# scientific reports

### OPEN



## Evaluating brain volume segmentation accuracy and reliability of FreeSurfer and Neurophet AQUA at variations in MRI magnetic field strengths

Hyunji Lee<sup>1</sup>, Hye Weon Kim<sup>1</sup>, Minho Lee<sup>1</sup>, Jimin Kang<sup>1</sup>, Donghyeon Kim<sup>1</sup>, Hyun Kook Lim<sup>2</sup>, Jun-Young Lee<sup>3,4</sup>, Eosu Kim<sup>5</sup> & Regina EY Kim<sup>1⊠</sup>

We aimed to compare the accuracy and reliability of two segmentation tools for magnetic resonance (MR) volumetry (FreeSurfer and Neurophet AQUA) at two magnetic field strengths (1.5T and 3T). We included 101 patients for the 1.5T–3T dataset and 112 for the 3T–3T dataset from three hospitals and five open-source datasets. The mean volume difference and average volume difference percentage with the change in magnetic field strength were compared between the methods. The hippocampus volume was larger with FreeSurfer than the Neurophet AQUA. In most brain regions, the Neurophet AQUA yielded a smaller average volume difference percentage (all < 10%) than FreeSurfer (all > 10%). The Neurophet AQUA exhibited more stable connectivity and regularity of the segmented components. Regarding volume, the Neurophet AQUA had effect sizes and ICCs comparable to those of FreeSurfer across the magnetic field strengths. With FreeSurfer, the original volume difference was small, whereas the average volume difference percentage was small with the Neurophet AQUA. Image segmentation took 1 h with FreeSurfer and 5 min with the Neurophet AQUA. When choosing an automatic segmentation method, the differences in image processing time and volume variability due to changes in the magnetic field strength of these methods should be considered.

Accurate and reliable analysis of structural brain volume using magnetic resonance (MR) imaging (MRI) is critical in clinical settings. In the long-term follow-up of patients with dementia, differences in image acquisition methods must be considered for clinical interpretation<sup>1</sup>. Within the same hospital, MRI equipment may be upgraded during the follow-up period, and patients can visit multiple medical centers that use different MR image-acquisition methods<sup>2,3</sup>. Such changes, including changes in scanner models and magnetic field strength, can introduce substantial variability in the results obtained from commonly used, MR-based, automatic brain segmentation software<sup>4–6</sup>. This variability is likely to exceed the actual effect of biological changes in the brain, potentially compromising the validity of clinical diagnoses and research results<sup>2,3</sup>.

Automated methods for brain segmentation from MR images are widely used for detailed and consistent analyses<sup>7,8</sup>. Numerous automated segmentation software programs have been clinically approved by regulatory bodies such as the U.S. Food and Drug Administration and the European Medicines Agency<sup>9</sup>. These programs are eligible for insurance claims and are extensively used in the clinical setting. For clinical use, such programs must process the results to ensure accurate interpretation of follow-up MR images, even with changes in MRI acquisition methods. However, among the many MRI parameters, a change in the magnetic field strength may have a major effect on MRI outcomes, resulting in inconsistent measurements across scans of the same patient and inconsistent results in automatic segmentation<sup>1,5</sup>.

<sup>1</sup>Research Institute, Neurophet Inc., 124, Teheran-ro, Gangnam-gu, Seoul 06234, Republic of Korea. <sup>2</sup>Department of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea. <sup>3</sup>Department of Psychiatry and Neuroscience Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea. <sup>4</sup>Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea. <sup>5</sup>Department of Psychiatry, Institute of Behavioral Science in Medicine, Brain Korea 21 FOUR Project for Medical Science, Yonsei University College of Medicine, Seoul, Republic of Korea. <sup>See</sup>email: reginaeunyoungkim@neurophet.com The performance of the automatic segmentation method in terms of the minimization of variability due to differences in the magnetic field strength can be evaluated from two main perspectives: accuracy and reliability<sup>10</sup>. Accuracy can be verified through visual assessment of segmentation quality and quantitative assessment of volume measurement. Reliability can be evaluated based on whether the volumetric measurements obtained from scans of the same patient by using different magnetic field strengths yield consistent results. Both aspects should be considered, as a bias toward either accuracy or reliability will not lead to valid results in clinical practice.

The aim of this study was to investigate the effects of differences in magnetic field strength and the test-retest reliability of structural MRI measurements. We compared the performance of two automated segmentation methods: the latest version of the widely used automated segmentation software, FreeSurfer<sup>11</sup>, and a commercially available software, Neurophet AQUA. Segmentation quality was visually evaluated for the two methods. The evaluation of reliability involved the scanning of patients by using both 1.5T and 3T MRI machines within 6 months. We determined the degree of reliability of the software for two images obtained via different magnetic field strengths and for two obtained with the same field strength (3T) within 6 months. We hypothesized that the reliability of Neurophet AQUA segmentation method for brain MRI between 1.5T and 3T would be similar to that of FreeSurfer and that use of the Neurophet AQUA segmentation method exhibited similar reliability to that of FreeSurfer, with its own benefits and disadvantages.

#### Results

#### **Demographic characteristics**

The 1.5T–3T dataset was utilized for reliability analysis, comprising a total of 101 patients: Asian  $(n=30; 4 \text{ men}, 26 \text{ women}; \text{age range: } 55–88 \text{ years}; \text{mean} \pm \text{standard deviation age: } 69.5 \pm 7.0 \text{ years}) \text{ collected from Yonsei}$ Severance Hospital (SEV) and Caucasian  $(n=71; 39 \text{ men}, 32 \text{ women}; \text{ age range: } 43–89 \text{ years}; 73.6 \pm 8.2 \text{ years})$ obtained from open-source databases, including the Alzheimer's Disease Neuroimaging Initiative (ADNI)<sup>12</sup>, Open Access Series of Imaging Studies 3 (OASIS3)<sup>13</sup>, and Single Individual volunteer for Multiple Observations across Networks databases (SIMON)<sup>14</sup>. The 3T–3T dataset was also utilized for reliability analysis, comprising a total of 112 patients, consisting of Asian  $(n=54; 18 \text{ men}, 36 \text{ women}; \text{ age range: } 56–85 \text{ years}; 74.0 \pm 7.8 \text{ years})$ collected from Yeouido St. Mary's Hospital (YID) and Boramae Hospital (BRM), Asian  $(n=45; 23 \text{ men}, 22 \text{ women}; \text{ age range: } 20–30 \text{ years}; 23.5 \pm 2.7 \text{ years})$  from the Beijing Connectivity-based Brain Imaging Research Database (C-BIRD)<sup>15</sup>, and Caucasian  $(n=13; 6 \text{ men}, 7 \text{ women}; \text{ age range: } 21–30 \text{ years}; 25.2 \pm 2.8 \text{ years})$  from the Precision In Neuroimaging database (PIN)<sup>4</sup>.

#### Segmentation quality

The segmentation quality was visually assessed for further comparison in this study. The evaluation was performed by comparing the ground truth (GT) produced through consensus by three radiologists with 15, 9, and 5 years of experience with Neurophet AQUA and FreeSurfer segmentation results. The evaluation variable was the Dice Similarity Coefficient (DSC), which is calculated by dividing the number of common voxels between the GT and segmentation tool result by the average number of voxels between the two results. The DSC is a number between 0 and 1, where 0 indicates that the two results have no elements in common, and 1 indicates that the two results are identical. The Dice Similarity Coefficient (DSC) was 0.8 or higher for both Neurophet AQUA and FreeSurfer regardless of the magnetic field strength in all participants (Table 1). The results of detailed regions are provided in Supplementary Table S1.

Segmentation with both methods was of sufficient quality, although subtle differences were observed. An example of 3T MR image segmentation performed on one patient using Neurophet AQUA and FreeSurfer methods is provided in Fig. 1. Compared with FreeSurfer, the Neurophet AQUA yielded stable connectivity, with the segmented region not invading other regions. As shown in the figure, this difference is especially visible in the hippocampus. FreeSurfer segmented the hippocampus by encroaching on the inferior lateral ventricle, whereas Neurophet AQUA method segmented the hippocampus into smaller parts without encroaching on the inferior lateral ventricle<sup>16</sup>.

	Neurophet AQUA	FreeSurfer	
	Mean ± SD Mean		
3T	$0.83 \pm 0.01$	$0.98 \pm 0.01$	
1.5T	$0.84 \pm 0.02$	$0.97 \pm 0.02$	
Total	$0.83 \pm 0.02$	$0.98 \pm 0.02$	

**Table 1**. Comparison of DSC against ground truth across automated segmentation methods. The mean and SD of the DSC values for 112 randomly selected patients compared to the ground truth results manually segmented by three experts from the 1.5T and 3T datasets. The DSC ranges from 0 to 1, with values closer to 1 indicating higher similarity to the ground truth. Abbreviation: DSC, dice similarity coefficient; Lt, left hemisphere; Rt, right hemisphere; SD, standard deviation.

.....



**Fig. 1**. Representative images of color-labeled images of the Neurophet AQUA and FreeSurfer segmentation methods. Examples of quality assessment for 3T MR images of one patient. On visual inspection, the Neurophet AQUA revealed stable connectivity (red arrows), and regularity of the segmented components compared to the FreeSurfer method. In the enlarged images, FreeSurfer showed rough boundaries. The differences in segmentation between the hippocampus and the inferior lateral ventricle are shown in detail in 3D. Abbreviation: MR, magnetic resonance.

	Lt.			Rt.			
	Neurophet AQUA	et AQUA FreeSurfer		Neurophet AQUA	FreeSurfer		
	Mean <u>+</u> SD	Mean <u>+</u> SD	p	Mean ± SD	Mean ± SD	p	
Cortical gray matter	218.0±32.2	$217.5 \pm 31.9$	0.816†	$217.3 \pm 31.5$	$217.3 \pm 31.6$	0.989†	
Cerebral gray matter	245.6±35.1	$244.3 \pm 35.2$	0.578†	$245.0 \pm 34.5$	$244.4 \pm 34.8$	0.798†	
Caudate	3.6 ± 0.5	3.5 ± 0.6	0.012	$3.6 \pm 0.6$	$3.6 \pm 0.6$	0.536†	
Putamen	$5.2 \pm 0.7$	$4.6 \pm 0.8$	< 0.001	$5.1 \pm 0.8$	$4.7 \pm 0.8$	< 0.001	
Pallidum	$1.8 \pm 0.3$	$2.0 \pm 0.3$	< 0.001	$2.3 \pm 0.3$	$1.9 \pm 0.3$	< 0.001	
Thalamus	$7.5 \pm 1.0$	7.0 ± 1.0	< 0.001	$6.9 \pm 0.9$	$7.0 \pm 1.0$	0.827†	
Amygdala	$1.6 \pm 0.3$	$1.4 \pm 0.4$	< 0.001	$1.8 \pm 0.3$	$1.6 \pm 0.3$	< 0.001	
Hippocampus	$3.4 \pm 0.5$	3.8 ± 0.6	< 0.001	$3.5 \pm 0.5$	$3.9 \pm 0.7$	< 0.001	
Inferior lateral ventricles	$1.4 \pm 0.9$	0.9 ± 0.9	< 0.001	$1.4 \pm 0.8$	$0.9 \pm 0.9$	< 0.001	
Cerebellum	67.3 ± 7.8	64.1 ± 7.9	< 0.001	66.5 ± 7.9	$64.0\pm8.0$	< 0.001	
Cerebral white matter	$226.3 \pm 26.9$	216.5 + 30.1	< 0.001	226.7 + 27.2	$215.6 \pm 30.4$	< 0.001	

**Table 2**. Comparisons of original volume obtained by automated segmentation methods. The analysis encompassed MR images of all patients (426 MR images of 213 patients), irrespective of the magnetic field strength, for comparison if the volume distributions. The volume units are cm<sup>3</sup>. † Not statistically significant. Abbreviation: Lt, left hemisphere; Rt, right hemisphere; SD, standard deviation.

#### Volume measurements

Table 2 provides a detailed comparison of the volumes measured using the two automated segmentation methods across various brain regions in both the right and left hemispheres<sup>17</sup>. In most regions (putamen, amygdala, inferior lateral ventricles, cerebellum, and cerebral white matter), slightly larger volumes were measured in both hemispheres via Neurophet AQUA than those measured via the FreeSurfer method (p < 0.001). In the case of

the pallidum, FreeSurfer yielded larger volumes in the left hemisphere than the Neurophet AQUA ( $1.8 \pm 0.3$  vs.  $2.0 \pm 0.3$  cm<sup>3</sup>, p < 0.001), and the Neurophet AQUA yielded larger volumes in the right hemisphere than FreeSurfer ( $2.3 \pm 0.3$  vs.  $1.9 \pm 0.3$  cm<sup>3</sup>, p < 0.001). In the left hemisphere, the Neurophet AQUA resulted in slightly larger volumes for the caudate nucleus ( $3.6 \pm 0.5$  vs.  $3.5 \pm 0.6$  cm<sup>3</sup>, p = 0.012), whereas no difference was observed in the right hemisphere ( $3.6 \pm 0.6$  vs.  $3.6 \pm 0.6$  cm<sup>3</sup>, p = 0.536). Similarly, the thalamus volume was deemed larger with the Neurophet AQUA in the left hemisphere ( $7.5 \pm 1.0$  vs.  $7.0 \pm 1.0$  cm<sup>3</sup>, p < 0.001), whereas no difference was detected in the right hemisphere ( $6.9 \pm 0.9$  vs.  $7.0 \pm 1.0$  cm<sup>3</sup>, p = 0.827). Notably, the hippocampus volume was deemed larger with FreeSurfer in both hemispheres ( $3.4 \pm 0.5$  vs.  $3.8 \pm 0.6$  cm<sup>3</sup>, p < 0.001 in the left hemisphere;  $3.5 \pm 0.5$  vs.  $3.9 \pm 0.7$  cm<sup>3</sup>, p < 0.001 in the right hemisphere). More detailed comparisons of the volume distributions of the 1.5T and 3T datasets are provided in Supplementary Tables S2 and S3, respectively.

#### Reliability

The comparison results of reliability across 1.5T-3T dataset are detailed in Table 3. For Neurophet AQUA, statistically significant differences in volume measurements were observed for most of the brain regions between the 1.5T and 3T datasets. Effect sizes were mostly small, ranging from -0.409 to 0.243. Similar results were obtained using the FreeSurfer method at the same magnetic field strengths. Statistically significant differences (p < 0.001) were detected for several brain regions, (the cortical gray matter, cerebral gray matter, pallidum, thalamus, amygdala, and cerebral white matter). Effect sizes ranged from -0.307 to 0.433, small to moderate, and ICCs ranged from 0.869 to 0.965, indicating satisfactory agreement.

The paired t-test results for both methods revealed a volume change with the magnetic field strength change. Considering both the effect size and ICC, both automated segmentation methods proved reliable in estimating brain region volumes across different magnetic field strengths.

Considering the significant volumetric discrepancies between the two automated segmentation methods (Table 2), a percentage-based comparison was added to the reliability evaluation (Table 4), to control for ROI size across the two methods.

Table 4 presents a comprehensive analysis of the average volume difference percentage (AVDP) between the Neurophet AQUA and FreeSurfer. The mean percentage difference in brain region volume between magnetic field strengths excluding relative volumetric trends between segmentation software, was assessed by calculating the amount of variance/bias in each of the datasets. The following equation was used:

$$AVDP = \frac{|1st MRI \operatorname{scan} ROI volume - 2nd MRI \operatorname{scan} ROI volume|}{(1st MRI \operatorname{scan} ROI volume + 2nd MRI \operatorname{scan} ROI volume)/2} \times 100 \,(\%) \,.$$

This evaluation metric provides insights into the range of volumetric variability. A smaller AVDP indicates better reliability. For this analysis, four subgroups were considered: (a) segmentation of the 1.5T–3T dataset by using the Neurophet AQUA and (c) FreeSurfer, and (b) segmentation of the 3T–3T dataset by using the Neurophet AQUA and (d) FreeSurfer. As no criterion was available for evaluation of the AVDP between the 1.5T and 3T datasets, the evaluation was based on the AVDP of the 3T–3T dataset: (b) was used as the criterion for the results of (a), and (d) was used as the criterion for (b). The difference in performance between the two automatic segmentation methods is confirmed by comparing (a) and (c).

Most AVDPs were low in the 3T-3T dataset for both methods, as the magnetic field strength did not change. In the case of the Neurophet AQUA, the AVDP for the 1.5T-3T dataset was lower than that for 3T-3T dataset in the following regions: the putamen in both hemispheres ( $2.91\%\pm2.48\%$  vs.  $3.04\%\pm2.04\%$ , p < 0.05 in the left hemisphere;  $2.73\%\pm2.49\%$  vs.  $3.21\%\pm2.09\%$ , p < 0.05 in the right hemisphere), the interior lateral ventricle ( $8.78\%\pm9.21\%$  vs.  $11.22\%\pm15.70\%$ , p < 0.05), and cerebellum ( $2.13\%\pm1.58\%$  vs.  $2.15\%\pm2.12\%$ , p < 0.05) of the right hemisphere. In the case of the AVDP of the 1.5T-3T dataset in the right thalamus of FreeSurfer was significantly lower than that of the 3T-3T dataset ( $5.05\%\pm3.74\%$  vs.  $5.41\%\pm4.48\%$ , p < 0.05).

The performances of the two automated segmentation methods were compared in terms of reliability across the 1.5T and 3T datasets. In all regions in which a significant difference in the AVDP was detected between the two methods, the Neurophet AQUA exhibited better reliability than FreeSurfer. In most regions (the caudate, putamen, pallidum, thalamus, amygdala, inferior lateral ventricles, cerebellum, and cerebral white matter) the Neurophet AQUA yielded a significantly smaller AVDP than FreeSurfer (p < 0.05). No significant differences were observed in the cortical gray matter, hippocampus, or cerebral gray matter AVDP. Figure 2 provides a detailed visualization of the differences in magnetic field strength between the hippocampus and cerebral white matter, confirming the reliability and stability of the two automated segmentation methods. In the case of the cerebral white matter, in which a significant AVDP difference was observed, the two methods greatly differed in terms of mean and SD ( $3.26\% \pm 2.54\%$  vs.  $6.99\% \pm 4.42\%$  in the left hemisphere and  $3.42\% \pm 2.43\%$  vs.  $7.18\% \pm 4.58\%$  in the right hemisphere).

#### Discussion

The present study investigated the impact of varying magnetic field strengths on the reliability and accuracy of automated brain segmentation methods, specifically comparing FreeSurfer (version 7.1.0) and an Neurophet AQUA deep-learning-based approach. Our findings, consistent with prior research, emphasize the importance of considering MRI acquisition variations, including changes in magnetic field strength, when analyzing structural brain volumes.

Results from the latest FreeSurfer version (7.1.0) were analyzed, revealing slight differences in volume determination between FreeSurfer and our Neurophet AQUA. Although this discrepancy is statistically

	Neurophet AQUA 1.5T-3T							
	1.5T	3T	Р	Effect size	ICC			
Cortical gray matter Lt.	201.05 [188.42, 216.64]	208.18 [194.7, 225.03]	< 0.001	-0.356	0.953			
Cortical gray matter Rt.	202.16 [186.31, 215.59]	207.09 [195.09, 226.94]	< 0.001	-0.415	0.936			
Cerebral gray matter Lt.	227.63 [213.30, 244.63]	234.62 [220.40, 253.76]	< 0.001	-0.350	0.955			
Cerebral gray matter Rt.	228.38 [211.80, 243.60]	232.94 [220.82, 254.40]	< 0.001	-0.409	0.938			
Caudate Lt.	3.46 [3.23, 3.77]	3.52 [3.32, 3.88]	< 0.001	-0.202	0.944			
Caudate Rt.	3.40 [3.21, 3.67]	3.52 [3.24, 3.83]	< 0.001	-0.320	0.925			
Putamen Lt.	4.88 [4.64, 5.26]	4.92 [4.59, 5.33]	0.139†	-0.048	0.973			
Putamen Rt.	4.73 [4.53, 5.11]	4.73 [4.47, 5.12]	0.542†	-0.019	0.976			
Pallidum Lt.	1.74 [1.59, 1.95]	1.77 [1.58, 1.92]	0.673†	0.022	0.925			
Pallidum Rt.	2.17 [1.99, 2.38]	2.20 [2.06, 2.43]	< 0.001	-0.209	0.944			
Thalamus Lt.	7.00 [6.57, 7.62]	7.18 [6.7, 7.82]	< 0.001	-0.261	0.948			
Thalamus Rt.	6.59 [6.18, 7.06]	6.70 [6.28, 7.26]	< 0.001	-0.264	0.922			
Amygdala Lt.	1.55 [1.36, 1.72]	1.67 [1.47, 1.80]	< 0.001	-0.371	0.923			
Amygdala Rt.	1.70 [1.46, 1.83]	1.75 [1.54, 1.92]	< 0.001	-0.325	0.938			
Hippocampus Lt.	3.22 [2.82, 3.55]	3.41 [2.96, 3.70]	< 0.001	-0.251	0.966			
Hippocampus Rt.	3.41 [3.04, 3.72]	3.58 [3.12, 3.81]	< 0.001	-0.320	0.948			
Inferior lateral ventricles Lt.	1.29 [1.03, 2.22]	1.27 [0.95, 2.06]	< 0.001	0.102	0.989			
Inferior lateral ventricles Rt.	1.34 [1.04, 1.95]	1.22 [1.00, 1.87]	< 0.001	0.068	0.994			
Cerebellum Lt.	65.30 [60.83, 68.97]	65.58 [61.87, 70.03]	< 0.001	-0.082	0.986			
Cerebellum Rt.	63.98 [60.49, 68.29]	64.86 [61.19, 68.89]	< 0.001	-0.105	0.986			
Cerebral white matter Lt.	222.58 [207.36, 241.05]	216.56 [199.52, 233.88]	< 0.001	0.243	0.967			
Cerebral white matter Rt.	226.53 [207.38, 241.43]	219.32 [199.33, 235.74]	< 0.001	0.241	0.966			
	FreeSurfer 1.5T-3T							
	FreeSurfer 1.5T-3T 1.5T	3T	Р	Effect size	ICC			
Cortical gray matter Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]	<b>3T</b> 206.81 [190.86, 225.31]	<b>P</b> < 0.001	Effect size -0.306	<b>ICC</b> 0.937			
Cortical gray matter Lt. Cortical gray matter Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66]	<b>P</b> < 0.001	Effect size -0.306 -0.305	ICC 0.937 0.930			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74]	P           < 0.001	Effect size           -0.306           -0.305           -0.284	ICC 0.937 0.930 0.943			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55]	P           < 0.001	Effect size -0.306 -0.305 -0.284 -0.287	ICC 0.937 0.930 0.943 0.937			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55] 3.33 [3.12, 3.88]	P         < 0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044	ICC 0.937 0.930 0.943 0.937 0.943			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55] 3.33 [3.12, 3.88] 3.52 [3.23, 3.95]	P         < 0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084	ICC 0.937 0.930 0.943 0.937 0.943 0.925			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55] 3.33 [3.12, 3.88] 3.52 [3.23, 3.95] 4.35 [4.01, 4.75]	P           < 0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.106	ICC 0.937 0.930 0.943 0.937 0.943 0.925 0.894			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55] 3.33 [3.12, 3.88] 3.52 [3.23, 3.95] 4.35 [4.01, 4.75] 4.42 [4.02, 4.75]	P           < 0.001	Effect size           -0.306           -0.305           -0.284           -0.287           -0.044           -0.084           -0.106           -0.096	ICC 0.937 0.930 0.943 0.937 0.943 0.925 0.894 0.915			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55] 3.33 [3.12, 3.88] 3.52 [3.23, 3.95] 4.35 [4.01, 4.75] 4.42 [4.02, 4.75] 1.93 [1.77, 2.13]	P           < 0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.106 -0.096 -0.296	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.925 0.894 0.915 0.869			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]	<b>3T</b> 206.81 [190.86, 225.31] 208.71 [190.33, 226.66] 232.35 [213.94, 251.74] 234.17 [215.82, 252.55] 3.33 [3.12, 3.88] 3.52 [3.23, 3.95] 4.35 [4.01, 4.75] 4.42 [4.02, 4.75] 1.93 [1.77, 2.13] 1.89 [1.73, 2.12]	P           <0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.084 -0.106 -0.096 -0.296 -0.257	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.925 0.894 0.915 0.869 0.900			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Thalamus Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.34 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.77 [6.32, 7.39]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]	P           < 0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.106 -0.096 -0.296 -0.257 0.406	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.943 0.925 0.894 0.915 0.869 0.900 0.908			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.34 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.77 [6.32, 7.39]           6.63 [6.17, 7.17]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]	P           < 0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.084 -0.106 -0.096 -0.296 -0.257 0.406 0.214	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.925 0.894 0.915 0.869 0.900 0.900 0.908 0.928			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt. Amygdala Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]	P           <0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.106 -0.096 -0.296 -0.257 0.406 0.214 -0.307	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.945 0.869 0.900 0.908 0.908 0.908 0.928 0.915			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]	P           <0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.106 -0.096 -0.296 -0.257 0.406 0.214 -0.307 -0.288	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.945 0.869 0.900 0.908 0.908 0.908 0.928 0.915 0.911			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Rt. Hippocampus Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]	P           <0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.084 -0.106 -0.096 -0.296 -0.296 -0.257 0.406 0.214 -0.307 -0.288 -0.071	ICC 0.937 0.930 0.943 0.943 0.925 0.894 0.915 0.869 0.900 0.900 0.908 0.928 0.915 0.911 0.911			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Lt. Cerebral gray matter Lt. Caudate Lt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Rt. Hippocampus Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]           3.69 [3.30, 4.06]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]           3.80 [3.38, 4.17]	P           <0.001	Effect size -0.306 -0.305 -0.284 -0.287 -0.044 -0.084 -0.106 -0.096 -0.296 -0.257 0.406 0.214 -0.307 -0.307 -0.288 -0.071 -0.103	ICC 0.937 0.930 0.943 0.943 0.925 0.894 0.915 0.869 0.900 0.900 0.908 0.928 0.915 0.911 0.911 0.965 0.957			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Rt. Hippocampus Lt. Hippocampus Rt. Inferior lateral ventricles Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]           3.69 [3.30, 4.06]           0.72 [0.48, 1.68]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]           3.80 [3.38, 4.17]           0.82 [0.43, 1.85]	P           <0.001	Effect size           -0.306           -0.305           -0.284           -0.287           -0.044           -0.084           -0.096           -0.297           -0.406           -0.297           -0.298           -0.297           0.406           -0.214           -0.307           -0.288           -0.071           -0.103	ICC 0.937 0.930 0.943 0.943 0.943 0.925 0.894 0.915 0.869 0.900 0.900 0.908 0.915 0.911 0.911 0.911 0.957 0.989			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Lt. Hippocampus Lt. Hippocampus Rt. Inferior lateral ventricles Lt. Inferior lateral ventricles Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.34 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]           3.69 [3.30, 4.06]           0.72 [0.48, 1.68]           0.64 [0.45, 1.40]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]           3.80 [3.38, 4.17]           0.82 [0.43, 1.85]           0.67 [0.43, 1.47]	P           <0.001	Effect size           -0.306           -0.287           -0.287           -0.044           -0.084           -0.096           -0.296           -0.257           0.406           0.214           -0.307           -0.288           -0.071           -0.103           -0.103	ICC 0.937 0.930 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.915 0.900 0.908 0.915 0.911 0.965 0.957 0.989 0.985			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Lt. Amygdala Lt. Amygdala Lt. Hippocampus Lt. Hippocampus Lt. Inferior lateral ventricles Lt. Inferior lateral ventricles Rt. Cerebellum Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.34 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]           3.69 [3.30, 4.06]           0.72 [0.48, 1.68]           0.64 [0.45, 1.40]           6.1,91 [57.50, 66.02]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]           3.80 [3.38, 4.17]           0.82 [0.43, 1.85]           0.67 [0.43, 1.47]           61.37 [57.62, 65.45]	P           <0.001	Effect size           -0.306           -0.287           -0.287           -0.044           -0.07           -0.084           -0.084           -0.106           -0.296           -0.297           0.406           0.214           -0.307           -0.288           -0.071           -0.206           -0.103           -0.103	ICC 0.937 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.945 0.869 0.900 0.908 0.908 0.915 0.911 0.965 0.957 0.989 0.985 0.974			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Lt. Hippocampus Lt. Hippocampus Lt. Inferior lateral ventricles Lt. Inferior lateral ventricles Rt. Cerebellum Lt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]           3.69 [3.30, 4.06]           0.72 [0.48, 1.68]           0.64 [0.45, 1.40]           61.91 [57.50, 66.02]           61.58 [57.15, 65.85]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]           3.80 [3.38, 4.17]           0.82 [0.43, 1.85]           0.67 [0.43, 1.47]           61.37 [57.62, 65.45]           61.95 [57.32, 66.11]	P           <0.001	Effect size           -0.306           -0.284           -0.287           -0.044           -0.070           -0.084           -0.071           -0.288           -0.071           -0.084           -0.307	ICC 0.937 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.943 0.915 0.869 0.900 0.908 0.908 0.915 0.911 0.965 0.957 0.989 0.985 0.985			
Cortical gray matter Lt. Cortical gray matter Rt. Cerebral gray matter Rt. Cerebral gray matter Rt. Caudate Lt. Caudate Rt. Putamen Lt. Putamen Rt. Pallidum Lt. Pallidum Rt. Thalamus Lt. Thalamus Rt. Amygdala Lt. Amygdala Rt. Hippocampus Lt. Hippocampus Lt. Inferior lateral ventricles Lt. Inferior lateral ventricles Rt. Cerebellum Rt. Cerebellum Rt.	FreeSurfer 1.5T-3T           1.5T           201.93 [184.47, 217.77]           201.25 [184.77, 220.52]           227.76 [208.22, 243.38]           227.37 [209.61, 246.9]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.38 [3.09, 3.82]           3.46 [3.24, 3.86]           4.24 [4.01, 4.60]           4.30 [4.04, 4.62]           1.90 [1.71, 2.06]           1.87 [1.69, 2.01]           6.63 [6.17, 7.17]           1.24 [1.05, 1.45]           1.49 [1.28, 1.61]           3.63 [3.20, 3.97]           3.69 [3.30, 4.06]           0.72 [0.48, 1.68]           0.64 [0.45, 1.40]           61.91 [57.50, 66.02]           61.58 [57.15, 65.85]           214.55 [197.71, 233.38]	3T           206.81 [190.86, 225.31]           208.71 [190.33, 226.66]           232.35 [213.94, 251.74]           234.17 [215.82, 252.55]           3.33 [3.12, 3.88]           3.52 [3.23, 3.95]           4.35 [4.01, 4.75]           4.42 [4.02, 4.75]           1.93 [1.77, 2.13]           1.89 [1.73, 2.12]           6.50 [5.93, 6.98]           6.49 [5.93, 7.03]           1.36 [1.10, 1.54]           1.52 [1.37, 1.69]           3.70 [3.14, 4.04]           3.80 [3.38, 4.17]           0.82 [0.43, 1.85]           0.67 [0.43, 1.47]           61.37 [57.62, 65.45]           61.95 [57.32, 66.11]           205.89 [184.1, 221.31]	P           <0.001	Effect size           -0.306           -0.305           -0.284           -0.287           -0.044           -0.044           -0.026           -0.106           -0.296           -0.297           -0.440           -0.296           -0.297           0.406           0.214           -0.307           -0.288           -0.071           -0.103           -0.026           0.014           0.055           -0.042           0.409	ICC 0.937 0.943 0.943 0.943 0.943 0.925 0.894 0.915 0.869 0.900 0.908 0.908 0.928 0.915 0.911 0.957 0.957 0.985 0.985 0.974 0.973 0.987			

**Table 3**. Original volume comparison results between 1.5 T and 3 T. The original volume (median [25th percentile, 75th percentile]) at each magnetic field strength and the p-value of the paired t-test for the comparison of the brain region volume of each patient, effect size, and ICCs are presented together. The volume units are cm<sup>3</sup>. †Not statistically significant. Abbreviation: ICC, interclass correlation coefficient; Lt, left hemisphere; Rt, right hemisphere.

	Neurophet AQUA		FreeSurfer		
	(a) 1.5T-3T	(b) 3T-3T	(c) 1.5T-3T	(d) 3T-3T	
Cortical gray matter Lt. <sup>1,3</sup>	$3.93 \pm 2.50$	$2.16 \pm 2.23$	$4.43 \pm 3.59$	$2.73 \pm 2.28$	
Cortical gray matter Rt. <sup>1,3</sup>	$4.58 \pm 3.04$	$2.59 \pm 2.78$	$4.75 \pm 3.63$	$3.01 \pm 2.52$	
Cerebral gray matter Lt. <sup>1,3</sup>	3.69±2.36	$2.07 \pm 2.17$	$4.07 \pm 3.27$	$2.61 \pm 2.24$	
Cerebral gray matter Rt. <sup>1,3</sup>	$4.36 \pm 2.87$	$2.48 \pm 2.69$	$4.35 \pm 3.30$	$2.90 \pm 2.45$	
Caudate Lt. <sup>1,2</sup>	3.76±3.32	$1.98 \pm 1.77$	$5.72 \pm 4.68$	$4.27 \pm 3.74$	
Caudate Rt. <sup>1,3</sup>	$4.82 \pm 4.36$	$2.07 \pm 2.22$	$6.14 \pm 5.16$	$3.65 \pm 4.89$	
Putamen Lt. <sup>2,3</sup>	$2.91 \pm 2.48$	$3.04 \pm 2.04$	$6.04 \pm 5.92$	$4.60 \pm 5.02$	
Putamen Rt. <sup>2,3</sup>	$2.73 \pm 2.49$	$3.21 \pm 2.09$	$5.64 \pm 6.44$	3.16±2.91	
Pallidum Lt. <sup>2,3</sup>	$5.07 \pm 4.68$	$4.09 \pm 2.86$	$7.02 \pm 5.38$	$4.68 \pm 3.42$	
Pallidum Rt. <sup>2,3</sup>	$4.45 \pm 3.45$	$5.07 \pm 2.83$	$6.51 \pm 5.21$	$4.87 \pm 4.50$	
Thalamus Lt. <sup>1,2,3</sup>	3.38±3.11	$2.12 \pm 1.67$	$5.69 \pm 4.26$	$3.56 \pm 2.83$	
Thalamus Rt. <sup>1,2,3</sup>	$4.87 \pm 2.80$	$2.61 \pm 2.22$	$5.05 \pm 3.74$	$5.41 \pm 4.48$	
Amygdala Lt. <sup>1,2,3</sup>	7.49±6.13	$3.75 \pm 3.12$	$11.15 \pm 7.71$	$6.74 \pm 6.25$	
Amygdala Rt. <sup>1,2,3</sup>	$6.99 \pm 5.64$	$3.79 \pm 3.04$	$9.37 \pm 7.33$	$5.58 \pm 5.17$	
Hippocampus Lt. <sup>1</sup>	$4.55 \pm 3.68$	$2.74 \pm 2.28$	$4.16 \pm 4.02$	$3.20 \pm 2.39$	
Hippocampus Rt. <sup>1</sup>	$5.56 \pm 4.55$	$3.29 \pm 2.77$	$4.60 \pm 7.38$	$4.02 \pm 4.32$	
Inferior lateral ventricles Lt. <sup>2</sup>	$9.90 \pm 7.80$	$8.85 \pm 8.74$	$14.25 \pm 13.41$	$14.85 \pm 12.44$	
Inferior lateral ventricles Rt. <sup>2</sup>	$8.78 \pm 9.21$	$11.22 \pm 15.70$	$12.61 \pm 15.42$	$13.76 \pm 14.05$	
Cerebellum Lt. <sup>2,3</sup>	$1.84 \pm 1.83$	$1.73 \pm 1.72$	$2.89 \pm 2.20$	$1.61 \pm 1.47$	
Cerebellum Rt. <sup>2,3</sup>	$2.13 \pm 1.58$	$2.15 \pm 2.12$	$2.73 \pm 2.56$	$2.35 \pm 2.27$	
Cerebral white matter Lt. <sup>1,2,3</sup>	$3.26 \pm 2.54$	$1.72 \pm 1.69$	$6.99 \pm 4.42$	$2.16 \pm 2.94$	
Cerebral white matter Rt. <sup>1,2,3</sup>	$3.42 \pm 2.43$	$2.03 \pm 2.16$	$7.18 \pm 4.58$	$2.20 \pm 3.37$	

**Table 4**. Comparisons of AVDP obtained via the Neurophet AQUA and FreeSurfer in each subgroup. The AVDP units are %. <sup>1</sup>Statistically significant difference between (a) and (b). <sup>2</sup>Statistically significant difference between (c) and (d). Abbreviation: AVDP, average volume difference percentage.

- -

significant, the effect size is generally small, suggesting that the Neurophet AQUA measures volume numerically like FreeSufer. Further visual evaluation confirmed that both FreeSufer and Neurophet AQUA produced segmentation results of sufficient quality when compared to GT from an experienced radiologist. However, FreeSufer showed less smooth segmentation results than the Neurophet AQUA. In FreeSurfer, the segmentation results showed rough and uneven boundary that invaded other regions. In quantitative terms, FreeSurfer consistently resulted in larger hippocampal volumes.

Reliability analysis indicated the robustness of both methods in estimating brain region volumes across diverse magnetic field strengths. Significantly, our study showcased that our Neurophet AQUA outperformed FreeSurfer in terms of both variability and rate of change, even when confronted with varying magnetic field strengths ranging from 1.5T to 3T. With the Neurophet AQUA, the average of AVDP did not exceed 10% in any region, but with FreeSufer, the AVDP of the left amygdala (11.15% $\pm$ 7.71%) and the inferior lateral ventricles in both hemispheres (14.25% $\pm$ 13.41% in the left hemisphere and 12.61% $\pm$ 15.42% in the right hemisphere) exceeded 10%. Although no significant difference in AVDP was observed between the two methods for the hippocampus, the Neurophet AQUA generally yielded a more consistent AVDP than FreeSurfer.

This study contributes to the field in that we evaluated the effect of differences in magnetic field strength on automated brain segmentation methods according to small brain regions. Additionally, utilizing data from multiple hospitals and public databases, our study represents real-world data variability. A thorough reliability analysis, considering effect size and ICC, comprehensively assessed segmentation method performance. Percentage-based (ADVP) comparisons highlighted the Neurophet AQUA's superior reliability across various brain regions.

However, this study also had several limitations. First, although we included a diverse dataset, we cannot rule out the potential influence of other uncontrolled variables, such as variations in image acquisition protocols within each magnetic field strength group. Additionally, we did not consider other factors that might have impacted the MRI data, such as differences in head position, motion artifacts, or scanner-specific hardware and software configurations<sup>4</sup>. Thus, we evaluated the variability of only the magnetic field strength. In future studies, these other factors should be explored in detail. Therefore, the generalizability of our results is limited, and they should be interpreted with caution. Second, the observed volume differences between the two automated segmentation methods, although statistically significant, may have limited clinical significance. Further research is required to determine the clinical relevance of these differences<sup>18</sup>. The same applies to the visual quality of the two methods. In this study, we targeted small brain regions. As visual confirmation of such small brain regions is challenging, the accuracy of such assessment may be improved if several experts evaluate the results of FreeSurfer and the Neurophet AQUA<sup>19</sup>.



**Fig. 2.** Bland–Altman plots at 1.5T and 3T X-axis: mean brain volume at 1.5T and 3T. The unit of volume on the X-axis is  $cm^3$ . Y-axis: AVDP between 1.5T and 3T. The mean, lower (-1.96 SDs), and upper (+1.96 SDs) limits of agreement are indicated by dotted lines. A negative difference on the y-axis indicates that the brain volume measured via 1.5T was smaller than that measured via 3T. Abbreviations: Lt, left hemisphere; Rt, right hemisphere; AVDP, average volume difference percentage; SD, standard deviation.

We demonstrated that MRI datasets with different magnetic field strengths can be used to obtain robust measurements of brain volume, using both the latest version of the widely used automatic segmentation software<sup>20</sup>, FreeSurfer, and our proposed Neurophet AQUA. Both methods yielded high accuracy and reliability across different brain regions. However, the strengths and weaknesses of the two methods differ and should be considered in their selection. FreeSurfer yields a small difference in the original volume between 1.5T and 3T MR images, but considered as a variable percentage, the difference is larger than that of the Neurophet AQUA, the operation time is 1 h per image, and some image failures occurred. The Neurophet AQUA has a relatively large difference in the original volume, but the AVDP percentage is significantly smaller than that of FreeSurfer, and the SD is small; therefore, it yields more consistent results. In addition, segmentation with the Neurophet AQUA was possible within 5 min per image, and no segmentation failures occurred.

These results provide valuable insights for clinicians and researchers who use automated segmentation methods in various settings. The exceptional reliability of the Neurophet AQUA in specific brain regions suggests its potential advantages for specific applications. Further research on 7T over 1.5T and 3T MRI is needed to improve the accuracy and reliability of automated brain segmentation in clinical practice and research<sup>21</sup>.

#### Methods

#### Study participants

A graphical representation of the participant inclusion process is provided in Fig. 3. Data were sourced from three domestic hospitals and five publicly available databases. The following exclusion criteria were applied to participants within these datasets: (1) contraindications to MRI (e.g., a pacemaker or metal in the body), (2) history of head trauma, and (3) own or first/second-degree family history of neurological or psychiatric disorders<sup>22</sup>. After applying these exclusion criteria, the data of 233 patients were evaluated in this study, comprising 112 patients with images obtained with both 1.5T and 3T magnetic field strengths (the 1.5T–3T dataset) and 121 with both images obtained with a magnetic field strength of 3T (3T–3T dataset). To ensure data



**Fig. 3**. Flow chart of the study population. The flow chart of this study is organized from initial collection to final analysis. Abbreviations: CN, cognitive normal; MCI, mild cognitive impairment; AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative database; OASIS3, Open Access Series of Imaging Studies 3 database; SIMON, Single Individual volunteer for Multiple Observations across Networks database; C-BIRD, Connectivity-based Brain Imaging Research Database at Beijing Normal University; PIN, Precision in Neuroimaging database.

quality, we excluded images that failed the analysis in FreeSurfer and those that were blurry from the analysis, resulting in the inclusion of 223 patients in the study.

#### Informed consent

Use of the data obtained from the hospitals in Korea were reviewed by each institutional review board (Yonsei Severance Hospital: 1-2021-0025, Yeouido St. Mary's Hospital: SC20DSDP0121, and Boramae Hospital: 30-2020-325) according to ethical research practices. All patients provided informed consent for the use of their data for research, and we strictly adhered to Korean data protection and privacy regulations, ensuring confidentiality and security of the data throughout the research process.

#### Image acquisition

MRI data were acquired using scanners from three well-known manufacturers in the field of MRI technology: Philips, SIEMENS, and GE, representing a wide range of instrument models, according to the multicenter nature of the study. The image acquisition protocols are summarized in Table 5. We believe that this dataset effectively mirrors the variability that occurs in in clinical practice.

#### Automated methods for image volumetry processing

T1-weighted volumetric images were processed using automated segmentation. Each patient's brain MRI was uploaded to the server of the software tool, and both methods employed a cross-sectional pipeline for test-retest analysis. This approach involved processing the scans at every time point independently for each patient, allowing for a comprehensive evaluation of the brain structures in each MRI session.

#### FreeSurfer

The FreeSurfer software version 7.1.0 (https://surfer.nmr.mgh.harvard.edu, Harvard University, Boston, MA, USA) uses a template-driven approach for volumetric and surface-based segmentation<sup>11,23</sup>. The processing involved automated steps for quantification of brain region volumes and detection of potential abnormalities. Initially, data preprocessing, including motion correction, skull stripping<sup>24</sup>, and intensity normalization, was performed to enhance image quality and consistency. Subsequently, the brain was inflated to a spherical shape for precise alignment to a common spherical space by using Talairach atlas coordinates. FreeSurfer's advanced segmentation algorithms were used to identify various brain regions, such as cortical structures, subcortical structures, and ventricles. The software provided 110 volume labels, and additional cortical or cerebral gray matter volumes were calculated using the provided formulae. However, image volumetry processing with this software demonstrated relatively low efficiency, with a processing time of 3 h per patient.

Dataset	Site	Magnetic field strength	Manufacturer	Model	RT	ET	ST	FA
SEV	CEV	1.5T	Philips	Achieva	7.274	3.287	1	8
	SEV	3T	Philips	Ingenia CX	7.127	3.253	1	8
		1.5T	Philips	Achieva	8.619	4.003	1.2	8
		3T	Philips	Achieva	6.807	3.161	1.2	8
		1.5T	Siemens	Sonata	3000	3.540	1.2	8
		3T	Siemens	Allegra	2300	2.910	1.2	9
1 FT 2T pair	ADNI	1.5T	Philips	Intera	8.614	4.001	1.2	8
1.51–51 pair		3T	Philips	Intera	6.800	3.158	1.2	8
0.1070		1.5T	Siemens	Trio	3000	3.610	1.2	8
		3T	Siemens	Symphony	2300	2.940	1.2	9
	OASTS2	1.5T	Siemens	Sonata	1.900	0.004	1	15
	UASIS5	3T	Siemens	Trio	2.400	0.003	1	8
	SIMON	1.5T	Siemens	Sonata	3.000	0.004	1	8
	SIMON	3T	Siemens	Trio	2300	2.980	1	9
VID	3T	Siemens	Skyra	1860	2.530	1	9	
		3T	Siemens	Magnetom VIDA	1860	2.520	1	9
	DDM	3T	Philips	Ingenia CX	6.503	2.976	1	9
27 27	BRM	3T	Siemens	Magnetom VIDA	2300	2.980	1	9
C-BIRD	C DIDD	3T	Siemens	Trio	2530	3.390	1.33	7
	C-DIRD	3T	Siemens	Trio	2530	3.390	1.33	7
	DIN	3T	GE	Discovery MR750	7.360	3.050	1	11
	PIIN	3T	GE	Discovery MR750	7.360	3.050	1	11

**Table 5**. Site-specific image acquisition. Abbreviation: RT, repeat time; ET, echo time; ST, slice thickness; FA, flip angle; SEV, Yonsei Severance Hospital; ADNI, Alzheimer's Disease Neuroimaging Initiative; OASIS3, Open Access Series of Imaging Studies 3; SIMON, Single Individual volunteer for Multiple Observations across Networks; YID, Yeouido St. Mary's Hospital; BRM, Boramae Hospital; C-BIRD, Connectivity-based Brain Imaging Research Database; PIN, Precision in Neuroimaging.

#### **Neurophet AQUA**

For the purpose of AQUA segmentation, we used the training data of 746 T1-weighted images with a high quality of 300, a medium quality of 363 and a low quality of 83. The training data also gathered considering multicenter and multi-vendor data as described in previous study<sup>25</sup>. Based on the Nested U-Net<sup>26</sup>, we developed a deep learning-based segmentation algorithm using the SAU-Net architecture<sup>25</sup>. As part of preprocessing, image resampling was performed with 1 mm iso-cubic and right-anterior-superior reorientation. A histogram-based intensity regularization was applied before the images were cropped to local patches by using uniform sampling. The local patches were fed into the network and the network was iteratively trained using data augmentation techniques such as Gaussian noise, random blur, bias field, affine transformation, and elastic deformation. In the next step, we used the 3D ResNeSt block with a channel-wise attention module and an Evolving Normalization-Activation (EvoNorm) layer in each convolution block. After this, we performed hierarchical transfer learning with freeze the weights except for the normalization and output layers using a level-wise training dataset for image quality (Fig. 4). Then, the predicted results from the 3D local patches were segmented and combined with a fixed size and overlay to segment the entire MR image more efficiently (Fig. 4). With these processes, approximately 100 regions of interest (ROIs) were obtained in the cortex and subcortex in approximately five minutes.

#### Statistical analysis

Statistical analysis was performed using R version 4.2.2 and Python 3.10.12 to assess brain region volumes and identify significant differences between groups. Parametric and non-parametric statistical tests were used according to the results of normality (the Shapiro–Wilk test) and equal variances (Levene's test) tests.

We determined the differences in volume measurements by region between the two automated segmentation methods by subdividing the 1.5T and 3T datasets for each method and obtaining the original volume for each region. The volume difference between the segmentation methods for the entire dataset was deemed more important than the volume difference per patient; therefore, an independent t-test was conducted.

Reliability was evaluated for both the 1.5T-3T dataset and the 3T-3T dataset. The reliability result of the 3T-3T dataset was used as reference during evaluation of the 1.5T-3T dataset reliability. The evaluation largely comprised a comparison of the original volume and a comparison of volume percentages.

First, we used paired t-tests to evaluate the deflection of the original volume that occurred as a function of the field strength for each automated segmentation method. The ICC and effect size were calculated for accurate interpretation, considering not only the statistical significance of the differences but also their magnitude and concordance<sup>27</sup>. The reliability of each method for each magnetic field strength was assessed as the absolute agreement of single measurements by using a two-way mixed-model ICC. We interpreted the ICC values according to the following guideline: values less than 0.50, between 0.50 and 0.75, between 0.75 and 0.90, and greater than 0.9 are indicative of poor, moderate, good, and excellent reliability, respectively<sup>28</sup>. Cohen's d was calculated as indicator of effect size to determine the practical significance of the observed differences in brain region volume. The following guidelines were used to interpret effect sizes: an effect size  $\leq 0.2$  is small, 0.5 is medium, and  $\geq 0.8$  is large<sup>29</sup>.

The non-parametric Kruskal–Wallis test was used to compare AVDP distributions between multiple groups, and the Holm test for post hoc analysis was performed to identify specific between-group differences. We conducted a more comprehensive evaluation of these disparities via Bland–Altman plots, focusing on critical regions. These plots provide a visual representation of the potential systematic biases and extent of variability among the measurements. A mean difference of approximately zero in the plot implies the absence of systematic bias. Additionally, the compactness of the agreement limits reveals the degree of disparity between the measurements at 3T and 1.5T.



Fig. 4. Illustrations of hierarchical transfer learning using a level-wise training dataset for image quality.

Scientific Reports | (2024) 14:24513

All statistical analyses were two-sided and performed with the significance level set at p < 0.05. The results of the statistical tests were interpreted to draw meaningful conclusions regarding the impact of the studied factors on brain region volume.

#### Data availability

The data that support the findings will be available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Received: 15 January 2024; Accepted: 27 September 2024 Published online: 18 October 2024

#### References

- 1. Lysandropoulos, A. P. et al. Quantifying brain volumes for multiple sclerosis patients follow-up in clinical practice comparison of 1.5 and 3 Tesla magnetic resonance imaging. Brain Behav. 6, e00422. https://doi.org/10.1002/brb3.422 (2016).
- 2. Jovicich, J. et al. Brain morphometry reproducibility in multi-center 3T MRI studies: A comparison of cross-sectional and longitudinal segmentations. Neuroimage 83, 472-484 (2013).
- 3. Lee, H., Nakamura, K., Narayanan, S., Brown, R. A. & Arnold, D. L. Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. Neuroimage 184, 555-565 (2019).
- 4. Hedges, E. P. et al. Reliability of structural MRI measurements: The effects of scan session, head tilt, inter-scan interval, acquisition sequence, FreeSurfer version and processing stream. Neuroimage 246, 118751. https://doi.org/10.1016/j.neuroimage.2021.118751 (2022)
- 5. Ammari, S. et al. Influence of magnetic field strength on magnetic resonance imaging radiomics features in brain imaging, an in vitro and in vivo study. Front. Oncol. 10, 541663. https://doi.org/10.3389/fonc.2020.541663 (2021).
- 6. Hu, Y. et al. A multi-modal MRI data processing pipeline with test-retest reliability and predicative validity assessments. Hum. Brain Mapp. 44, 2062-2084 (2023).
- 7. Brewer, J. B., Magda, S., Airriess, C. & Smith, M. E. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. Am. J. Neuroradiol. 30, 578-580 (2009).
- 8. van den Heuvel, T. L. A. et al. Automated detection of cerebral microbleeds in patients with traumatic brain Injury. Neuroimage Clin. 12, 241-251 (2016).
- 9. Lee, J. Y., Park, J. E., Chung, M. S., Oh, S. W. & Moon, W. J. Expert opinions and recommendations for the clinical use of quantitative analysis software for MRI-based brain volumetry. Taehan Yongsang Uihakhoe Chi 82, 1124-1139 (2021)
- 10. Heinen, R. et al. Robustness of automated methods for brain volume measurements across different MRI field strengths. PLoS ONE 11, e0165719, https://doi.org/10.1371/journal.pone.0165719 (2016).
- 11. Fischl, B. FreeSurfer. Neuroimage 62, 774-781 (2012).
- 12. Petersen, R. C. et al. Alzheimer's disease neuroimaging initiative (ADNI): Clinical characterization. Neurology 74, 201-209 (2010).
- 13. Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C. & Buckner, R. L. Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults. J. Cognit. Neurosci. 22, 2677-2684 (2010).
- 14. Duchesne, S. et al. The Canadian dementia imaging protocol: Harmonizing national cohorts. J. Magn. Reason Imaging 49, 456-465 (2019).
- 15. Lin, Q. et al. A connectivity-based test-retest dataset of multi-modal magnetic resonance imaging in young healthy adults. Sci. Data 2, 150056. https://doi.org/10.1038/sdata.2015.56 (2015).
- 16. Ochs, A. L., Ross, D. E., Zannoni, M. D., Abildskov, T. J. & Bigler, E. D. For the Alzheimer's disease neuroimaging initiative. Comparison of automated brain volume measures obtained with NeuroQuant and FreeSurfer. J. Neuroimaging 25, 721-727 (2015).
- 17. Lee, J. Y. et al. Clinically available software for automatic brain volumetry: Comparisons of volume measurements and validation of intermethod reliability. Korean J. Radiol. 22, 405-414 (2021).
- 18. Marchewka, A. et al. Influence of magnetic field strength and image registration strategy on voxel-based morphometry in a study of Alzheimer's disease. Hum. Brain Mapp. 35, 1865-1874 (2014).
- 19. Mulder, E. R. et al. Hippocampal volume change measurement: quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. Neuroimage 92, 169-181 (2014).
- 20. Gronenschild, E. H. B. M. et al. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS ONE 7, e38234. https://doi.org/10.1371/journal.pone.0038234 (2012).
- 21. Kerchner, G. A. Ultra-high field 7T MRI: A new tool for studying Alzheimer's disease. J. Alzheimers Disease 26, 91-95 (2011).
- 22. Reijmer, Y. D., Leemans, A., Brundel, M., Kappelle, L. J. & Biessels, G. J. Disruption of the cerebral white matter network is related to slowing of information processing speed in patients with type 2 diabetes. Diabetes 62, 2112-2115 (2013).
- 23. Dale, A. M. & Fischl, B. Sereno, M. I. Cortical surface-based analysis. Neuroimage 9, 179-194 (1999).
- Ségonne, F. et al. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22, 1060–1075 (2004).
   Lee, M. et al. Split-attention U-Net: A fully convolutional network for robust multi-label segmentation from brain MRI. *Brain Sci.* 10, 974. https://doi.org/10.3390/brainsci10120974 (2020).
- 26. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. & Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. Deep. Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support (2018) 11045, 3-11 (2018).
- 27. Olejnik, S. & Algina, J. Measures of effect size for comparative studies: Applications, interpretations, and limitations. Contemp. Educ. Psychol. 25, 241-286 (2000).
- 28. Koo, T. K. & Li, M. Y. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15, 155-163 (2016)
- 29. Sullivan, G. M. & Feinn, R. Using effect size-or why the P value is not enough. J. Grad Med. Educ. 4, 279-282 (2012).

#### Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute(KHIDI), funded by the Ministry of Health&Welfare, Republic of Korea (grant number : RS-2020-KH107144).

#### **Author contributions**

H.L., R.E.Y.K. and H.W.K. contributed to conceptualization.H.L. drafted the original manuscript, conducted statistical analysis, performed data collection, and prepared tables and figures. R.E.Y.K. and D.K. provided project leadership, supervised the study, granted approval, and curated the data. R.E.Y.K. and H.W.K. participated in study supervision, formal analysis. M.L. and J.K. contributed to software support. D.K. managed resources and funding.H.K.L., J.Y.L., and E.K. were responsible for patient data collection and care. All authors critically reviewed and approved the final manuscript.

#### Funding

We are grateful for the financial support that made this research possible, provided by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2020-KH107144).

#### Declarations

#### **Competing interests**

The authors declare no competing interests.

#### Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi. org/10.1038/s41598-024-74622-y.

Correspondence and requests for materials should be addressed to R.E.K.

#### Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024