Neuroinformatics (2019) 17:443–450
https://doi.org/10.1007/s12021-018-9408-7

ORIGINAL ARTICLE

Automatic Thalamus Segmentation on Unenhanced 3D T1 Weighted Images: Comparison of Publicly Available Segmentation Methods in a Pediatric Population

Salem Hannoun[1,2] · Rayyan Tutunji[3] · Maria El Homsi[3] · Stephanie Saaybi[3] · Roula Hourani[3]

Published online: 14 December 2018

Abstract

The anatomical structure of the thalamus renders its segmentation on 3DT1 images harder due to its low tissue contrast, and not well-defined boundaries. We aimed to investigate the differences in the precision of publicly available segmentation techniques on 3DT1 images acquired at 1.5 T and 3 T machines compared to the thalamic manual segmentation in a pediatric population. Sixty-eight subjects were recruited between the ages of one and 18 years. Manual segmentation of the thalamus was done by three junior raters, and then corrected by an experienced rater. Automated segmentation was then performed with FSL Anat, FIRST, FreeSurfer, MRICloud, and volBrain. A mask of the intersections between the manual and automated segmentation was created for each algorithm to measure the degree of similitude (DICE) with the manual segmentation. The DICE score was shown to be highest using volBrain in all subjects (0.873 ± 0.036), as well as in the 1.5 T (0.871 ± 0.037), and the 3 T (0.875 ± 0.036) groups. FSL-Anat and FIRST came in second and third. MRICloud was shown to have the lowest DICE values. When comparing 1.5 T to 3 T groups, no significant differences were observed in all segmentation methods, except for FIRST ($p = 0.038$). Age was not a significant predictor of DICE in any of the measurements. When using automated segmentation, the best option in both field strengths would be the use of volBrain. This will achieve results closest to the manual segmentation while reducing the amount of time and computing power needed by researchers.

Keywords Thalamus · Magnetic resonance imaging · Pediatric imaging · Manual and automated segmentation · Similarity index

Abbreviations

BBB	Blood-Brain Barrier
DICE	DICE Similarity Index
GM	Gray Matter
FIRST	FMRIB's Integrated Registration and Segmentation Tool
FSL	FMRIB Software Library
MRI	Magnetic resonance imaging
WI	Weighted Images
WM	White Matter

Salem Hannoun and Rayyan Tutunji Indicative of shared first authorship/ Equal Contribution

✉ Roula Hourani
rh64@aub.edu.lb

[1] Abu-Haidar Neuroscience Institute, Faculty of medicine, American University of Beirut, Riad El-Solh, Beirut 1107 2020, Lebanon

[2] Nehme and Therese Tohme Multiple Sclerosis Center, Faculty of medicine, American University of Beirut, Riad El-Solh, Beirut 1107 2020, Lebanon

[3] Department of Diagnostic Radiology, American University of Beirut Medical Center, Riad El-Solh, P.O.Box: 11-0236, Beirut 1107 2020, Lebanon

Introduction

Magnetic resonance imaging (MRI) has been extensively used to monitor the pathophysiology of the brain in specific regions of interest. In that scope, brain segmentation has been applied, particularly for the assessment of age related volumetric changes in patients across the lifespan (Courchesne et al. 2000), or as a comparative tool for the clinical assessment of different diseases versus control populations (Csernansky et al. 2004). The manual tracing of subcortical gray matter (GM) such as the thalamus, requires a high level of expertise. Their involvement is increasingly recognized as an important pathophysiological feature. Indeed, the thalamus mediates communication among sensory, motor, and associative brain

regions. It plays a major role in the regulation of consciousness, alertness, arousal, and attention and is considered part of the limbic system (Duan et al. 2007). Damage to the thalamus can be associated with motor and somatosensory disturbances, and cognitive decline. Thalamic changes in terms of volume and intensity are reported in various neurological disorders such as hypoxic-ischemic brain injury, some metabolic diseases, schizophrenia (Ganzola et al. 2014), bipolar disorder (Sassi et al. 2002), obsessive compulsive disorder (Rotge et al. 2009), post-traumatic stress disorder (Jatzko et al. 2006), and depression (Koolschijn et al. 2009). The normal development of the thalamus in children is therefore of great importance to insure its major role of hub for the majority of incoming sensory and motor functions.

Several methods have been previously developed and introduced to perform automatic and semi-automatic regional segmentation as accurately and specifically as possible. Such tools accelerate data analysis in large studies, and deliver reproducible and consistent outcomes, which are crucial for obtaining reliable results (Mulder et al. 2014). However, manual segmentation by expert operators of brain structures namely the thalamus, remain the gold standard. Still, it is a tedious, laborious, time-consuming, and not reproducible task, constituting a difficult challenge even for an expert radiologist.

When performing brain segmentation, 3DT1-weighted images (WI) are optimal. This pulse sequence fulfills the requirements for spatial resolution and contrast-to-noise ratio between GM and white matter (WM) structures at a tolerable measuring time. However, the anatomical structure of the thalamus renders its segmentation harder on conventional MRI due to the thalamic low tissue contrast, noise, and unclear and not well-defined boundaries. Thalamic segmentation requires therefore the use of higher resolution anatomical MR images or magnetization transfer images. Unfortunately, with the increase of the images' resolution, the acquisition time is also increased, which is challenging in a clinical setting. Similarly, some segmentation methods such as thresholding, and region growing give poor results because they only use image intensity information. It is therefore why, in this study, we aimed to emphasize the importance of using recently developed or updated, publicly available, automated segmentation algorithms (FIRST, FSL-Anat, Freesurfer, volBrain, and MRICloud) by testing their precision and accuracy to segment the thalamus.

Materials and Methods

Subjects

In this retrospective study, 126 MRI scans for subjects between the ages of one and 18 years were identified from our brain development database study, from October 2008 to August 2016. Only subjects reported to have normal brain MRIs by a neuroradiologist with more than 20 years of experience, were included in this study. Exclusion criteria included: subjects under one year old due to the confounding factors of incomplete myelination; presence of a brain tumor; history of repeated seizures and epilepsy; indications of mental disabilities charts; brain infarcts or hemorrhage; diffuse brain abnormalities on MRI like leukodystrophy or atrophy; lesions around and in the thalamus and gray matter; periventricular lesions; subjects with only gadolinium enhanced images performed. The reports and clinical indications for getting an MRI of patients were reviewed, and those who met the criteria had their images uploaded and reviewed. The most common indications for undergoing MRIs were: headache, hearing problems, retinoblastomas, and suspicion of one time only seizure-like activity. The Institutional Review Board (IRB) of our institution approved this retrospective cross-sectional study and waived the requirement to obtain informed consent. Research has also been conducted in full accordance with the World Medical Association Declaration of Helsinki.

MRI Processing

MRI images were acquired on either a 1.5 T ($N = 39$ subjects) or a 3 T ($N = 29$ subjects) scanner (Ingenia, Phillips). All subjects underwent only one MRI acquisition on either the 1.5 T or the 3 T machines. Positioning is done parallel to the inferior border of the corpus callosum on the axial plane and perpendicular to the hippocampus on the coronal plane. Image parameters for the 3DT1 at the 1.5 T scanner were: TR/TE = 7.5/3.4 ms, matrix = 220x216mm, spatial resolution = 0.859 × 0.859 mm, slice thickness = 1.1 mm. Image parameters for the 3 T scanner were: TR/TE = 8.2/3.7 ms, matrix = 240x222mm, spatial resolution = 0.937 × 0.937 mm, slice-thickness = 1 mm. Before introducing MRI images to the battery of segmentations, a quality control step was performed to rule out and exclude any images with major artifacts that could implicate an error during segmentation. Upon visual inspection of the images by an experienced operator with more than twelve years of experience, 19 out of the 126 subjects were excluded due to the bad quality of their images (movement artifacts and bad signal to noise ratio). Thirty nine other subjects had only 3DT1 images acquired after gadolinium administration, making them unsuitable for segmentation as gadolinium affects the segmentation outcomes (Hannoun et al. 2018). This brought the total number of normal subjects with 3DT1-WI to 68 (36 females (age = 10.98 ± 4.66 years) & 32 males (age = 10.71 ± 5.54 years)).

Manual Segmentation

Manual segmentation of the thalamus was performed by three junior raters on 3DT1-WI using the Medical Imaging Interaction Toolkit (MITK v.2016.11). A fourth senior

experienced rater (neuroscientist with more than 12 years of experience in brain anatomy) reviewed, corrected, and approved all manually segmented masks of the thalamus. To test for interrater reliability for the manual segmentation, all raters segmented the thalami of a randomly selected set of 33 subjects. Thalamic boundaries were defined as followed: The anterior and posterior boundary of the thalamus were defined by the posterior aspect of the interventricular foramen the Pulvinar nucleus respectively. The wall of the third-ventricle served as medial boundary while the lateral aspect was defined by the posterior limb of the internal capsule. The superior boundaries were measured at the fornix and the inferior at the level of the hypothalamic sulcus (Felten et al. 2010). Thalamic masks were adjusted in all three planes: axial, coronal and sagittal. Thalamus manual delineation took around 30 to 45 min per subject.

Automatic Segmentation

The following provides a brief overview of the methods used for thalamic automatic segmentation: volBrain, MRICloud, FSL (FIRST & FSL_Anat), and Freesurfer. Default parameters were used for all segmentation algorithms.

volBrain volBrain is an online MRI brain volumetry system (http://volbrain.upv.es). It is a pipeline of processes aimed to automatically analyze MRI brain data (around ten minutes of processing time). It provides the volumes of the main intracranial cavity tissues and some macroscopic areas. Finally, automatic subcortical structure segmentation is performed, and related volumes and label maps are provided.

MRICloud MRICloud provides a fully automated cloud service for brain parcellation of MPRAGE images based on Multiple-Atlas Likelihood Fusion algorithm, JHU multi-atlas inventories with 286 defined structures, and an Ontology Level Control technology (https://mricloud.org). The atlas used for the processing of our data was the pediatric_286labels_11atlases_V5L.

FSL FMRIB Software Library (FSL5.0) is a comprehensive library of analysis tools for MRI brain imaging data (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL). FSL comprises two tools for subcortical GM segmentation: FMRIB's Integrated Registration and Segmentation Tool (FIRST) and FSL_Anat.

I) **FIRST.** FIRST is a model-based segmentation/registration tool that uses manually segmented models. The shape and appearance model are based on multivariate Gaussian assumptions. Shape is expressed as a mean with modes of variation. FIRST searches through linear combinations of shape modes of variation for the most probable shape instance given the observed intensities in a T1-WI.

II) **FSL Anat.** FSL_Anat provides a general pipeline for processing anatomical images. Most of the pipeline involves standard use of FSL tools. The stages in order are: 1-reorient the images to the standard (MNI) orientation, 2-automatically crop the image, 3-bias-field correction, 4-registration to standard space (linear and non-linear), 5-brain-extraction, 6-tissue-type segmentation, 7-subcortical structure segmentation using FIRST.

Freesurfer FreeSurfer image analysis suite (v5.3) provides a full processing stream for structural MRI data that includes skull stripping, B1 bias field correction, and GM/WM segmentation (http://freesurfer.net). It also includes reconstruction of cortical surface models, labeling of regions on the cortical surface, subcortical brain structures, nonlinear registration of the cortical surface of an individual with a stereotaxic atlas and statistical analysis of group morphometry differences.

Statistical Analysis

To create a reliable value for statistical analysis, we generated the Sorensen-Dice similarity index (DICE). This index measures the similarity between two masks, more in detail, it compares the number of identical voxel pairs between the two masks (Næss-Schmidt et al. 2016). The resulting values are bounded between zero and one. These were created by multiplying the manual segmentation with the automated one for each subject; this generates a mask consisting of shared areas of each measurement. This was then divided by the total volume of the manual and automated measurements. The resulting equation is as follows:

$$DICE = \frac{2(Manual \cap Auto)}{Manual + Auto}.$$

Data were then analyzed using Stata14.1 (StataCorp, TX, USA). Interrater reliability for the manual segmentation performed by all raters was measured using a weighted Kappa. The Shapiro-Wilk test was first applied to test the normality of our data. Subjects were divided in two groups depending on the field strength: 1.5 T ($N = 39$ subjects) or a 3 T ($N = 29$ subjects). Gender and age comparisons were made using the Wilcoxon rank-sum test. This test was also used to test the segmentation differences between both 1.5 T an 3 T groups. A linear regression was conducted to examine the relationship between DICE values for each method and age and repeated for each field strength. A One-way Anova test with a Bonferroni correction for multiple comparisons was performed to test the differences between the DICE values of the techniques. An intraclass

correlation (ICC) between the five segmentation techniques DICE values was also performed.

Results

The average volume of the manually segmented gold standard thalamus in all subjects was 13.23 ± 2.34 cm^3 (females = 13.24 ± 2.40 cm^3; males = 13.21 ± 2.32 cm^3). Results of the automatic segmentations are illustrated in Fig. 1 and Table 1. There were no significant sex or age differences in the similitude indices. Interrater reliability for manual segmentation as measured by weighted Kappa was high (0.908, $p < 0.001$), indicating a high level of agreement between all raters (agreement = 98.04%; expected agreement = 82.60%).

There was an overestimation of the mean thalamic volumes in all methods except for MRICloud and volBrain, which showed smaller volumes compared to the gold standard.

The DICE score was shown to be highest using volBrain in all subjects as well as in the 1.5 T and 3 T group (Table 2, Fig. 2). FSL-Anat and FIRST came in second and third, followed by Freesurfer. MRICloud was shown to have the lowest median DICE (Min-Max) value at 0.736 (0.580–0.808) for all subjects, 0.736 (0.580–0.808) for the 1.5 T group, and 0.738 (0.641–0.805) for the 3 T group.

We first examined the differences in DICE scores between the 1.5 T and 3 T groups. The Wilcoxon rank-sum test conducted for each segmentation method revealed no significant differences in DICE values when comparing both field strengths except with FIRST where the 3 T segmentations had higher DICE than the 1.5 T segmentations ($p = 0.038$). Similitude indexes tended to be higher in the 3 T compared to the 1.5 T group (Table 2, Fig. 2).

We secondly studied the impact of age on the DICE values generated from the manual and automated segmentation techniques. Results show no significant effect of age on DICE values in any of the applied segmentation methods.

One-way Anova test showed volBrain DICE values to be significantly the highest (FSL-Anat and FIRST $p < 0.01$; Freesurfer and MRICloud $p < 0.001$) and MRICloud DICE values to be the lowest ($p < 0.001$) when compared to the other techniques. No significant differences were observed when comparing FSL-Anat, FIRST and Freesurfer between themselves. The estimated correlation between individual ratings was 0.124, indicating little similarity between ratings within a target (subject), low reliability of individual target ratings, or no target effect. The estimated intraclass correlation between ratings averaged over $k = 5$ judges (segmentation techniques) is higher, 0.414. The estimated intraclass correlation measures the similarity or reliability of mean ratings from groups of five judges. We also have statistical evidence to reject the null hypothesis that neither ICC is zero based on confidence intervals and the F test.

Discussion

Along with the advances of medical imaging, MRI has become a non-invasive soft tissue contrast imaging modality; it provides information about shape and size of brain structures

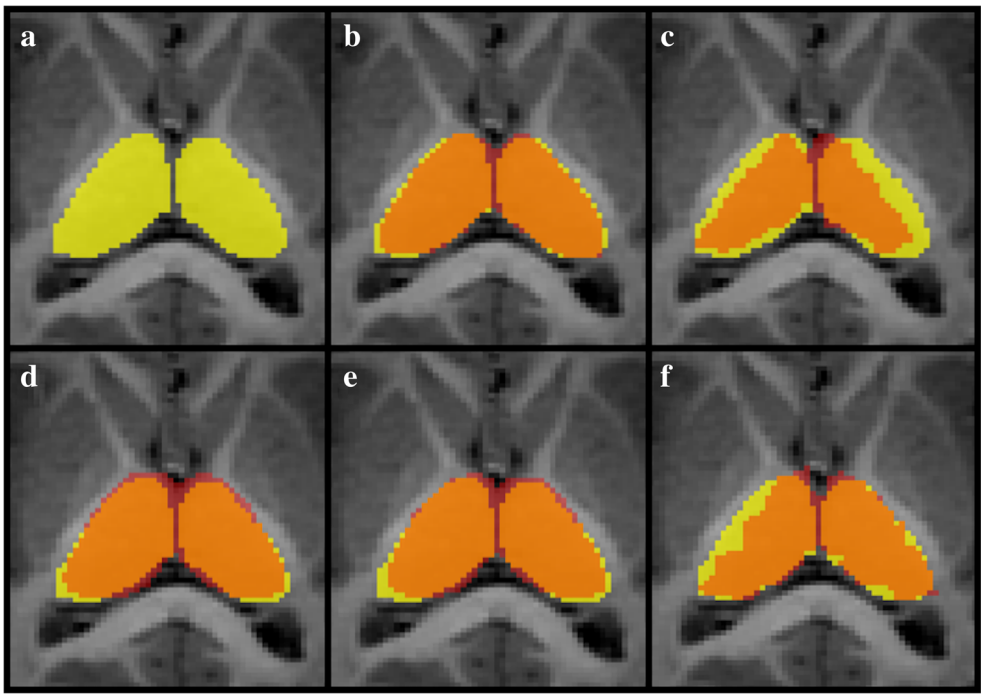

Fig. 1 Image of the manually segmented Thalamus in yellow (**a**) superimposed on the automatic segmentation done with volBrain (**b**), MRICloud (**c**), FSL Anat (**d**), FIRST (**e**), and FreeSurfer (**f**). Areas in red represent voxels where no intersection was seen between manual and automatic segmentations. Areas in orange represent voxels where both manual and automatic segmentations intersected

Table 1 Median (Min-Max) volume values in mm^3 for manual and automatic segmentation methods in all subjects as well as in in the 1.5 T and 3 T groups

	All subjects	1.5 T	3 T
Manual	12912.4 (6942.1–21291.4)	12815.3 (6942.1–20035.2)	13595.8 (8391.5–21291.4)
FSL-Anat	15279.4 (9764.5–20786.4)	15371.0 (11207.9–20786.4)	14540.7 (9764.5–18169.5)
FIRST	15150.0 (9101.0–20178.4)	15708.6 (11218.2–20178.4)	14189.8 (9101.0–17522.8)
FreeSurfer	14183.2 (7672.3–18734.9)	14277.6 (9745.6–18734.9)	13783.8 (7672.3–16788.8)
MRICloud	10141.6 (5136.9–13037.9)	10251.4 (6394.9–13037.9)	9764.6 (5136.9–12520.8)
volBrain	11978.1 (7183.2–16574.6)	12007.0 (7945.1–16574.6)	11882.0 (7183.2–13708.2)

without exposing the patient to ionization radiation (Liang and Lauterbur 2000). In current clinical routine, the images of different MRI sequences are employed for the diagnosis and delineation of brain abnormalities (lesions, tumors, hyperintensities, etc.…). Among these sequences, the non-enhanced 3DT1-WI allows an easy delineation of brain anatomy. However, field strength plays a major role in providing a better segmentation quality.

In this study, we aimed to investigate the performance of publicly available automatic segmentation techniques on non-enhanced 3DT1-WI acquired on 1.5 T or 3 T MRIs. The implementation of automated techniques makes large scale populations studies much easier to conduct. Manual delineation of specific regions of interest or even whole brain segmentation could be tedious and time consuming. To this end, several segmentation techniques have been developed, each based on different algorithms, some being semi-automatic, others fully automatic. Automatic segmentation has the potential to positively impact clinical medicine by freeing physicians from the burden of manual labeling. It can also provide robust and quantitative measurements to aid in diagnosis and better understanding of the disease. Indeed, the MRI evaluation of thalamic volume can help differentiate multiple sclerosis from other diseases that cause white matter abnormalities (Solomon et al. 2017). It has also been identified early in the disease course of multiple sclerosis, including pediatric and pre-symptomatic cases as well as clinically and radiologically isolated syndromes (Aubert-Broche et al. 2011; Azevedo et al. 2015; Bergsland et al. 2012). Thalamic volume loss has also been shown to be a predictive marker of progression in various diseases and disorders including Parkinson's disease (Lee et al. 2011), Alzheimer's disease (De Jong et al. 2008), traumatic brain injuries (Fearing et al. 2008), and some psychological disorders such as bipolar disorder (Radenbach et al. 2010), autism spectrum disorders (Tsatsanis et al. 2003) and Obsessive Compulsive Disorder (Rosenberg et al. 2000). Different acute and chronic clinical situations including systemic metabolic diseases such as Tay Sach's, Krabbe's disease, and Refsum, and vascular conditions such as hypoxic-ischemic brain injury (Ricci et al. 2006) can additionally affect the basal ganglia and the thalami bilaterally causing change in thalamic size.

A key problem in medical imaging is the automatic segmentation of MRI images via different processes. Some labs such as FMRIB, have even developed multiple tools including FAST (for brain tissue segmentation), FIRST (morphometric subcortical segmentation) and FSL-Anat (anatomical processing pipeline combining both FIRST and FAST) (Smith et al. 2004). These tools can differ in their outcome as well as their precision, accuracy, parameters, and atlases they use to segment structures of interest. The use of different registration techniques and several adult atlases can even improve the reliability of segmentation (Aljabar et al. 2009). In contrast, pediatric populations segmentation improves when applying manually delineated pediatric atlases (Murgasova et al. 2007). Furthermore, specialized segmentation algorithms as previously mentioned prove to be superior when used with pediatric populations (Weisenfeld and Warfield 2009). In highly variable brain MRI data, it may be difficult for these segmentation tools to properly model the thalamus.

Table 2 DICE Median (Min-Max) values for each segmentation method in all subjects as well as in in the 1.5 T and 3 T groups

	All subjects	1.5 T	3 T	p*
FSL-Anat	0.859 (0.726–0.907)	0.851(0.726–0.907)	0.862(0.781–0.892)	0.152
FIRST	0.860 (0.724–0.897)	0.851 (0.724–0.896)	0.865 (0.796–0.897)	0.038
FreeSurfer	0.838 (0.753–0.883)	0.836 (0.753–0.883)	0.841 (0.814–0.882)	0.213
MRICloud	0.736 (0.580–0.808)	0.736 (0.580–0.808)	0.738 (0.641–0.805)	0.669
volBrain	0.882 (0.772–0.924)	0.879 (0.775–0.924)	0.888 (0.772–0.913)	0.642

*Wilcoxon rank-sum test applied to test the significance of differences between the 1.5 T and 3 T groups

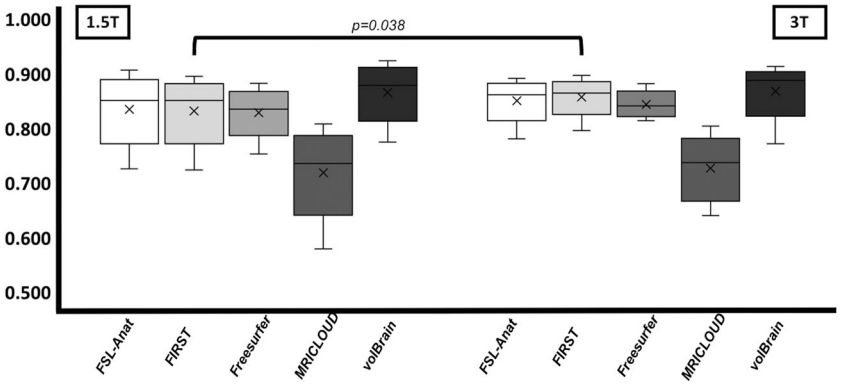

Fig. 2 Box plot showing the DICE for the 1.5 T and the 3 T groups. The DICEs differences for each method between both field strengths are not significant except for the FIRST segmentation

To address this issue, several technics such as multi-atlas label fusion, multi-atlas segmentation techniques, patch-based methods have been introduced (Næss-Schmidt et al. 2016). However, these new approaches are still not yet publicly available, thereby less used in clinical research. In the case of brain development, one problem is the automatic segmentation of certain brain regions such as the thalamus. We have therefore examined the degree of overlap between five different segmentation techniques (FSL-Anat, FIRST, volBrain, FreeSurfer, MRICloud) and the gold standard manual segmentation. We found that volBrain segmentation had the best outcome in terms of accuracy with regard to the manual segmentation with a median DICE of 0.882 (0.772–0.924) for all subjects, 0.879 (0.775–0.924) for the 1.5 T group, and 0.888 (0.772–0.913) for the 3 T group. On the other end of the spectrum, MRICloud proved to have the lowest values with a DICE of 0.736 (0.580–0.808)for all subjects, 0.736 (0.580–0.808) for the 1.5 T group, and 0.738 (0.641–0.805)for the 3 T group. Our findings in volBrain are in accordance with a previous study that found the highest accuracy in volBrain as well when using other MRI sequences (Næss-Schmidt et al. 2016). Other studies using T1-WI, found higher DICE (0.887) using FSL (Patenaude et al. 2011). This difference may be related to the importance of coherent labelling protocols and similar imaging parameters within the template library (Næss-Schmidt et al. 2016).

When comparing the DICE values obtained from the 1.5 T to those from the 3 T group, no significant differences were observed except with FIRST ($p = 0.038$). These differences could be due to the MRI field strength itself. This transition from 1.5 T to 3 T is accompanied by technical differences in signal-to-noise and contrast-to-noise. Several studies have already shown that such field strength differences also appear to influence volume measurements (Jovicich et al. 2009; Tardif et al. 2010). The effect of field strength differences is even added to the already known variability induced by the acquisition parameters and processing pipelines that confound the measurement of brain volumes (Bakshi et al. 2005). Indeed, even though FSL-Anat uses FIRST to segment the subcortical GM structures, it did not show any significant differences. This is probably due to the additional pre-processing steps that FSL-Anat performs (such as reorientation of the images to the standard (MNI) space, bias-field correction, …) before applying FIRST, that could impact the quality of the segmentation.

Since our population was of a brain development study, a further variable to consider in using automated segmentation techniques is age. The relationship with age between the different methods remained relatively stable throughout, with no major shifts. Indeed, none of the segmentations were influenced by age. Although age plays a major role in volumetric assessment in brain development studies, it does not have any significance in our study. Indeed, since the measure in question is the DICE which indicates the degree of similitude between two segmentation methods, no matter the volume of the segmented structure, the outcome will be the same.

Besides the imaging modality, the performance of the software can be associated with the image quality, the segmentation parameters, and the algorithm. It may therefore be useful to address the above-mentioned variables separately to test their effect on the software's outcome. A possibility to take this study further would be to use data that has been optimized for use with FSL and FreeSurfer. Studying differences in optimized data can also provide us with better imaging guidelines. It could also be more interesting to establish an image-processing pipeline to meet a practical clinic requirement. Testing other techniques that were not accessible to our lab would also be good to examine other methods available. Additionally, while most studies usually use T1-WI for structural analysis, it is often the case in retrospective studies that non-enhanced images are not available, as is the case with the 30% of identified subjects that we excluded from our study. Therefore, there is a need to assess the extent of the effects of gadolinium-enhancement on automated subcortical GM segmentation tools. This point constitutes one of our study's limitations. Indeed, the lack of pre- and post-contrast images obtained on the same subject are required to test the validity of the segmentations on

gadolinium enhanced images. This is due to the selection criteria from a clinical database. Such studies may not be feasible, especially with pediatric populations as they mean longer scan times which may negatively impact image quality due to increased possibilities of motion artifacts. A future study in our center will target the segmentation differences between T1 enhanced and non-enhanced images using the same subjects. This would give an even better understanding of how enhancement can affect automated segmentation.

Conclusion

Among five automated segmentation techniques, volBrain proved to have the best outcomes in non-enhanced 3DT1-WI. T1 non-enhanced image segmentation using volBrain would appear to be the ideal methodology for segmentation of the thalamus.

Information Sharing Statement

All software distributions utilized in this work can be accessed by the general public:

volBrain (http://volbrain.upv.es); MRICloud (https://mricloud.org); FMRIB Software Library (FSL5.0) (RRID: SCR_002823, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL); FreeSurfer image analysis suite (v5.3) (RRID:SCR_001847, http://freesurfer.net).

References

Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., & Rueckert, D. (2009). Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage, 46*(3), 726–738. https://doi.org/10.1016/j.neuroimage.2009.02.018.

Aubert-Broche, B., Fonov, V., Ghassemi, R., Narayanan, S., Arnold, D. L., Banwell, B., et al. (2011). Regional brain atrophy in children with multiple sclerosis. *NeuroImage, 58*(2), 409–415. https://doi.org/10.1016/j.neuroimage.2011.03.025.

Azevedo, C. J., Overton, E., Khadka, S., Buckley, J., Liu, S., Sampat, M., et al. (2015). Early CNS neurodegeneration in radiologically isolated syndrome. *Neurology(R) neuroimmunology & neuroinflammation, 2*(3), e102. https://doi.org/10.1212/NXI.0000000000000102.

Bakshi, R., Dandamudi, V. S. R., Neema, M., De, C., & Bermel, R. a. (2005). Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. *Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging, 15*(4 Suppl), 30S–45S. https://doi.org/10.1177/1051228405283901.

Bergsland, N., Horakova, D., Dwyer, M. G., Dolezal, O., Seidl, Z. K., Vaneckova, M., et al. (2012). Subcortical and cortical gray matter atrophy in a large sample of patients with clinically isolated syndrome and early relapsing-remitting multiple sclerosis. *AJNR. American Journal of Neuroradiology, 33*(8), 1573–1578. https://doi.org/10.3174/ajnr.A3086.

Courchesne, E., Chisum, H. J., Townsend, J., Cowles, A., Covington, J., Egaas, B., et al. (2000). Normal brain development and aging: Quantitative analysis at in vivo MR imaging in healthy volunteers. *Radiology, 216*(3), 672–682. https://doi.org/10.1148/radiology.216.3.r00au37672.

Csernansky, J. G., Schindler, M. K., Splinter, N. R., Wang, L., Gado, M., Selemon, L. D., et al. (2004). Abnormalities of thalamic volume and shape in schizophrenia. *American Journal of Psychiatry, 161*(5), 896–902. https://doi.org/10.1176/appi.ajp.161.5.896.

De Jong, L. W., Van Der Hiele, K., Veer, I. M., Houwing, J. J., Westendorp, R. G. J., Bollen, E. L. E. M., et al. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: An MRI study. *Brain, 131*(12), 3277–3285. https://doi.org/10.1093/brain/awn278.

Duan, Y., Li, X., & Xi, Y. (2007). Thalamus segmentation from diffusion tensor magnetic resonance imaging. *International Journal of Biomedical Imaging, 2007*, 90216. https://doi.org/10.1155/2007/90216.

Fearing, M. A., Bigler, E. D., Wilde, E. A., Johnson, J. L., Hunter, J. V., Xiaoqi, L., et al. (2008). Morphometric MRI findings in the thalamus and brainstem in children after moderate to severe traumatic brain injury. *Journal of Child Neurology, 23*(7), 729–737. https://doi.org/10.1177/0883073808314159.

Felten, D. L., Shetty, A. N., & Felten, D. L. (2010). Netter's atlas of neuroscience. Saunders/Elsevier.

Ganzola, R., Maziade, M., & Duchesne, S. (2014, June). Hippocampus and amygdala volumes in children and young adults at high-risk of schizophrenia: Research synthesis. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2014.03.030.

Hannoun, S., Baalbaki, M., Haddad, R., Saaybi, S., El Ayoubi, N. K., Yamout, B. I., et al. (2018). Gadolinium effect on thalamus and whole brain tissue segmentation. *Neuroradiology*. https://doi.org/10.1007/s00234-018-2082-5.

Jatzko, A., Rothenhöfer, S., Schmitt, A., Gaser, C., Demirakca, T., Weber-Fahr, W., et al. (2006). Hippocampal volume in chronic posttraumatic stress disorder (PTSD): MRI study using two different evaluation methods. *Journal of Affective Disorders, 94*(1–3), 121–126. https://doi.org/10.1016/j.jad.2006.03.010.

Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., et al. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage, 46*(1), 177–192. https://doi.org/10.1016/j.neuroimage.2009.02.010.

Koolschijn, P. C. M. P., Van Haren, N. E. M., Lensvelt-Mulders, G. J. L. M., Hulshoff Pol, H. E., & Kahn, R. S. (2009). Brain volume abnormalities in major depressive disorder: A meta-analysis of magnetic resonance imaging studies. *Human Brain Mapping, 30*(11), 3719–3735. https://doi.org/10.1002/hbm.20801.

Lee, S. H., Kim, S. S., Tae, W. S., Lee, S. Y., Choi, J. W., Koh, S. B., & Kwon, D. Y. (2011). Regional volume analysis of the Parkinson disease brain in early disease stage: Gray matter, white matter, striatum, and thalamus. *American Journal of Neuroradiology, 32*(4), 682–687. https://doi.org/10.3174/ajnr.A2372.

Liang, Z. P., & Paul C. Lauterbur. (2000). Principles of magnetic resonance imaging: A signal processingperspective., Wiley-IEEE Press. https://doi.org/10.1109/978047054565.

Mulder, E. R., de Jong, R. A., Knol, D. L., van Schijndel, R. A., Cover, K. S., Visser, P. J., et al. (2014). Hippocampal volume change measurement: Quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST.

NeuroImage, 92, 169–181. https://doi.org/10.1016/j.neuroimage.2014.01.058.

Murgasova, M., Dyet, L., Edwards, D., Rutherford, M., Hajnal, J., & Rueckert, D. (2007). Segmentation of brain MRI in young children. *Academic Radiology, 14*(11), 1350–1366. https://doi.org/10.1016/j.acra.2007.07.020.

Næss-Schmidt, E., Tietze, A., Blicher, J. U., Petersen, M., Mikkelsen, I. K., Coupé, P., et al. (2016). Automatic thalamus and hippocampus segmentation from MP2RAGE: Comparison of publicly available methods and implications for DTI quantification. *International Journal of Computer Assisted Radiology and Surgery, 11*(11), 1979–1991. https://doi.org/10.1007/s11548-016-1433-0.

Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage, 56*(3), 907–922. https://doi.org/10.1016/j.neuroimage.2011.02.046.

Radenbach, K., Flaig, V., Schneider-Axmann, T., Usher, J., Reith, W., Falkai, P., et al. (2010). Thalamic volumes in patients with bipolar disorder. *European Archives of Psychiatry and Clinical Neuroscience, 260*(8), 601–607. https://doi.org/10.1007/s00406-010-0100-7.

Ricci, D., Anker, S., Cowan, F., Pane, M., Gallini, F., Luciano, R., et al. (2006). Thalamic atrophy in infants with PVL and cerebral visual impairment. *Early Human Development, 82*(9), 591–595. https://doi.org/10.1016/j.earlhumdev.2005.12.007.

Rosenberg, D. R., Benazon, N. R., Gilbert, A., Sullivan, A., & Moore, G. J. (2000). Thalamic volume in pediatric obsessive-compulsive disorder patients before and after cognitive behavioral therapy. *Biological psychiatry, 48*(4), 294–300. https://doi.org/10.1016/S0006-3223(00)00902-1.

Rotge, J.-Y., Guehl, D., Dilharreguy, B., Tignol, J., Bioulac, B., Allard, M., et al. (2009). Meta-analysis of brain volume changes in obsessive-compulsive disorder. *Biological Psychiatry, 65*(1), 75–83. https://doi.org/10.1016/j.biopsych.2008.06.019.

Sassi, R. B., Nicoletti, M., Brambilla, P., Mallinger, A. G., Frank, E., Kupfer, D. J., et al. (2002). Increased gray matter volume in lithium-treated bipolar disorder patients. *Neuroscience Letters, 329*(2), 243–245. https://doi.org/10.1016/S0304-3940(02)00615-8.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23*(SUPPL. 1), S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051.

Solomon, A. J., Watts, R., Dewey, B. E., & Reich, D. S. (2017). MRI evaluation of thalamic volume differentiates MS from common mimics. *Neurology(R) neuroimmunology & neuroinflammation, 4*(5), e387. https://doi.org/10.1212/NXI.0000000000000387.

Tardif, C. L., Collins, D. L., & Pike, G. B. (2010). Regional impact of field strength on voxel-based morphometry results. *Human Brain Mapping, 31*(7), 943–957. https://doi.org/10.1002/hbm.20908.

Tsatsanis, K. D., Rourke, B. P., Klin, A., Volkmar, F. R., Cicchetti, D., & Schultz, R. T. (2003). Reduced thalamic volume in high-functioning individuals with autism. *Biological Psychiatry, 53*(2), 121–129. https://doi.org/10.1016/S0006-3223(02)01530-5.

Weisenfeld, N. I., & Warfield, S. K. (2009). Automatic segmentation of newborn brain MRI. *NeuroImage, 47*(2), 564–572. https://doi.org/10.1016/j.neuroimage.2009.04.068.