two possible methods for propagating the atlas labels into subject space: computing pairwise registrations between each ALBERT template and the target subject, or calculating one probabilistic atlas computed from all ALBERT subjects at once, followed by registering the subject to this atlas. NeBSS takes advantage of both of these methods, but in a less rigid approach to allow for increased variability in clinically abnormal populations.

2. Methods
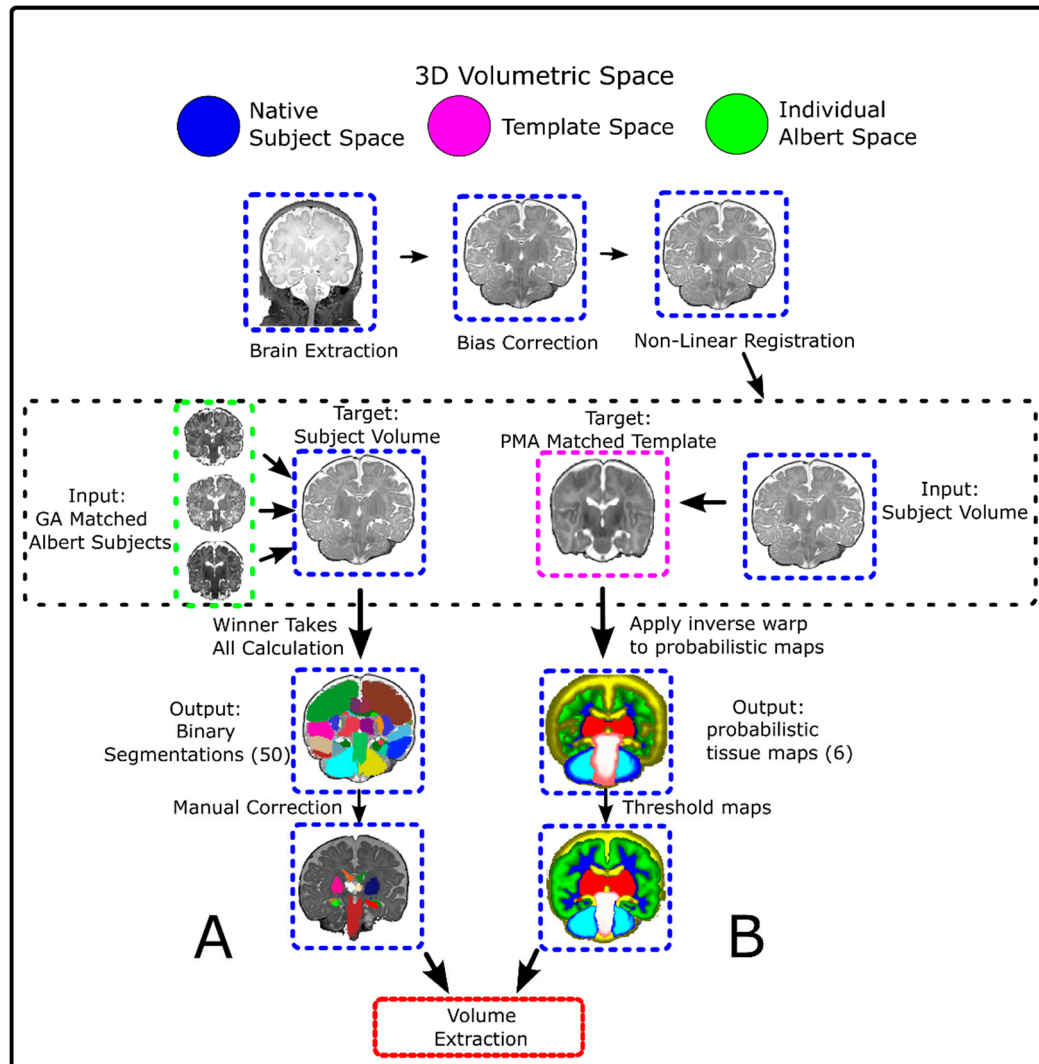
2.1 NeBSS: Neonatal Brain Structure Segmentation

Figure 1. Overview of NeBSS Pipeline. After brain extraction and bias correction, the pipeline splits into two branches (**A**) Non-linear registration of ALBERT subjects into subject space (**B**) non-linear registration of subject image into probabilistic atlas space, followed by the transformation of the atlas labels back into subject space. A manual correction step is required to ensure proper registration and correct for any additional errors that may arise due to pathology or low image quality.

Figure 1 shows an overview of the NeBSS pipeline. The first step in the pipeline is to pre-process the input MRI in order to remove unnecessary tissue and noise prior to the registration. We first run a brain extraction with FSL's Brain Extraction Tool (BET) [11]. Next, we use FSL's bias correction algorithm [12,13] to normalize the brain image, removing any intensity gradient artifact due to field inhomogeneity. We are left with an intensity normalized, clean subject brain still in the native scanner space, retaining its original volume and relative structural proportions. This will be the input to the subsequent registration steps. At this point the pipeline is effectively split into two independent, parallel branches. In branch A, the ALBERT neonatal parcellation dataset is used as the source to propagate onto the subject space the 50 discrete brain structures modeled by the atlas. Many clinical questions, however, also rely on the measurement of broader, global tissue volumes (branch B). Our solution to this is to use a probabilistic neonatal atlas created by Serag et. al [14], in similar fashion to the FS adult atlas. This atlas contains gestational age specific probability maps of gross tissue

volumes, namely total CSF, cortical grey matter, deep grey matter, white matter, brainstem and cerebellum, which will be propagated onto the subject's MRI in native space.

Both branches use the Advanced Normalization Tools (ANTS) algorithm [15,16] to calculate a non-linear transformation between a target image and the source image. ANTS is a robust symmetrical, diffeomorphic registration algorithm that continues to gain popularity in a plethora of neuroimaging (and other) applications [17]. A diffeomorphic registration maps an input to a target, where each point in the input has a one-to-one equivalent in the target. This is in contrast to elastic transformation algorithms, in which this property does not hold, making them inconvenient when mapping pre-defined segmentation labels between two spaces. A symmetric transformation means it has an invertible function, allowing us to both map our target to a template, as well as mapping the template back into the target.

Branch A uses ANTS to calculate a non-linear transformation of the pre-parcellated ALBERT subjects into native subject space. We choose four ALBERT subjects closest in gestation age to our subject to increase our registration accuracy and yield segmentations that more closely match the developmental stage of the subject. We then perform a voxelwise winner-takes-all calculation across the four transformed ALBERT subject labels to determine the structure classification at each voxel of our subject. If there is a tie, we randomly select one of the labels. We are left with 50 non-overlapping discrete regions mapped onto our subject space. Finally, before we extract our features, a manual correction step is necessary to ensure anatomical accuracy of our label propagation. If we start with a good quality image, minimal manual correction will be needed. However, severely abnormal brains or poor data quality will require more laborious manual correction. Once we are satisfied with the segmentation, we extract the structures of interest and calculate their volumes.

Branch B works in the reverse direction as branch A. We non-linearly transform our subject brain into the gestation age matched probabilistic atlas space. This is done because the higher resolution atlas provides a better registration target, particularly when dealing with lower quality images. The resulting transformation is then inversed, a convenient feature of the ANTS algorithm, allowing us to bring the tissue probability maps into the subject's native space. This gives probability maps for each tissue type in the atlas, transformed into the subject's native MRI space. The final tissue volumes are calculated by either binarizing each tissue map at a user defined probability threshold, or using a partial volume method by weighting each voxel's contribution to the final volume by its probability of being classified as that tissue. Higher quality input images yield higher probability thresholds, giving us more confidence in our final volume.

This pipeline is written in Python, and is structured using the Nipype framework [18]. Nipype is an open source, neuroimaging specific processing pipeline engine designed to create transparent and reproducible data processing graphs. Nipype contains built-in interfaces to the vast majority of existing neuroimaging software libraries, including FSL and ANTS. Additionally, Nipype provides tools for extending their library and creating project specific classes and interfaces. One of the biggest strengths of Nipype's pipeline engine is the exhaustive attention to efficient reproducible work. File checksums are calculated prior to each node in the processing graph, allowing us to rerun any step in the pipeline without having to compute all the previous nodes. This is achieved by verifying the checksum of the inputs at each node. If they match the previously computed files, that node will be skipped and only nodes containing a change in the input have to be re-run.

2.2 Evaluation of Pipeline Performance

We tested the performance of NeBSS by calculating the Dice similarity coefficient between the initial output of the pipeline and the manually corrected output using two representative structures: the cerebellum and hippocampus. The Dice similarity coefficient is a similarity metric calculated as

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

DSC is a measure of agreement between two image segmentations, ranging from 0 (no overlapping voxels) to 1 (perfect agreement). As our pipeline is designed to work with highly varied protocol and pathology, we believe that this measure is serves as a more realistic measure of the performance of the pipeline. We chose the cerebellum, as it is one of the largest structures segmented by NeBSS, and is susceptible to high registration errors due to large variation in size and position, especially in abnormally developing cohorts. We also measured the hippocampus, as it is a very small, centrally located structure with a highly complex shape. This was performed on three different cohorts: healthy, term born neonates (n=56), term born neonates with diagnosis of congenital heart disease (n=59), and preterm born neonates (n=65). Finally, we measured the inter- and intra-rater reliability of the pipeline by randomly choosing two subjects from each cohort, and measuring the DSC between repeated manual corrections performed by two users.

2.3 Comparison between 1.5 and 3.0 Tesla Field Strength

One common problem that arises when analyzing clinically acquired data, especially in retrospective studies of clinically abnormal cohorts, is that the available data is often acquired using a heterogeneous set of protocols and/or magnet field strength. Clinical translational multi-center neuroprotection neonatal brain studies can be conducted at different field strengths across multiple vendor platforms, leading to heterogeneity in imaging acquisition. The latter is a large enough confounder that datasets often need to be split between field strength, or outright excluded from analysis. To test the reliability of NeBSS across magnet strength, we performed a proof of concept analysis by comparing the gross structure segmentation outputs across three neonatal cohorts who had brain MRI's at our institution, using a variety of T2 weighted images acquired using both clinical and research optimized protocols on multiple vendors. The null hypothesis was that we would not observe any significant differences in volume between the cohorts when controlling for post-menstrual age (PMA) and normalizing each structure by total brain volume (to attenuate any influences due to pathology or additional confounders). When pooling the subjects into one group, we added the subject's clinical cohort as an added covariate. Statistical significant was tested using ANOVA in R [19]. Table 1 shows the distribution of subjects across field strengths. The cohorts are as follows: term born neonates with congenital heart disease (CHD), preterm-born neonates imaged at approximately term-equivalent age, and term born healthy controls.

Table 1.   Summary of three cohorts scanned at 1.5 and 3.0 Tesla field strength. Term born neonates with congenital heart disease (CHD), preterm-born neonates imaged at approximately term-equivalent age, and term born healthy controls who were scanned at our institution, varied between standard clinical protocols and research specific sequences.

	Field Strength (Tesla)	
Cohort	**1.5**	**3.0**
All Combined	17	71
CHD	7	25
Control	6	27
Preterm	4	19

3. Results

3.1 Workflow

NeBSS works optimally with T2 weighted images, as they have much better tissue contrast in neonates, especially grey-white matter differentiation. Volumetric T2 imaging, however, can be more susceptible to motion artifacts compared to T1 weighted images. As a compromise, NeBSS provides the user an alternative. We can instead acquire motion corrected T2 images, such as fast spin echo

(FSE), up to 2 mm slice thickness. By linearly registering this image to a volumetric T1, we retain the volumetric accuracy while leveraging the better tissue contrast from the T2 image.

The structural MRI can be fed into the pipeline directly from acquisition using a graphical user interface (GUI). Figure 2 shows the GUI used in NeBSS. The only inputs required from the user are the subject's gestation age, post-menstrual age, and the coordinates for a bounding box in order to crop excessive extra-cranial structures from the input image. This last parameter is necessary when imaging neonates, as we don't always have control over what position they are in; as long as they are comfortable and sleeping. Because of this, often the field of view ends up capturing the infant's arm and shoulders. This can be problematic for the brain extraction and registration steps.
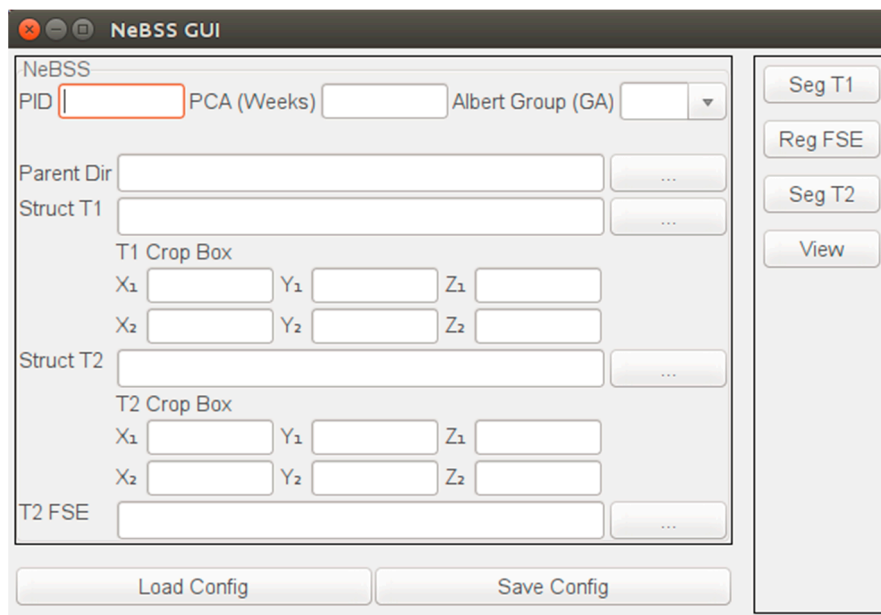

Figure 2. NeBSS Graphical User Interface (GUI). Simple GUI that allows the user to input all the necessary parameters for the neonatal segmentation. The GUI also provides a convenient method of registering a non-volumetric T2 weighted image to a volumetric T1 weighted image for better volumetric accuracy while leveraging the better tissue contrast observed in T2 weighted images in neonates.

Once NeBSS finishes running, the user is required to verify the label registrations, and perform any manual corrections if necessary. Any imaging specific software can be used for this step, but we recommend ITK-SNAP [20] as it is purpose built for volumetric label creation and editing. If the user is satisfied with the segmentations, the final step is simply to extract the volumes and 3D structures of interest for subsequent analysis.

3.2 Evaluation of Pipeline Performance

Figure 3 shows the Dice coefficient between the pipeline output and manual correction for all subjects in each cohort. Cerebellum showed higher Dice coefficient, indicating relatively less manual correction was necessary compared to hippocampus. The Control subjects required less correction than both CHD and preterm cohorts (p< 0.001 for Cerebellum, p < 0.025 for Hippocampus), due to fewer structural abnormalities and more consistent imaging protocol. This is summarized in Table 2. We then inspected the subjects with the lowest Dice scores, and as expected these subjects often presented with poor image quality and/or structural abnormalities that would cause them to fail more rigorous segmentation methods. Figure 4 shows sample images from the subjects with the lowest Dice scores across all cohorts.
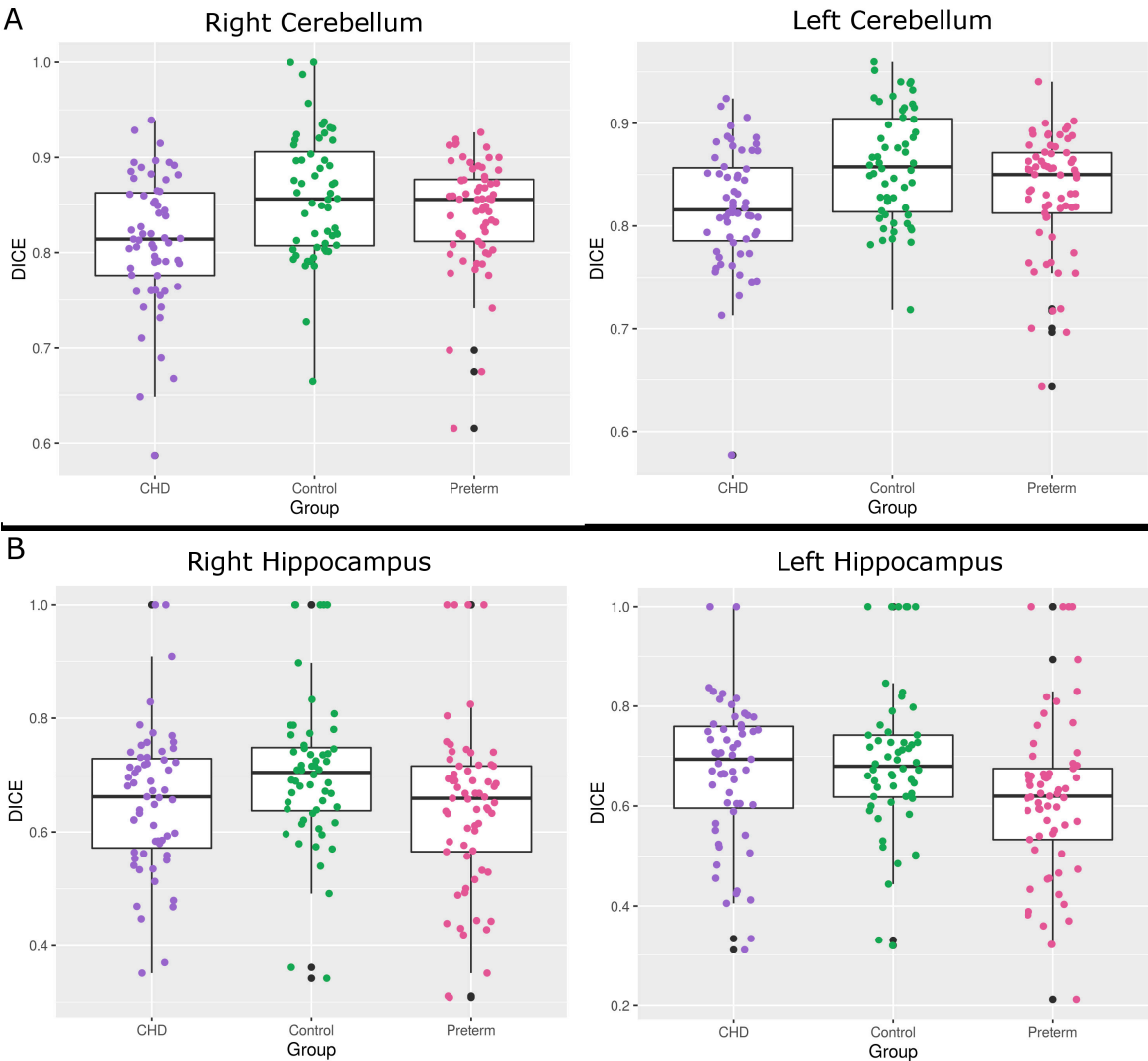
Figure 3. Distribution of Dice Similarity Scores between pipeline output and manual correction in A) cerebellum and B) hippocampus. Cerebellum showed higher Dice coefficient, indicating relatively less manual correction was necessary compared to hippocampus. The Control subjects required less correction than both CHD and preterm cohorts, due to less structural abnormalities and more consistent imaging protocol.

Table 2. Mean Dice similarity score between pipeline output and manual correction in two representative structures.

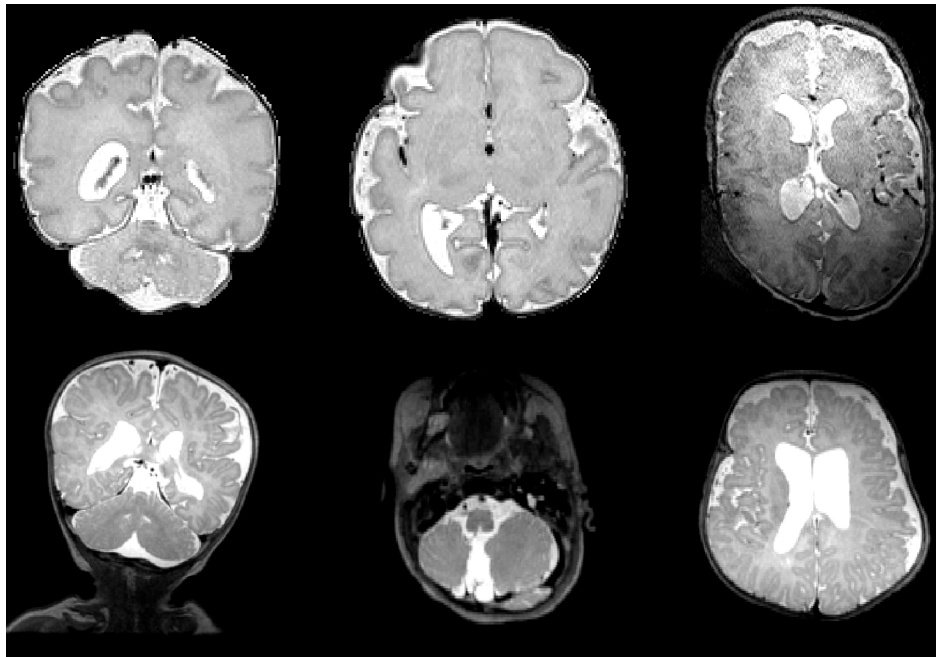|	Structure			
	L Cerebellum	**R Cerebellum**	**L Hippocampus**	**R Hippocampus**
Cohort	mean (sd)	mean (sd)	mean (sd)	mean (sd)
Control	0.859 (0.05)	0.858 (0.07)	0.690 (0.15)	0.705 (0.13)
CHD	0.818 (0.06)	0.812 (0.07)	0.667 (0.14)	0.655 (0.13)
Preterm	0.833 (0.06)	0.842 (0.06)	0.615 (0.17)	0.641 (0.15)

Figure 4. Sample images from the subjects with the lowest Dice coefficient across all cohorts. Note the asymmetric and dilated ventricles pose a challenge for registration based segmentation methods. Additionally, noise and motion can reduce the accuracy of these methods. However, the exclusion of these subjects would greatly limit clinical research. These subjects would likely fail more constrained automated segmentation methods.

The inter- and intra-rater reliability (Table 3) of the pipeline showed strong similarity coefficient in both structures, with the cerebellum again showing more consitent segmentation across repeated measures.

Table 3. Mean Dice coefficient of repeated manual correction across six subjects. For inter-rater reliability, two users performed manual correction on the same set of six subjects. For intra-rater reliability, one user performed an additional set of corrections on the same set of subjects.

	Structure			
	L Cerebellum	**R Cerebellum**	**L Hippocampus**	**R Hippocampus**
Measure	mean (sd)	mean (sd)	mean (sd)	mean (sd)
Inter-Rater	0.941 (0.01)	0.951 (0.01)	0.811 (0.19)	0.84 (0.11)
Intra-Rater	0.951 (0.03)	0.941 (0.03)	0.89 (0.03)	0.87 (0.09)

3.3 Comparison between 1.5 and 3.0 Tesla Field Strength

Table 4 shows the ANOVA p-values of each structure of interest when compared between acquisitions at 1.5 and 3 Tesla field strength. No significant volumetric differences were observed in any of the structures of interest across the cohorts. This supports the goal of NeBSS of being agnostic to both protocol and field-strength.

Table 4 ANOVA p-values comparing three cohorts scanned at 1.5 and 3.0 Tesla field strength. Volumes were extracted from T2 weighted images acquired using both clinical and research optimized protocols on multiple vendors. Volumes were corrected for post-menstrual age and normalized by total brain volume. The combined analysis also controlled for group.

Cohort	Structure			
	Cerebellum	CSF	Cortex	WM
All Combined	0.605	0.998	0.450	0.957
CHD	0.119	0.692	0.795	0.242
Control	0.513	0.736	0.080	0.340
Preterm	0.731	0.805	0.765	0.194

4. Discussion

We have introduced NeBSS, a robust semi-automated neonatal brain segmentation pipeline. NeBSS fills a need in clinical translational research in neonatal imaging, where existing automated or semi-automated implementations are too rigid to be successfully applied to clinically significant cohorts. NeBSS is able to reliably parcellate neonatal images acquired using heterogeneous imaging protocols and magnet field strength, with minimal manual correction, especially when compared to traditional manual segmentation methods. The output volumes show no observable bias as a result of magnet field strength, and we are confident that it can be applied to heterogeneous clinical data. The software is further strengthened by being fully open source, allowing the research community to incrementally improve on its performance and usability transparently.

The software implementation is still in its infancy, with room for improvement in both computational efficiency and usability. Future work will focus on reducing the computational time required for segmenting poor quality images, improving the overall accuracy of the initial segmentation to further reduce the labor required to parcellate neonatal brain images, and streamlining software installation and workflow.

5. Availability

NeBSS is open source and published under the BSD License. The source code is available at www.github.com/PIRCImagingTools/NeBSS.

Author Contributions: R.C. developed the software, collected and analyzed the data, performed the experiments, and contributed to the manuscript; A.Z. collected and analyzed the data; V.G. contributed to the writing of the manuscript and provided technical support; A.P. contributed to the writing of the manuscript, provided technical support and designed the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI: Multidisciplinary Digital Publishing Institute
DOAJ: Directory of open access journals
CHD: Congenital Heart Disease
NeBSS: Neonatal Brain Structure Segmentation
CSF: Cerebral Spinal Fluid
FS: Freesurfer
ANTS: Advanced Normalization Tools
PMA: Post-Menstrual Age
GUI: Graphical User Interface

References

1. Li G, Li Y, Wang Y, Shen D. Consistent sulcal parcellation of longitudinal cortical surfaces. In: *Proceedings - International Symposium on Biomedical Imaging*. Vol 57. NIH Public Access; 2011:337-340. doi:10.1109/ISBI.2011.5872418.

2. Kazemi K, Noorizadeh N. Quantitative Comparison of SPM, FSL, and Brainsuite for Brain MR Image Segmentation. *J Biomed Phys Eng*. 2014;4(1):13-26. http://www.ncbi.nlm.nih.gov/pubmed/25505764. Accessed September 8, 2016.

3. Kochunov P, Rogers W, Mangin JF, Lancaster J. A library of cortical morphology analysis tools to study development, aging and genetics of cerebral cortex. *Neuroinformatics*. 2012;10(1):81-96. doi:10.1007/s12021-011-9127-9.

4. Kersbergen KJ, Makropoulos A, Aljabar P, et al. Longitudinal Regional Brain Development and Clinical Risk Factors in Extremely Preterm Infants. *J Pediatr*. 2016. doi:10.1016/j.jpeds.2016.08.024.

5. Paus T, Collins DL, Evans AC, Leonard G, Pike B, Zijdenbos A. Maturation of white matter in the human brain: A review of magnetic resonance studies. *Brain Res Bull*. 2001;54(3):255-266. doi:10.1016/S0361-9230(00)00434-2.

6. Lenroot RK, Giedd JN. Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neurosci Biobehav Rev*. 2006;30(6):718-729. doi:10.1016/j.neubiorev.2006.06.001.

7. Fischl B, Van Der Kouwe A, Destrieux C, et al. Automatically Parcellating the Human Cerebral Cortex. *Cereb Cortex*. 2004;14(1):11-22. doi:10.1093/cercor/bhg087.

8. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002;33(3):341-355. doi:10.1016/S0896-6273(02)00569-X.

9. Gousias IS, Edwards AD, Rutherford MA, et al. Magnetic resonance imaging of the newborn brain: Manual segmentation of labelled atlases in term-born and preterm infants. *Neuroimage*. 2012;62(3):1499-1509. doi:10.1016/j.neuroimage.2012.05.083.

10. Gousias IS, Hammers A, Counsell SJ, et al. Magnetic Resonance Imaging of the Newborn Brain: Automatic Segmentation of Brain Images into 50 Anatomical Regions. Rodriguez-Fornells A, ed. *PLoS One*. 2013;8(4):e59990. doi:10.1371/journal.pone.0059990.

11. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. *Neuroimage*. 2012;62(2):782-790.

12. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging*. 2001;20(1):45-57. doi:10.1109/42.906424.

13. Vovk U, Pernuš F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging*. 2007;26(3):405-421. doi:10.1109/TMI.2006.891486.

14. Serag A, Aljabar P, Ball G, et al. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. *Neuroimage*. 2012;59(3):2255-2265. doi:10.1016/j.neuroimage.2011.09.062.

15. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12(1):26-41. doi:10.1016/j.media.2007.06.004.

16. Avants BB, Tustison NJ, Stauffer M, Song G, Wu B, Gee JC. The Insight ToolKit image registration framework. *Front Neuroinform*. 2014;8(April):44. doi:10.3389/fninf.2014.00044.

17. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 2011;54(3):2033-2044. doi:10.1016/j.neuroimage.2010.09.025.

18. Gorgolewski K, Burns CD, Madison C, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*. 2011;5. doi:10.3389/fninf.2011.00013.

19. R Development Core Team. R: A Language and Environment for Statistical Computing. 2011. http://www.r-project.org/.

20. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-1128. doi:10.1016/j.neuroimage.2006.01.015.