Neuroinformatics
https://doi.org/10.1007/s12021-021-09520-z

ORIGINAL ARTICLE

A Comparative Analysis of MRI Automated Segmentation of Subcortical Brain Volumes in a Large Dataset of Elderly Subjects

Jaime Gomez Ramirez[1] Javier Quilis Sancho[1] Miguel A. Fernandez Blazquez[1]

Accepted: 11 March 2021

bstract

In this study, we perform a comparative analysis of automated image segmentation of subcortical structures in the elderly brain. Manual segmentation is very time-consuming and automated methods are gaining importance as a clinical tool for diagnosis. The two most commonly used software libraries for brain segmentation -FreeSurfer and FSL- are put to work in a large dataset of 4,028 magnetic resonance imaging (MRI) scans collected for this study. We find a lack of linear correlation between the segmentation volume estimates obtained from FreeSurfer and FSL. On the other hand, FreeSurfer volume estimates tend to be larger thanFSL estimates of the areas putamen, thalamus, amygdala, caudate, pallidum, hippocampus, and accumbens. The characterization of the performance of brain segmentation algorithms in large datasets as the one presented here is a necessary step towards partially or fully automated end-to-end neuroimaging workflow both in clinical and research settings.

Keywords Brain imaging · magnetic resonance imaging · Segmentation · FreeSurfer · FSL

Introduction

Healthy aging and neurodegenerative and psychiatric disorders show changes in the structure of the brain (Gado et al. 1983; Snowdon 2003; Carlson et al. 2008; Fjell et al. 2009). These changes may include differences observed in either the morphology of subcortical structures or in alterations of the thickness and morphology of the cortex (Yang et al. 2019; Azevedo et al. 2019; Vollmer et al. 2015).

Brain atrophy, either the organ's volume or specific subfields, can work as a biomarker to reveal the existence of neurological disorders, such as Alzheimer's disease (Pini et al. 2016), Parkinson's disease (Zhou et al. 2020), and multiple sclerosis (Losseff et al. 1996; Fox et al. 2000; Andravizou et al. 2019) among others. Two well-established markers of dementia are loss of gray matter volume (Beyer et al. 2007; Peter et al. 2014) and the shrinking of the hippocampus (Wang et al. 2003; de Flores et al. 2015). Being able to detect significant volumetric changes before behavioral deterioration begins is crucial for the early diagnosis of the disease.

In the era of big data, researchers and clinicians have at their disposal a large number of datasets that are being fed at an ever-increasing frequency (Bug 2005; Yan et al. 2020; Esteva et al. 2021). Neuroscience deserves a particular mention as one of the most data-rich scientific endeavors. Repositories with thousands of MRI scans freely accessible are already available (Jack Jr et al. 2008; Miller et al. 2016; Mazurowski et al. 2019), with larger, and more complex repositories coming up in the near future. Automated procedures will be an essential element in the toolkit of radiologists and other health professionals working on medical imaging (Hosny et al. 2018; Thrall et al. 2018). The reproducibility and time-saving characteristics of automated segmentation are some of the strongest points of AI in the field of medical imaging. However, the field is still in its adolescence and, AI cannot, so far, replace an expert eye in every domain (Topol 2019).

Nevertheless, the unique capabilities of AI can provide a formidable companion to the expert physician, in particular radiologists, dermatologists, and pathologists, all three information-based medicinal practices.

The main premise of big data applied to brain imaging in aging research is that given enough data, we may be able to fit the data to a model that makes predictions about the

Jaime Gomez-Ramirez
jd.gomezramirez@gmail.com

[1] Instituto de Salud Carlos III, Centro de Alzheimer Fundación Reina Sof´a, Madrid, Spain

Published online: 30 March 2021

dynamics of the atrophy, that is, characterize atrophy and its progression in an individual basis. Manual segmentation of the brain is a time-consuming and prone to error task (Collier et al. 2003; Despotović et al. 2015). To put this in perspective, only to extract the hippocampus volume, an expert requires at least 30-40 minutes (Firbank et al. 2008). Therefore, an exhaustive segmentation of the whole brain could take hours or days of human work (Firbank et al. 2008; Starmans et al. 2020). Manual segmentation of hundreds or more brains is not a viable option. For this reason, in this work, we explore automatic procedures of brain segmentation using a large longitudinal dataset, contrasting the results of the volume estimates obtained by the two most commonly used automatic segmentation tools.

The main advantages of automatic procedures are at least two: i) the lack of bias inherent in manual segmentation, two different human experts may produce very different estimates for the same image and ii) automatic procedures are time-saving. The goal of this work is thus, to present the current capabilities and promise of automatic procedures for brain segmentation methods using automatic brain segmentation methods, systematically and consistently in a large longitudinal study of 4028 MRI scans collected in a 5-year period.

Methodology

Participants

The dataset for this work was collected in a multi-year research study of healthy aging based in Madrid, Spain Gómez-Ram´rez et al. (2020), Fernández-Blázquez et al. (2020). The dataset is part of a single-center, observational, longitudinal cohort study. The participants were recruited between 2011 and 2013, the cohort started with 990 home-dwelling volunteers, between 70 and 85 years old, without any relevant psychiatric, neurological, or systemic disorder. Since their inclusion in the study, the volunteers undergo a yearly systematic clinical evaluation which includes medical history, neurological and neuropsychological exam, blood collection, and brain MRI. Each subject is examined yearly and information regarding a wide range of factors such as MRI, genetic, demographic, socioeconomic, cognitive performance, subjective cognitive decline, neuropsychiatric disorders, cardiovascular, sleep, diet, physical exercise, and self-assessed quality of life is collected. The subjects are clinically diagnosed as healthy, with mild cognitive impairment or dementia.

The 990 subjects in the first year of the study have the following characteristics: Age 74.72 $\pm$ 3.86; 40% Male; 16% heterozygous for the 4 genotype of APOE and 0.7% homozygous for the 4 genotype of APOE; 17% unschooled, 26% with primary school, 24% with medium or high school and 23% with university studies. When a subject is diagnosed with dementia, he or she abandons the study, the subjects are free to leave the study by personal choice at any time. Because of that, the number of subjects has been reduced over the years.

Imaging Study

A total of 4028 MRIs were collected in 5 years, 990 in the first visit, 768 in the second, 723 in the third, 634 in the fourth, 542 in the fifth, and 371 in the sixth year. The imaging data were acquired on a 3T General Electric scanner (GE Milwaukee) utilizing the following parameters, T1-weighted inversion recovery, flip angle 12 , 3-D pulse sequence: echo time *Min. full*, time inversion 600 ms, receiver bandwidth 19.23 kHz, field of view = 24.0 cm, slice thickness 1 mm, Freq. $\times$ Phase 288 $\times$ 288. The preprocessing of MRI 3 Tesla images in this study consisted of generating an isotropic brain image with non-brain tissue removed. We used the initial, preprocessing step in the two computational segmentation tools used in this study: FSL pipeline (fsl-anat (FSL 2017)) and the FreeSurfer pipeline (recon-all FreeSurfer 2017).

FreeSurfer and FSL Pipelines

The two software libraries for brain segmentation used in this work are FSL (Woolrich et al. 2009; Smith et al. 2004; Jenkinson et al. 2012) and FreeSurfer (Dale and Sereno 1993; Dale et al. 1999). Here we will focus on the analysis of structural MRI by the two platforms showing their main similarities and differences.

FSL allows the processing of structural MRI, functional MRI, diffusion MRI and perfusion MRI. As said, we are focusing on structural MRI. The main tools for FSL-based analysis are the Brain Extraction Tool (BET), the FMRIB's Automated Segmentation Tool (FAST), and the FMRIB's Integrated Registration and Segmentation Tool (FIRST). The BET routine separates brain tissue from non-brain tissue (Smith 2002; Simpson et al. 2012). BET estimates the total brain volume after bone, fat, and muscle has been removed in the image. The FAST routine (Zhang et al. 2001) routine can classify the voxels of the brain into white matter, gray matter, and CSF. And FIRST (Patenaude et al. 2011) uses Bayesian modeling to segment the different subcortical structures of the brain. The model training set includes both normal and pathological brains (including cases of schizophrenia and Alzheimer's disease) and the age

range is between 4 and 87 (Patenaude et al. 2011). FIRST segments the following brain structures: putamen, thalamus, amygdala, caudate, pallidum, hippocampus, accumbens, and brainstem. FSL also includes a tool called Structural Image Evaluation (SIENA/X) (Smith et al. 2002) that using Normalisation of Atrophy allows the analysis of brain change in different time points. SIENA/X estimates the volumetric loss of brain tissue between two images of the same subject at different times.

FreeSurfer includes tools for processing structural MRI, functional MRI, diffusion MRI and PET data. Again, we are focusing on structural MRI. The cross-sectional analysis starts with the surface-based stream (Dale et al. 1999; Fischl et al. 1999), where the skull is stripped, the cerebellum and brain stern are removed, the two hemispheres are separated and brain voxels are classified as white matter, grey matter or CSF (Fischl et al. 2002; Fischl et al. 2004). Then, the volume-based stream segments the different subcortical structures of the brain using a subject-independent probabilistic atlas.

The FreeSurfer training set consists of 40 MRIs, spread in age (10 healthy young subjects, 10 healthy middle-aged subjects, and 10 healthy elderly subjects) and including pathological brains (10 subjects with AD) (Desikan et al. 2006). The pipeline segments the cortex surface (Fischl et al. 2004). It uses three different atlases to give tree different cortical parcellations: the Desikan-Killiany atlas (Desikan et al. 2006), the DKT atlas (Klein and Tourville 2012) and the Destrieux atlas (Fischl et al. 2004; Destrieux et al. 2010). This pipeline provides the total brain volume, white matter, gray matter, CSF, segmentation of subcortical structures and cortical parcellation which is not provided in FSL. The subcortical segmentation is more extensive in FreeSurfer than in FSL because apart from the structures in FSL, FreeSurfer also segments cerebellum, optic chiasm, ventricles, vessel, and choroid plexus. FreeSurfer also has a longitudinal protocol (Reuter et al. 2010; Reuter and Fischl 2011) which analyzes different MRIs from the same subject across time and reduces the variability between time points. The longitudinal analysis of FreeSurfer provides not only brain volume changes as in FSL SIENA/X, but also cortical and subcortical volumetric changes in time (Reuter et al. 2010; Reuter and Fischl 2011).

The stages in the FSL pipeline are (in order): reorient the images to the standard (MNI) orientation, automatically crop the image, bias-field correction (RF/B1-inhomogeneity-correction), registration to standard space (linear and non-linear), brain-extraction, tissue-type segmentation, and subcortical structure segmentation.

The stages in the FreeSurfer pipeline are (in order): surface-based stream, with skull-stripping cerebellum and brain stern removal, two hemispheres separation and brain voxels classification (white matter, gray matter, and CSF), and finally brain segmentation (cortical, and subcortical).

We run both pipelines in an identical computational setting: Operating System Mac OS X, product version 10.14.5 and build version 18F132. The version of FreeSurfer is FreeSurfer-darwin-OSX-ElCapitan-dev-20190328-6241d26. The version of the BET tool for FSL is v2.1 - FMRIB Analysis Group, Oxford, and the FIRST tool version is 6.0.

Figure 1 shows the automatic subcortical segmentation obtained for one subject in our study using FSL and FreeSurfer.

Results

We evaluate how FSL and FreeSurfer perform on a comparative basis using a dataset of 4028 T1-weighted MRIs. As mentioned before, our final goal is to evaluate the performance of fully automated methods in a large, mono-center dataset. For the sake of comparability between the two tools, we focus on the brain structures that both FSL and FreeSurfer segment automatically, i.e. the total brain volume, putamen, thalamus, amygdala, caudate, pallidum, hippocampus, and accumbens. In this analysis we only apply the cross-sectional protocol to all the images, that is, we treat every MRI as independent from the rest. Therefore, MRIs taken at different times from the same subject are treated independently.

First, we look at the estimated brain volumes and brain structures obtained in the first year of the study from every subject in the dataset with 990 subjects in total each of them with automatic segmentation performed using both FSL and FreeSurfer. Motion and other artifacts can result in poor

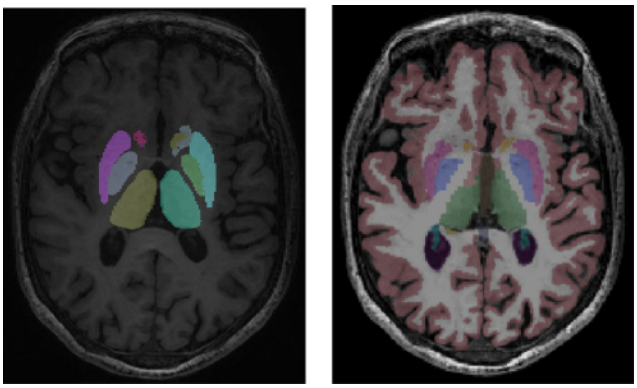

Fig 1 Axial view of automated subcortical segmentation of one participant in the study (71 year old, Female) using FSL( ) and FreeSrrfer(**b**). The figure shows the Putamen, Caudate, Pallidum, and Thalamus

estimates (Kecskemeti and Alexander 2020), we excluded 1% of the outliers in each structure, this means that we exclude the 0.5% larger and smaller interval of values.

Figure 2 depicts histograms of the volumetry estimated by both libraries for the total brain volume and the different subcortical structures. As expected, the distributions for each structure show a Gaussian shape for both FSL and FreeSurfer. However, it is worth mentioning that FreeSurfer distributions (in blue) have higher values in every structure compared to FSL distributions (in green). This fact is quantified in Table 1 including the mean and standard deviation for every structure and tool.

Visual inspection of the distributions in Fig. 2 suggests evidence against the null hypothesis -no significant difference between FreeSurfer and FSL samples. The statistical test shows a p-value lower than 0.01 in every structure, confirming the visual appreciation. Finally, we calculated the effect size with the Cohen's *d* (Cohen 1988) to study the standardized difference between the means, observing that except for the brain volume, the effect size is very large for every structure, shown in Table 1.

Interannual Volumetric Estimates with FreeSurfer and FSL

The differences in volume estimates obtained by both software packages might call into question the consistency of automated MRI segmentation methods. However, we need to take into account the experimental design and the characteristics of the collected dataset. As said, the study is longitudinal whereas the analysis applied is cross-sectional. Thus, the MRI scans for each subject were acquired with a time-lapse of about 1 year ±2 months between images but were analyzed as if they were independent scans. However, in normal conditions, one year (±2 months) is not time

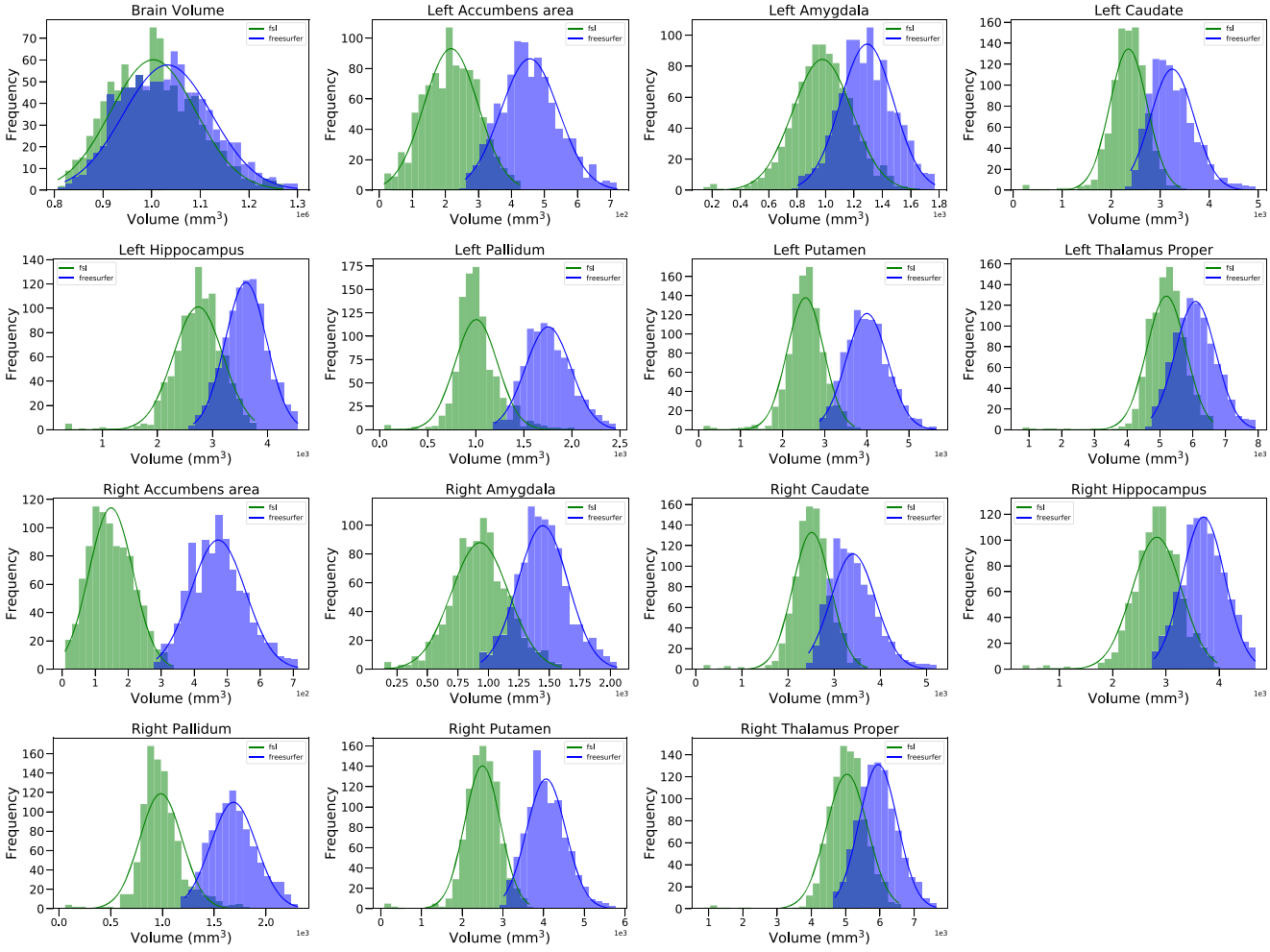

Fig 2 Distributions of the volume for every structure by the two automated methods. The distributions include every resonance obtained in the first visit of every subject (a total of 990). We compare the distributions provided by FSL (in green) and the ones obtained with FreeSurfer (in blue). The lines represent the theoretical Gaussian distribution calculated with the mean and the standard deviation of each distribution

Table 1 Means and standard deviations of the distributions of volumetric estimates computed with FSL and FreeSurfer shown in Fig. 2

Structure	μ_{fsl} (mm^3)	std$_{fsl}$	μ_{fr} (mm^3)	std$_{fr}$	p-val	Cohen d (95% CI)
Brain Volume	1003131.145	87388.925	1032458.59	91017.684	**	0.33 (0.24-0.42)
L Accumbens	216.333	82.316	454.939	87.938	**	2.80 (2.68-2.93)
L Amygdala	977.378	208.738	1292.643	186.856	**	1.59 (1.49-1.69)
L Caudate	2351.182	380.826	3244.433	444.103	**	2.16 (2.05-2.27)
L Hippo.	2739.324	451.388	3608.068	376.729	**	2.09 (1.98-2.20)
L Pallidum	1005.491	220.527	1753.616	236.138	**	3.27 (3.14-3.41)
L Putamen	2535.503	452.667	4001.368	490.511	**	3.11 (2.98-3.24)
L Thalamus	5194.977	594.592	6084.683	619.566	**	1.47 (1.37-1.56)
R Accumbens	147.22	67.014	472.016	83.022	**	4.31 (4.14-4.47)
R Amygdala	930.895	234.438	1446.186	206.795	**	2.33 (2.22-2.45)
R Caudate	2513.751	409.376	3410.172	485.98	**	2.00 (1.89-2.10)
R Hippo.	2829.544	461.866	3708.178	400.644	**	2.03 (1.92-2.14)
R Pallidum	983.657	206.276	1687.419	221.328	**	3.29 (3.15-3.42)
R Putamen	2496.482	443.494	4062.842	480.476	**	3.39 (3.25-3.53)
R Thalamus	5037.812	599.508	5952.144	559.42	**	1.58 (1.48-1.68)

The last two columns show the p-value and the Cohen's *d* from the comparison of the FSL and FreeSurfer distributions for each structure. Overall, FreeSurfer tends to estimate larger volume values than FSL

long enough to expect significant structural changes in the brain. Several studies (Fox and Freeborough 1997; O'Brien et al. 2001; Enzinger et al. 2005) show that the mean change in volume that a healthy adult brain experiences in a year is lower than 0.5%. Other studies of healthy aging (Fjell et al. 2009; Fjell et al. 2013) using similar methodology but focusing on the subcortical structures, showed 1% or less of annual shrinking for subcortical structures. Hence, a sensible way to test the self-consistency of both FSL and FreeSurfer is to measure the volumetric difference from the same subject between two consecutive years. We would expect, according to the literature, small changes ($< 1\%$) between the estimated volume for two-time points spanned one year time.

Following this argument, we calculated the variation of the volume of the cerebrum and the different subcortical structures, for each subject and in between every two consecutive years. This means that for a subject with n consecutive MRIs, we calculate

$$\Delta^{struc}_{y_i, y_{i-1}} = (Vol^{struc}_{y_i} - Vol^{struc}_{y_{i-1}})/Vol^{struc}_{y_{i-1}}, \quad (1)$$

where $Vol^{struc}_{y_i}$ is the volume of a structure, $struc$, between years i and i-1.

The study of the interannual volumetric variation estimated by FreeSurfer is shown in Fig. 3. The interannual variation of volumes is normalized, ie. defined within the range from 0 (total preservation in-between years) to 1 (total loss in-between years). Focusing first on the brain volume, we observe that the distribution is centered around 0 but skewed towards negative values. This is as expected since aging will produce a progressive shrinkage of the brain. Next, regarding the subcortical volume estimates in Fig. 3, we see a similar distribution shape for all the structures. However, the variability is more pronounced in subcortical structures than it is in the total brain volume, one order of magnitude larger in some cases. For example, the accumbens both left and right show the largest interannual variation (the accumbens is also the smallest subcortical structure studied here). The interannual volumetric variation is within 25%.

We proceed similarly for FSL and the results are shown in Fig. 4. We find strong similarities with the results obtained from FreeSurfer (Fig. 3), although FSL produced more outliers, five times more cases than FreeSurfer. Both FSL and FreeSurfer interannual volume variation distributions are centered around zero and skewed towards the left or negative values. The variability for the caudate, hippocampus, pallidum, putamen and thalamus, is again constrained below 25%. As it occurred with FreeSurfer segmentation, the accumbens presents larger variability than the rest of the structures, which as already noted it might be related to the fact that the nucleus accumbens is the smallest subcortical structure of the 7 structures segmented in the study. The amygdala shows significant interannual variability. Similarly, as it was observed in FreeSurfeer (Fig. 3), the total brain volume interannual change distribution (Top left Fig. 4) is below 5%, which again is what we would expect.

In Table 2, we assess the differences in the successive volume estimates from the two segmentation tools shown

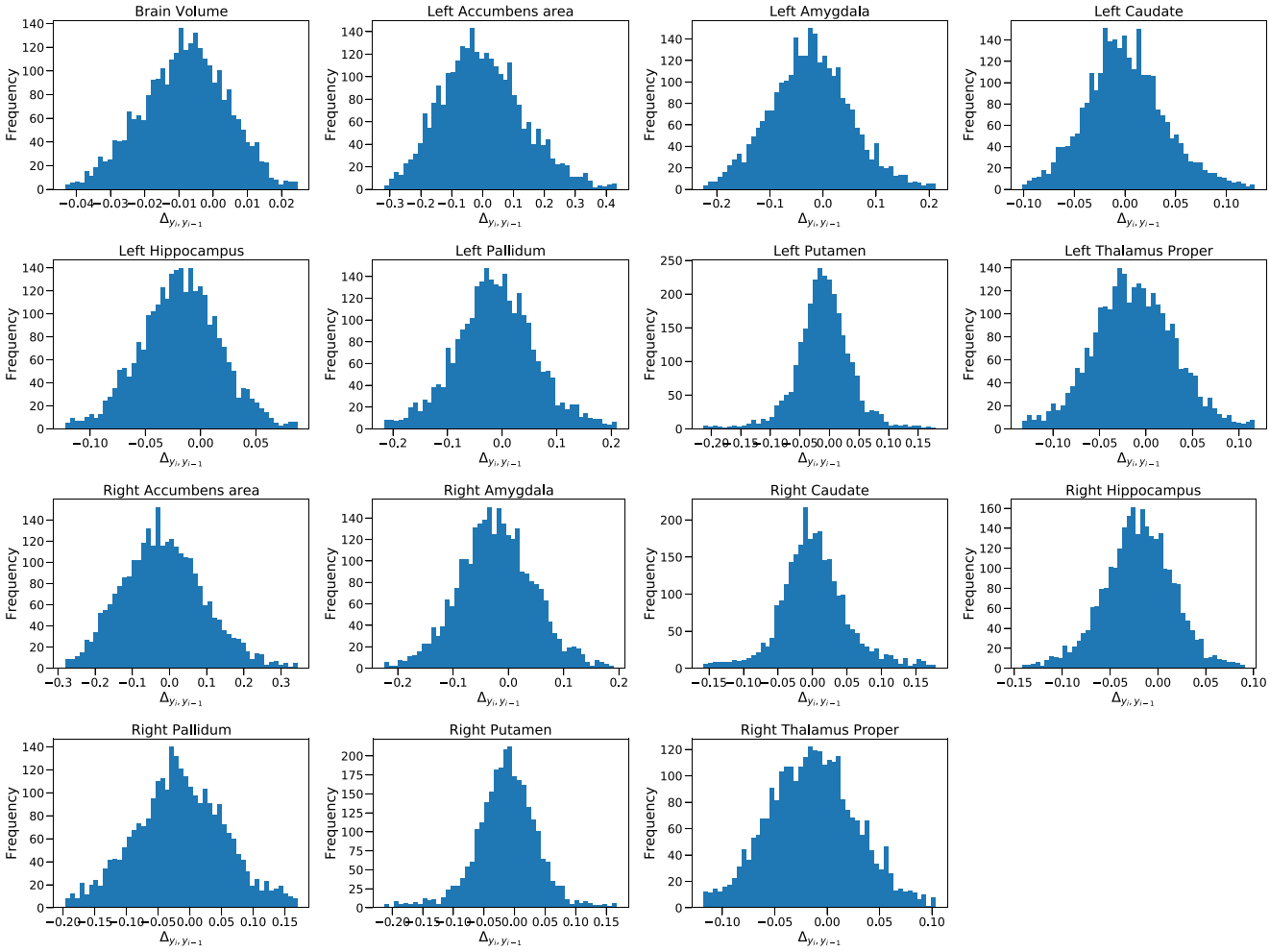

Fig 3 Variations in the volume of the different brain structures between two consecutive years (Equation 1) measured by FreeSurfer. The interannual volume variation is constrained around [-0.045, 0.025] (min. 0 and max. 1)

in Figs. 3, 4. The mean brain shrinking between two years is 0.007 for FSL and 0.008 for FreeSurfer, or less than one percentage point, which is similar to the results obtained in previous studies (Fox and Freeborough 1997; O'Brien et al. 2001; Enzinger et al. 2005). Concerning the subcortical structures, all the means are lower than 0.01 except for the right amygdala in both libraries and the right hippocampus in the case of FreeSurfer. In the case of FreeSurfer segmentation, only the mean of the caudate (left and right) is larger than zero, while the mean volume of the rest of the structures decreases in a year. This result is very similar to the results in References (Fjell et al. 2009; Fjell et al. 2013), commented before, where the mean shrinking of every structure is lower than 1%. Summarizing, we can conclude that both FreeSurfer and FSL are self-consistent in the sense that both present negligible inter-annual variability for the same subject's brain and subcortical volume estimates.

Discussion and Conclusions

We have studied the segmentation results of the two most commonly brain segmentation libraries, FSL and FreeSurfer. In Fig. 2, we checked that the volume distributions by each library follow a Gaussian distribution. But more important, in Figs. 3, 4 we showed that the changes in the brain volume of the different subjects within a year tend to be small, even when the images are treated as independent from each other, that is to say not using a longitudinal protocol for segmentation of two successive MRI scans for the same subject. These results are consistent with References Fjell et al. (2009, 2013), where segmentation was performed manually.

Other comparative analyses of FSL and FreeSurfer volumetric estimates of subcortical structures show similarities but also inconsistencies with our findings. For example, Rane et al. (2017) in a cohort of elderly subjects found

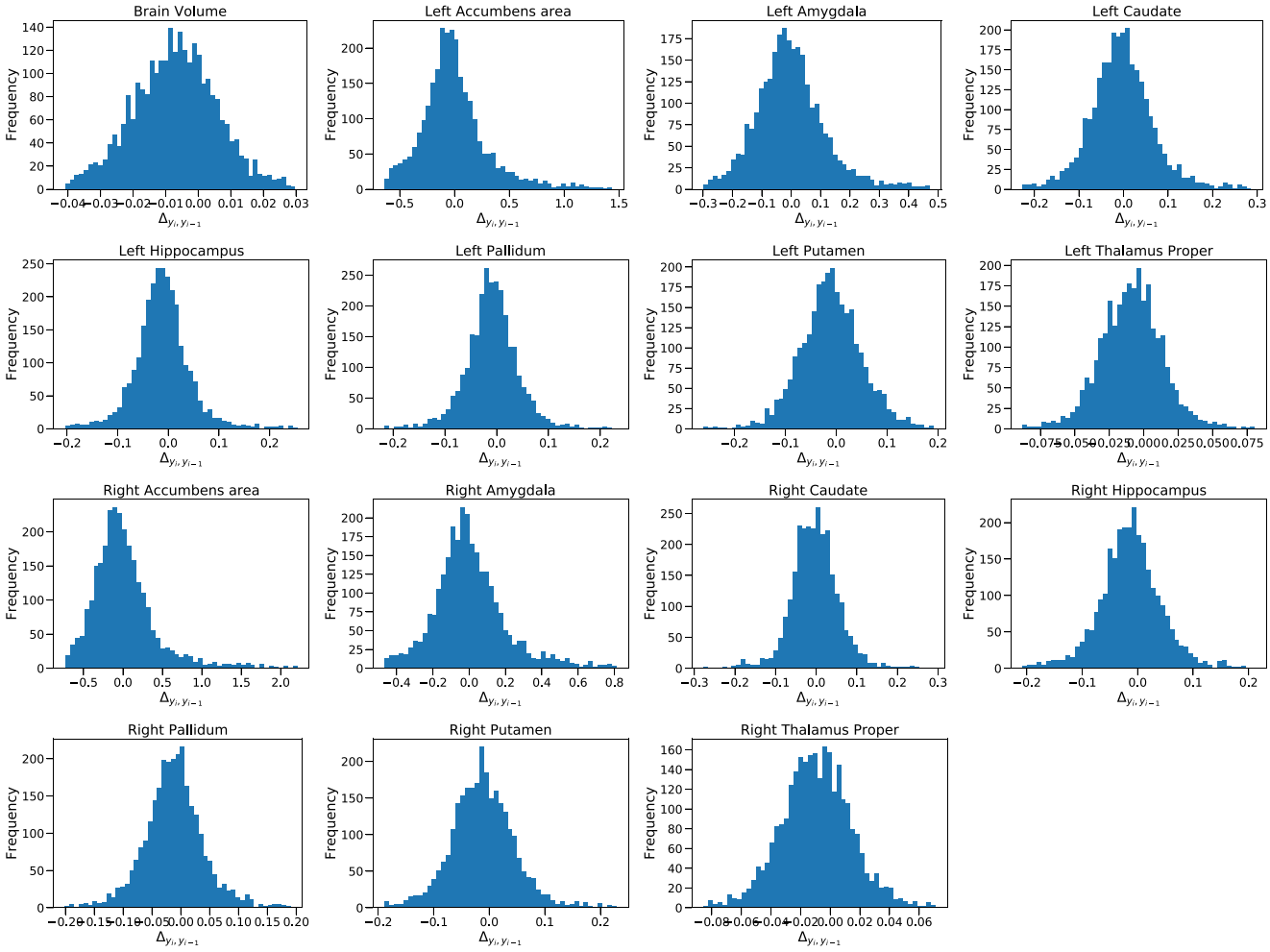

Fig 4 Variations in the volume of the different brain structures between two consecutive years (Equation 1) measured by FSL. The distributions of the interannual variation in FSL, present as it is the case of FreeSurfer, a Gaussian shape

larger volume estimates for the caudate, putamen, hippocampus, amygdala, and accumbens from FreeSurfer in comparison with FSL, which is in line with our results. However, the estimations of the pallidum and thalamus were larger for FSL, which is not in agreement with our results. In Seixas et al. (2010), the authors studied the volume of the hippocampus of healthy adults, finding that the volume estimates from FSL were larger than the ones obtained from FreeSurfer, which also deviates from our results. The hippocampus and amygdala have been studied also in (Morey et al. 2009), and the results are compatible with ours, FreeSurfer estimates larger values for the amygdala and the hippocampus. In Makowski et al. (2018) the automated segmentation in both the amygdala and the hippocampus, were larger in FreeSurfer than in FSL. And to conclude, in Morey et al. (2010), the authors found larger volumes measured by FreeSurfer in the amygdala with FSL volume estimates larger for the accumbens, caudate, hippocampus, pallidum, putamen, and thalamus.

The rationale behind the origin of the inconsistencies between studies may respond to different factors. First, most of the studies mentioned are modest in terms of dataset dimensionality, typically less than a hundred subjects. Differences found in studies with few MRIs could be a statistical effect and therefore we should be cautious in inferring conclusions. Last, differences in the operating system and software library versions have been related to inconsistencies in the results across studies (Gronenschild et al. 2012). Our study, on the other hand, does not

Table 2 Means and standard deviations of the relative changes for every structure within a year (ie. successive MRIs acquired in the span of 1 year ±2 months)

Structure	μ_{fsl}	std_{fsl}	μ_{fr}	std_{fr}
Brain Volume	−0.007	0.015	−0.008	0.012
Left Accumbens area	0.018	0.414	−0.001	0.133
Left Amygdala	0.004	0.169	−0.024	0.076
Left Caudate	0.005	0.187	0.0004	0.039
Left Hippocampus	−0.007	0.122	−0.018	0.036
Left Pallidum	−0.001	0.163	−0.012	0.073
Left Putamen	−0.001	0.218	−0.01	0.047
Left Thalamus Proper	−0.008	0.061	−0.013	0.044
Right Accumbens area	0.069	0.671	−0.016	0.11
Right Amygdala	0.02	0.291	−0.022	0.067
Right Caudate	0.005	0.18	0.003	0.05
Right Hippocampus	−0.01	0.113	−0.021	0.036
Right Pallidum	−0.004	0.17	−0.016	0.07
Right Putamen	0.006	0.315	−0.014	0.05
Right Thalamus Proper	−0.008	0.071	−0.014	0.041

The distributions are shown in Figs. 3, 4

suffer from any of these shortcomings since each of the 4028 MRIs scans analyzed were acquired using identical scanner, computational setting, and personnel which gives a particular quality to our study vis a vis studies with smaller samples or that combine MRIs from different scanners.

Lastly, we argue that the main reason for the inconsistencies between the estimates of FSL and FreeSurfer might be found in differences in the segmentation protocol. Different boundaries for the same brain structure would translate into different volume estimates. An additional factor that deserves attention is the dataset used to train the segmentation algorithm. Our dataset is entirely from elderly people and depending on how well-represented is this sector of the population in the library's brain template, the results could show inconsistencies between the output of the FSL and FreeSurfer algorithms.

Information Sharing Statement

Code and all data used in this research are publicly available on the Github repository under an Apache 2.0 license at https://github.com/grjd/automaticsegmentation. Part of the pre-processing code depends on FSL and FreeSurfer. Both software libraries are only licensed for non-commercial use and are freely available.

Acknowledgements The authors would like to thank the generous persons that volunteered to participate in the study and *Fundación Reina Sof´a* for their support. The authors acknowledge funding from *Ministerio de Ciencia, Innovación y Universidades* (CONNECT-AD) RTI2018-098762-B-C31 and and Structural Funds ERDF (INTERREG V-A Spain-Portugal (POCTEP) Grant: 0348CIE6E).

eclarations

Competing interests The authors declare no competing interests.

References

Andravizou, A., Dardiotis, E., Artemiadis, A., Sokratous, M., Siokas, V., Tsouris, Z., Aloizou, A.-M., Nikolaidis, I., Bakirtzis, C., Tsivgoulis, G., et al. (2019). Brain atrophy in multiple sclerosis: mechanisms, clinical relevance and treatment options. *Autoimmunity Highlights*, *10*(1), 7.

Azevedo, C.J., Cen, S.Y., Jaberzadeh, A., Zheng, L., Hauser, S.L., Pelletier, D. (2019). Contribution of normal aging to brain atrophy in ms. *Neurology-Neuroimmunology Neuroinflammation*, *6*(6), e616.

Beyer, M.K., Larsen, J.P., Aarsland, D. (2007). Gray matter atrophy in parkinson disease with dementia and dementia with lewy bodies. *Neurology*, *69*(8), 747–754.

Bug, W. (2005). The impact of the nih public access policy on literature informatics. *Neuroinformatics*, *3*(2), 81–91.

Carlson, N.E., Moore, M.M., Dame, A., Howieson, D., Silbert, L.C., Quinn, J.F., Kaye, J.A. (2008). Trajectories of brain loss in aging and the development of cognitive impairment. *Neurology*, *70*(11), 828–833.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences, 2nd edn. á/l.

Collier, D.C., Burnett, S.S., Amin, M., Bilton, S., Brooks, C., Ryan, A., Roniger, D., Tran, D., Starkschall, G. (2003). Assessment of consistency in contouring of normal-tissue anatomic structures. *Journal of Applied Clinical Medical Physics*, *4*(1), 17–24.

Dale, A., Fischl, B., Sereno, M.I. (1999). Cortical surface-based analysis: i. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194.

Dale, A.M., & Sereno, M.I. (1993). Improved localizadon of cortical activity by combining eeg and meg with mri cortical

surface reconstruction: a linear approach. *Journal of Cognitive Neuroscience*, *5*(2), 162–176. PMID 23972151.

de Flores, R., Joie, R.L., Chetelat, G. (2015). Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. *Neuroscience*, *309*, 29–50. Hippocampal vulnerability: from molecules to disease.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980.

Despotović, I., Goossens, B., Philips, W. (2015). Mri segmentation of the human brain: challenges, methods, and applications. Computational and mathematical methods in medicine.

Destrieux, C., Fischl, B., Dale, A., Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15.

Enzinger, C., Fazekas, F., Matthews, P.M., Ropele, S., Schmidt, H., Smith, S., Schmidt, R. (2005). Risk factors for progression of brain atrophy in aging. *Neurology*, *64*(10), 1704–1711.

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., Socher, R. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, *4*(1), 1–9.

Fernández-Blázquez, M.A., Noriega-Ruiz, B., Ávila Villanueva, M., Valent´-Soler, M., Frades-Payo, B., Ser, T.D., Gómez-Ram´rez, J. (2020). Impact of individual and neighborhood dimensions of socioeconomic status on the prevalence of mild cognitive impairment over seven-year follow-up. *Aging & Mental Health*, *0*(0), 1–10. PMID 32067489.

Firbank, M.J., Barber, R., Burton, E.J., O'Brien, J.T. (2008). Validation of a fully automated hippocampal segmentation method on patients with dementia. *Human Brain Mapping*, *29*(12), 1442–1449.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*, 341–355.

Fischl, B., Sereno, M.I., Dale, A. (1999). Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, *14*(1), 11–22.

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B. (2013). Brain changes in older adults at very low risk for Alzheimer's disease. *Journal of Neuroscience*, *33*(19), 8237–8242.

Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M. (2009). One-year brain atrophy evident in healthy aging. *Journal of Neuroscience*, *29*(48), 15223–15231.

Fox, N., Jenkins, R., Leary, S., Stevenson, V., Losseff, N., Crum, W., Harvey, R.J., Rossor, M., Miller, D., Thompson, A. (2000). Progressive cerebral atrophy in ms: a serial study using registered, volumetric mri. *Neurology*, *54*(4), 807–812.

Fox, N.C., & Freeborough, P.A. (1997). Brain atrophy progression measured from registered serial mri: Validation and application to Alzheimer's disease. *Journal of Magnetic Resonance Imaging*, *7*(6), 1069–1075.

FreeSurfer cortical reconstruction and parcellation process (2017). Anatomical processing script: recon-all. https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all, Last accessed on 2020-15-30.

FSL (2017). Anatomical processing script: fsl_anat. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/fsl_anat, Last accessed on 2020-15-30.

Gado, M., Hughes, C.P., Danziger, W., Chi, D. (1983). Aging, dementia, and brain atrophy: a longitudinal computed tomographic study. *American Journal of Neuroradiology*, *4*(3), 699–702.

Gómez-Ram´rez, J., Ávila-Villanueva, M., Fernández-Blázquez, M.Á. (2020). Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods. *Scientific Reports*, *10*(1), 1–15.

Gronenschild, E.H.B.M., Habets, P., Jacobs, H.I.L., Mengelers, R., Rozendaal, N., van Os, J., Marcelis, M. (2012). The effects of freesurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements. *PLOS ONE*, *7*(6), 1–13.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J.W.L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510.

Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., et al. (2008). The Alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *27*(4), 685–691.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M. (2012). Fsl. *NeuroImage*, *62*(2), 782–790. 20 YEARS OF fMRI.

Kecskemeti, S.R., & Alexander, A.L. (2020). Test-retest of automated segmentation with different motion correction strategies: a comparison of prospective versus retrospective methods. *NeuroImage*, *209*, 116494.

Klein, A., & Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, *6*, 171.

Losseff, N., Wang, L., Lai, H., Yoo, D., Gawne-Cain, M., McDonald, W., Miller, D., Thompson, A. (1996). Progressive cerebral atrophy in multiple sclerosis a serial mri study. *Brain: A Journal of Neurology*, *119*(6), 2009–2019.

Makowski, C., Béland, S., Kostopoulos, P., Bhagwat, N., Devenyi, G.A., Malla, A.K., Joober, R., Lepage, M., Chakravarty, M.M. (2018). Evaluating accuracy of striatal, pallidal, and thalamic segmentation methods: Comparing automated approaches to manual delineation. *NeuroImage*, *170*, 182–198. Segmenting the Brain.

Mazurowski, M.A., Buda, M., Saha, A., Bashir, M.R. (2019). Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on mri. *Journal of Magnetic Resonance Imaging*, *49*(4), 939–954.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al. (2016). Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature Neuroscience*, *19*(11), 1523–1536.

Morey, R.A., Petty, C.M., Xu, Y., Hayes], J.P., Wagner, H.R., Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G. (2009). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *NeuroImage*, *45*(3), 855–866.

Morey, R.A., Selgrade, E.S., Wagner II, H.R., Huettel, S.A., Wang, L., McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping*, *31*(11), 1751–1762.

O'Brien, J.T., Paling, S., Barber, R., Williams, E.D., Ballard, C., McKeith, I., Gholkar, A., Crum, W.R., Rossor, M.N., Fox, N.C. (2001). Progressive brain atrophy on serial mri in dementia with

lewy bodies, ad, and vascular dementia. *Neurology*, *56*(10), 1386–1388.

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.

Peter, J., Scheef, L., Abdulkadir, A., Boecker, H., Heneka, M., Wagner, M., Koppara, A., Kloppel, S., Jessen, F. (2014). Gray matter atrophy pattern in elderly with subjective memory impairment. *Alzheimer's and Dementia*, *10*(1), 99–108.

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., Frisoni, G.B. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews*, *30*, 25–48.

Rane, S., Plassard, A., Landman, B.A., Claassen, D.O., Donahue, M.J. (2017). Comparison of cortical and subcortical measurements in normal older adults across databases and software packages. *Journal of Alzheimer's Disease Reports*, *1*, 59–70.

Reuter, M., & Fischl, B. (2011). Avoiding asymmetry-induced bias in longitudinal image processing. *NeuroImage*, *57*(1), 19–21.

Reuter, M., Rosas, H.D., Fischl, B. (2010). Highly accurate inverse consistent registration: a robust approach. *NeuroImage*, *53*(4), 1181–1196.

Seixas, F.L., Débora, S., Saade, C., Conci, A., Souza, A., Tovar, F., Bramati, I. (2010). Anatomical Brain mri segmentation methods: Volumetric assessment of the hippocampus. IWSSIP 2010–17 Th International conference on systems, signals and image processing; 2010 Jan 17–19.

Simpson, M.I.G., Woods, W.P., Prendergast, G., Johnson, S.R., Green, G.G.R. (2012). Stimulus variability affects the amplitude of the auditory steady-state response. *PLOS ONE*, *7*(4), 1–10.

Smith, S.M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, *23*, S208 – S219. Mathematics in Brain Imaging.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., Stefano], N.D. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, *17*(1), 479–489.

Snowdon, D.A. (2003). Healthy aging and dementia: Findings from the nun study. *Annals of Internal Medicine*, *139*(5 Part 2), 450–454.

Starmans, M.P., van der Voort, S.R., Tovar, J.M.C., Veenland, J.F., Klein, S., Niessen, W.J. (2020). Chapter 18 - radiomics: Data mining using quantitative medical image features. In Zhou, S.K., Rueckert, D., Fichtinger, G. (Eds.) *Handbook of medical image computing and computer assisted intervention* (pp. 429–456): Academic Press.

Thrall, J.H., Li, X., Li, Q., Cruz, C., Do, S., Dreyer, K., Brink, J. (2018). Artificial intelligence and machine learning in radiology: Opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology*, *15*(3 Part B), 504–508. Data Science: Big Data Machine Learning and Artificial Intelligence.

Topol, E. (2019). Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK.

Vollmer, T., Signorovitch, J., Huynh, L., Galebach, P., Kelley, C., DiBernardo, A., Sasane, R. (2015). The natural history of brain volume loss among patients with multiple sclerosis: a systematic literature review and meta-analysis. *Journal of the Neurological Sciences*, *357*(1-2), 8–18.

Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G. (2003). Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. *NeuroImage*, *20*(2), 667–682.

Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M. (2009). Bayesian analysis of neuroimaging data in fsl. *NeuroImage*, *45*(1, Supplement 1), S173 – S186. Mathematics in brain imaging.

Yan, C., Gong, B., Wei, Y., Gao, Y. (2020). Deep multi-view enhancement hashing for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Yang, H., Xu, H., Li, Q., Jin, Y., Jiang, W., Wang, J., Wu, Y., Li, W., Yang, C., Li, X., et al. (2019). Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *General Psychiatry*, 32(2).

Zhang, Y., Brady, M., Smith, S. (2001). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57.

Zhou, C., Guan, X.-J., Guo, T., Zeng, Q.-L., Gao, T., Huang, P.-Y., Xuan, M., Gu, Q.-Q., Xu, X.-J., Zhang, M.-M. (2020). Progressive brain atrophy in parkinson's disease patients who convert to mild cognitive impairment. *CNS Neuroscience & Therapeutics*, *26*(1), 117–125.