Improving performance robustness of subject-based brain segmentation software

Jong-Hyeok Park[1], Kyung-Il Park[2,3], Dongmin Kim[1], Myungjae Lee[1], Shinuk Kang[1], Seung Joo Kang[4], Dae Hyun Yoon[5]; for the Alzheimer's Disease Neuroimaging Initiative*

[1]JLK, Seoul, Korea
[2]Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea
[3]Division of Neurology, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Korea
[4]Division of Gastroenterology, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Korea
[5]Division of Psychiatry, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Korea

Purpose

Artificial intelligence (AI)-based image analysis tools to quantify the brain have become commercialized. However, insufficient data for learning and scanner specificity is a limitation for achieving high quality. In the present study, the performance of personalized brain segmentation software when applied to multicenter data using an AI model trained on data from a single institution was improved.

Methods

Preindicators of brain white matter (WM) information from the training dataset were utilized for preprocessing. During learning, data of cognitively normal (CN) individuals from a single center were utilized, and data of CN individuals and Alzheimer disease (AD) patients enrolled in multiple centers were considered the test set.

Results

The preprocessing based on the preindicator (dice similarity coefficient [DSC], 0.8567) resulted in a better performance than without (DSC, 0.7921). The standard deviation (SD) of the WM region intensity (DSC, 0.8303) had a more substantial influence on the performance than the average intensity (DSC, 0.6591). When the SD of the test data WM intensity was smaller than the learning data, the performance improved (0.03 increase in lower SD, 0.05 decrease in higher SD). Furthermore, preindicator-based pretreatment increased the correlation of mean cortical thickness of the entire gray matter between Atroscan and FreeSurfer, and data augmentation without preprocessing did not.Both preindicator processing and data augmentation improved the correlation coefficient from 0.7584 to 0.8165.

Conclusion

Data augmentation and preindicator-based preprocessing of training data can improve the performance of AI-based brain segmentation software, both increasing the generalizability and stability of brain segmentation software.

Keywords: Alzheimer disease, Artificial intelligence, Data augmentation, Segmentation

Encephalitis

Jong-Hyeok Park et al. Performance of brain segmentation

Introduction

Cortical thinning and regional atrophy of the brain is associated with various health conditions, such as long-standing hypertension [1], diabetes [2], obesity [3], neurodegenerative disease, and normal aging [4]. Neuroinflammation has been suggested as a plausible mechanism of cortical thinning [4].

Quantitative characterization of the brain in neurodegenerative diseases, such as Alzheimer disease (AD), has been performed using cortical and hippocampal segmentation and visualized on magnetic resonance imaging (MRI). For brain imaging, statistical parametric mapping and FreeSurfer have been widely used [5]. Recently, image data with processing techniques and deep learning methods have been used for research in various ways. Studies based on convolutional neural networks (CNNs) have mainly focused on deep learning using medical images [6–8]. These methods perform well in detecting brain tumors and stroke lesion segmentation [9,10]. In dementia, factors such as segmentation, gray matter (GM) tissue maps, and cortical thickness, have high predictive value in AD. Therefore, determining how segmentation can be improved is critical to achieving accurate clinical decisions. Deep learning analysis of MRI can potentially assist in detecting biomarkers before the appearance of dementia symptoms which can help clinicians determine a confirmatory diagnosis. In addition, image biomarkers can be utilized for disease assessment, differential diagnosis, monitoring, and treatment [11]. However, obtaining a sufficient number of medical images for artificial intelligence (AI) learning is difficult. Furthermore, the AI models may have low performance if the images utilized for learning are acquired using different procedures and scanners at multiple institutions with insufficient variability in resolution, noise, and tissue appearance [12,13].

To address these issues, images should be standardized with preprocessing methods, such as brain registration, warping, and voxel-based morphometry [14–16]. In addition, data augmentation, which increases the sample responsiveness of a model, can be applied [17–19]. However, verifying whether an AI model exhibits satisfactory performance before implementing these methods is difficult. Therefore, we developed a system that complements model performance using preliminary indicators. The preindex system used in this study includes GM as the detection object and white matter (WM) as the preindicator.

In our preliminary indicator hypothesis, signals acquired in medical images have characteristic information based on devices and acquisition methods. Data derived from a specific area have a higher signal-to-noise ratio (SNR) than that obtained from the entire image. The performance of AI models generally decreases when external data are used without training. Furthermore, converting the external data based on the characteristics of the training data improves performance.

In the present study, brain WM was used as a preindicator because it occupies an extensive area in the brain compared with the cortex or other specific areas. Furthermore, the intensity of this area appears relatively uniform on MRI scans. Because this area is large, even if some portions of the WM are not detected, or other small areas are falsely detected, the effect on the average and standard deviation (SD) of all the divided pixels would be small. Preindicator-based preprocessing and data augmentation were utilized to improve model-based segmentation processes and improve software-based clinical evaluations and diagnostic efficacy.

Methods

This study was performed in accordance with the Declaration of Helsinki. The study was approved by the Institutional Review Board of Seoul National University Hospital (No. 1712-038-905) and written informed consent was waived due to its retrospective nature.

Dataset preparation

Alzheimer's Disease Neuroimaging Initiative
Data used in these analyses were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI/Data/index.shtml, adni.loni.usc.edu), a publicly available database launched in 2003 as a public-private partnership. MRI T1 sagittal data (n = 480) were collected. The data were labeled during the data acquisition process. Subjects were divided into cognitively normal

Table 1 The demographic data for the training dataset

Variable	Male (n)	Female (n)	Age (yr)
ADNI (model evaluation data)			
AD	144	132	76.63 ± 5.84
CN	122	82	77.51 ± 5.71
One institute (model training data)			
CN	137	164	63.51 ± 9.50

ADNI, Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu); AD, Alzheimer disease; CN, cognitively normal.

Jong-Hyeok Park et al. Performance of brain segmentation

𝑒ncephalitis

(CN, n = 204) and AD (n = 276) groups. The demographics of the study population are shown in Table 1.

Local medical check-up center

Another dataset of MRIs from 301 individuals (male, 137 and female, 164) was randomly selected from a healthy cohort in the Seoul National University Hospital Healthcare System Gangnam Center (IRB number, H-1712-038-905). Subjects who had undergone brain MRI including three-dimensional (3D) magnetization-prepared rapid gradient-echo sequence as part of dementia screening were considered. Whether the subjects were CN was determined with a comprehensive neuropsychological battery using the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease and clinical counseling by an experienced neurologist (Park KI) and psychiatrist (Yoon DH).

The local dataset was used for training and model selection and the ADNI data were used to evaluate the final model performance.

Data processing

Ground truth image segmentation

GM and WM in the brain were annotated using the FreeSurfer software package (version 6.0.0, available at http://surfer.nmr. mgh.harvard.edu), which was recently summarized by Fischl [20]; this software facilitates the automated parcellation of the brain cortex and subcortical structures. The GM and WM areas were automatically measured using FreeSurfer on a Linux workstation. The sagittal MRI was reformatted to an axial image and then converted to FreeSurfer mgz format. The procedure for GM and WM parcellation automatically assigns a neuroanatomical label to each voxel in an MRI volume based on probabilistic information estimated automatically from a manually labeled training set. The optimal linear transform was calculated by maximizing the likelihood of the input image. An atlas constructed from manually labeled images of 14 young and middle-aged subjects was utilized and nonlinear transformation was applied. Bayesian segmentation was performed and the maximum posteriori estimate of the labeling was determined.

Data augmentation

The factor considered during the data augmentation process was whether the image could occur. In general, for image data augmentation, several processes, such as translation, flipping, rotation, stretching, sharing, elastic deformation, and contrast augmentation were performed. However, because sharing

and elastic deformation can alter the shape of the brain, and flipping did not significantly occur, 3D translation and rotation (n = 5) and contrast augmentation (n = 5) were performed. Finally, a dataset that was 10 times the data held was obtained. Translation and 3D rotation were performed by applying randomly generated values from the normal distribution of the average of 0 and the scale of 0.1 to the transformation matrix. The contrast augmentation shifted the average of the entire image by adding randomly generated values from the average distribution of 0 and the scale of 5 to intensity values between –20 and 20.

Gray matter and white matter segmentation training

Atroscan (http://jlkgroup.com/#/platform/aihub), which provides complete segmentation of the brain cortex and subcortical structures, was used for GM and WM segmentation. The Atroscan segmentation algorithm is shown in Supplementary Figure 1. The 3D U-Net has a downsampling and an upsampling path, each with four resolution steps. In the downsampling path, each layer has three $3 \times 3 \times 3$ convolutions and batch normalization, which is followed by a rectified linear unit (ReLU) activation function, and a $2 \times 2 \times 2$ max pooling layer for downsampling. In the upsampling path, each layer uses a deconvolution with a kernel size of $2 \times 2 \times 2$, followed by three $3 \times 3 \times 3$ convolutions and batch normalization with ReLU. The network has shortcut connections between corresponding layers with the same resolution in the downsampling and upsampling paths.

For the loss function, dice similarity coefficient (DSC) was used, a key measure to determine the performance for semantic segmentation and calculated as follows:

$$DSC = \frac{2 \times \text{area of overlap}}{\text{Total combined pixels}} \quad \text{(Equation 1)}$$

For model training, the initial value of the model was set as a random variable. During learning, 100 epochs were performed and the training model that showed the highest DSC in the test set was selected as the final model. Input data for learning were downloaded from data $256 \times 256 \times 256$ to $128 \times 128 \times 128$ in size. Each data point was used for learning after rescaling to have a maximum intensity of 255 and dividing the rescaled data point by the mean and SD of the entire data.

Before data augmentation, 80% of the single-center data were randomly selected. These data were reorganized into a data set multiplied 10 times with a data augmentation process and

a data set that had not undergone augmentation. The remaining 20% of the data were used as a validation set to verify the performance of the model and applied to the collected ADNI data to evaluate whether a change in performance based on the pretreatment and data augmentation process on the WM occurred.

The DSC and correlation coefficient of the cortical thickness determined model performance. DSC evaluated model segmentation and the correlation coefficient determined model capabilities such as structural characteristics extraction.

Preprocessing methods using white matter
Brain WM was detected with a model of the same structure as the CNN model designed to segment the detection target. The average and SD of the area were measured. Accordingly, the 3D MRI data were shifted to the training mean and SD of the data. For comparison, the cortical GM, WM, and cerebrospinal fluid areas were determined.

$$x_{output} = \frac{\sum \frac{\sqrt{\frac{\sum |x_{train\_wm} - \bar{x}_{train\_wm}|^2}{n_{voxel\_train\_wm}}}}{n_{train}}}{\sqrt{\frac{\sum |x_{input} - \bar{x}_{input}|^2}{n_{voxel\_input\_wm}}}} \times (x_{input} - \bar{x}_{input\_wm}) + \bar{x}_{train\_wm}$$

$$\text{(Equation 2)}$$

where x_{input}, x_{output}, x_{train_wm}, x_{input_wm}, x_{input_wm}, $n_{voxel\_train\_wm}$, $n_{voxel\_input\_wm}$, n_{train} denote the input MRI T1 image, preprocessing image, signal intensity under WM area of training data, signal intensity under WM area of input data, voxel number under WM area of training data, and the number of training data, respectively. The 255-upper area of x_{output} was converted to 255 and the 0-under area of x_{output} was converted to 0. Figure 1 depicts the preprocessing changes in representative images.
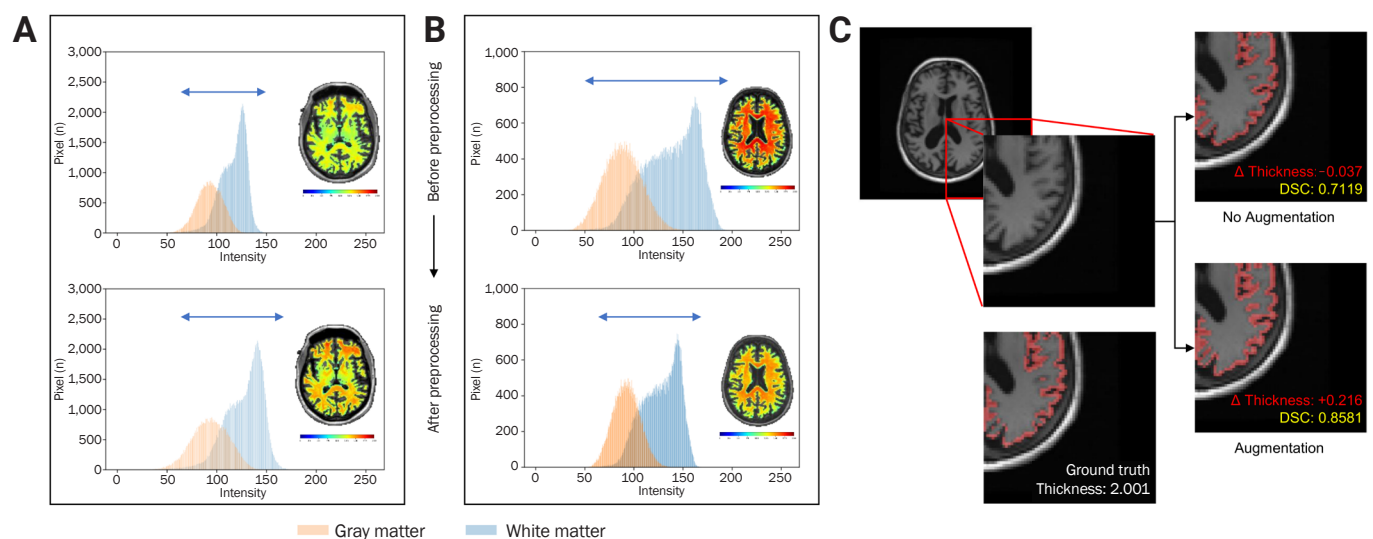
Cortical thickness measurement
In the 3D space, 173 direction vectors were calculated, representing all the direction vectors that can occur in a $6 \times 6 \times 6$ space and were measured under the assumption that cortical thickness does not exceed 6 mm. In addition, each pixel of the GM determined the minimum fulfilled direction [21]. The minimum length of each voxel eliminated the upper 90% and lower 10%. Furthermore, the minimum length group was averaged. When comparing our data with the FreeSurfer results, a Pearson correlation of 0.9623 (r-value) was obtained (Supplementary Figure 2).

Statistics
For statistical comparison, multicenter MRI data from ADNI were used. Two tests were performed to compare the performance of the preindicators. The change in DSC (ΔDSC) was evaluated with different techniques and the cortical thickness in dementia and nondementia groups was measured using each technique. A paired t-test and an independent t-test

Figure 1 Representative images depicting changes due to preprocessing based on signal intensity of white matter

In cases with a small standard deviation (SD) of signal intensity, **(A)** white matter SD increased and **(B)** decreased in cases with large SD. **(C)** Representative figure of cortical thickness measurement. Images have been color-coded for visualization purposes only in this figure.
DSC, dice similarity coefficient.

Jong-Hyeok Park et al. Performance of brain segmentation

*e*ncephalitis

were used to compare the cortical thickness between dementia and nondementia groups. In addition, Cohen's D was used to compare effect sizes based on differences between means as follows:

$$Cohe's\ D = \frac{\bar{x}_1 - \bar{x}_2}{s},\ s = \sqrt{\frac{(n_1 - 1)\,s_1^{\,2} + (n_2 - 1)\,s_2^{\,2}}{n_1 + n_2 - 2}}\quad \text{(Equation 3)}$$

where x_1, n_1, and s_1 denote the average of sample, sample number, and SD of the sample, respectively.

Results

### Model performance evaluation using dice similarity coefficient

Whether DSC, an indicator of the model performance of GM segmentation, changes with data augmentation and WM preprocessing was investigated. Model performance was lower when multicenter data were validated using a model trained on data from a single institution. However, the performance improved with data augmentation during learning and WM-based preprocessing of the test data. The average DSC increased by approximately 0.06 when both methods were implemented (Table 2). When using a test dataset of dementia patients, the model algorithm performed well although it was

trained on a CN database (Table 2). The ΔDSC after WM-based preprocessing (p < 0.001) and data augmentation (p < 0.001) was statistically significant with a t-statistic of 7.6762 after preprocessing and 6.2039 after data augmentation (paired t-test). Specifically, WM-based preprocessing enhanced the performance on test data with a lower WM intensity SD than training data. In the AD and CN groups, similar patterns were observed within each group. In the AD group, ΔDSC after treatment for lower SD (n = 164) was +0.0536, and ΔDSC for higher SD (n = 114) was –0.0863. In the CN group, ΔDSC at lower SD (n = 149) was +0.0457 and ΔDSC at higher

Table 2 Dice similarity coefficients according to preindicator case

Preprocessing	Augmentation	
	No	Yes
Cognitively normal		
No	0.7921	0.836
Yes		
White matter	0.8298	0.8567
Entire brain	0.5255	0.6989
Alzheimer disease		
No	0.7629	0.802
Yes		
White matter	0.8024	0.8293
Entire brain	0.3145	0.5276

Figure 2 The difference in DSC based on white matter preprocessing
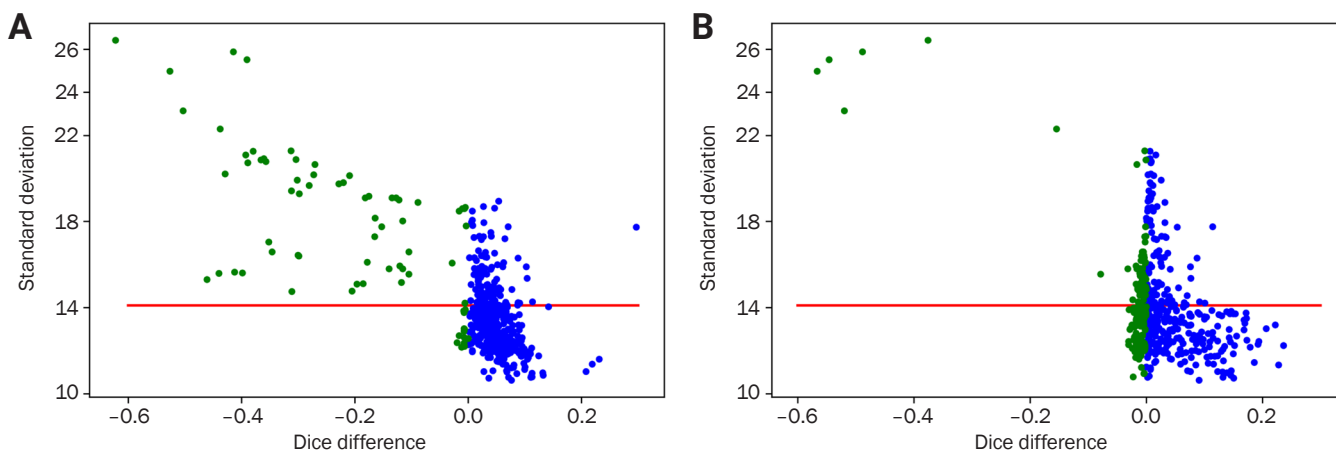
(A) Without augmentation and (B) with augmentation (red line denotes the average standard deviation [SD] of the training set, green circles indicate DSC lower than 0, indicating worse performance by preprocessing). The population with higher SD than the training set showed decreased DSC after preprocessing; augmentation complemented this effect.
DSC, dice similarity coefficient.

encephalitis

Jong-Hyeok Park et al. Performance of brain segmentation

SD (n = 55) was –0.0129. However, the differences between the groups were not statistically significant. Unlike WM-based preprocessing, data augmentation of learning data consistently increased DSC regardless of test data WM intensity SD (Figure 2). However, this pattern was not evident based on the average WM intensity of the training data (Supplementary Figure 3). Based on the above observations, we believe that adjusting SD with data augmentation, which performs intensity shift and rotation, can improve model performance. The performance with the single-center data was the highest (validation DSC, 0.9142), followed by the external multicenter CN validation (DSC, 0.8347) and AD (DSC, 0.7663) compared with pretreatment. This result indicated the internal validation performance improved from 0.9097 to 0.9142 compared with the model trained with augmentation and without SD adjustment; however, the external performance declined (DSC of CN validation, 0.8360 to 0.8347; DSC of AD validation, 0.8020 to 0.7663). In addition, model performance was analyzed using entire brain values as an indicator. Although the analysis was performed using the same process, the DSC with data augmentation was 0.6989 and 0.5255 without augmentation. Thus, considering brain WM, model performance was better.

When parcellation was taken into consideration, WM-based preprocessing did not show any meaningful positive effects. However, augmentation showed a significant increase in performance, where the values improved from 0.6475, 0.6436, 0.6383, and 0.5730 to 0.7226, 0.7075, 0.7293, and 0.6250 for the frontal, temporal, parietal, and occipital region, respectively (paired t-test, p < 0.001) (Supplementary Figure 4).

Data augmentation and white matter preprocessing improve correlation coefficient

In addition, whether data augmentation and WM preprocessing can alter the correlation coefficient that indicates model performance in the entire GM segmentation was investigated. When comparing the cortical thickness in AD and CN groups, the AD group showed significantly lower cortical thickness than the CN group in all index methods. Regarding t-values with augmentation and preprocessing, only preprocessing, only augmentation, and without augmentation and preprocessing were 16.764, 14.474, 12.206, and 13.404, respectively. Regarding Cohen's D values with augmentation and preprocessing, only preprocessing, only augmentation, and without augmentation and preprocessing were 1.737, 1.499, 1.264, and 1.389, respectively (Figure 3). However, the correlation of cortical thickness between FreeSurfer and Atroscan varied in different situations. Although data augmentation increased the segmentation DSC, the correlation of cortical thickness decreased. However, for index-based preprocessing, the correlation of cortical thickness increased (Figure 4). The highest correlation (0.8165) between FreeSurfer and Atroscan was obtained with indicator processing as well as data augmenta-

Figure 3 Cortical thickness comparison between AD and CN groups based on index methods
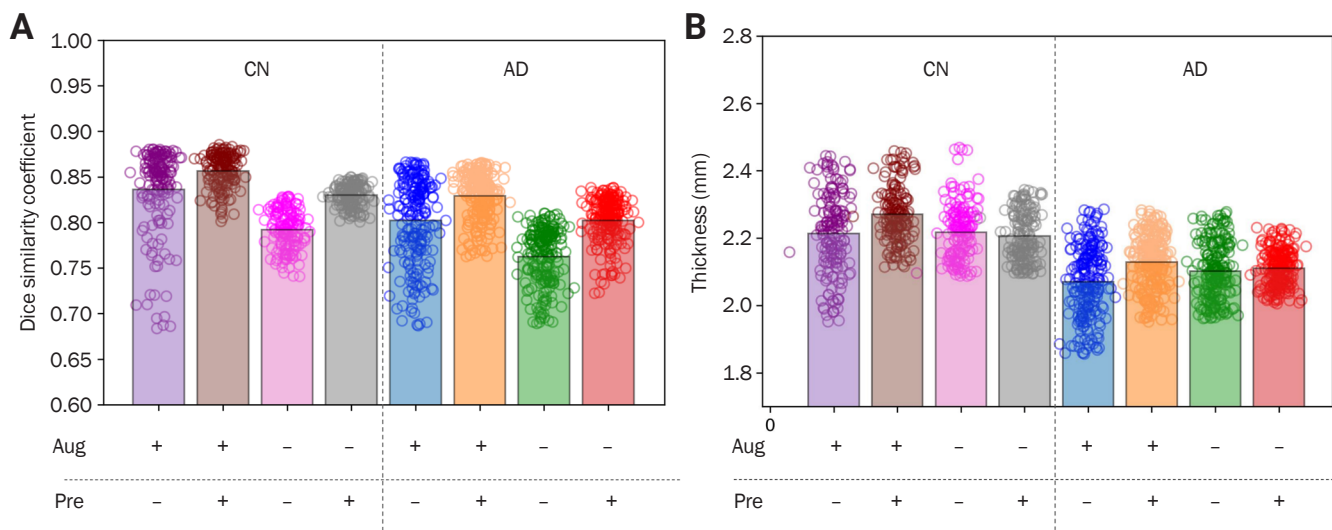
Dice similarity coefficient **(A)** and mean cortical thickness **(B)** based on different procedures.
Aug, augmentation; Pre, preprocessing; CN, cognitively normal; AD, Alzheimer disease.

Jong-Hyeok Park et al. Performance of brain segmentation

encephalitis

Figure 4 Correlation of cortical thickness obtained from the developed model with FreeSurfer results
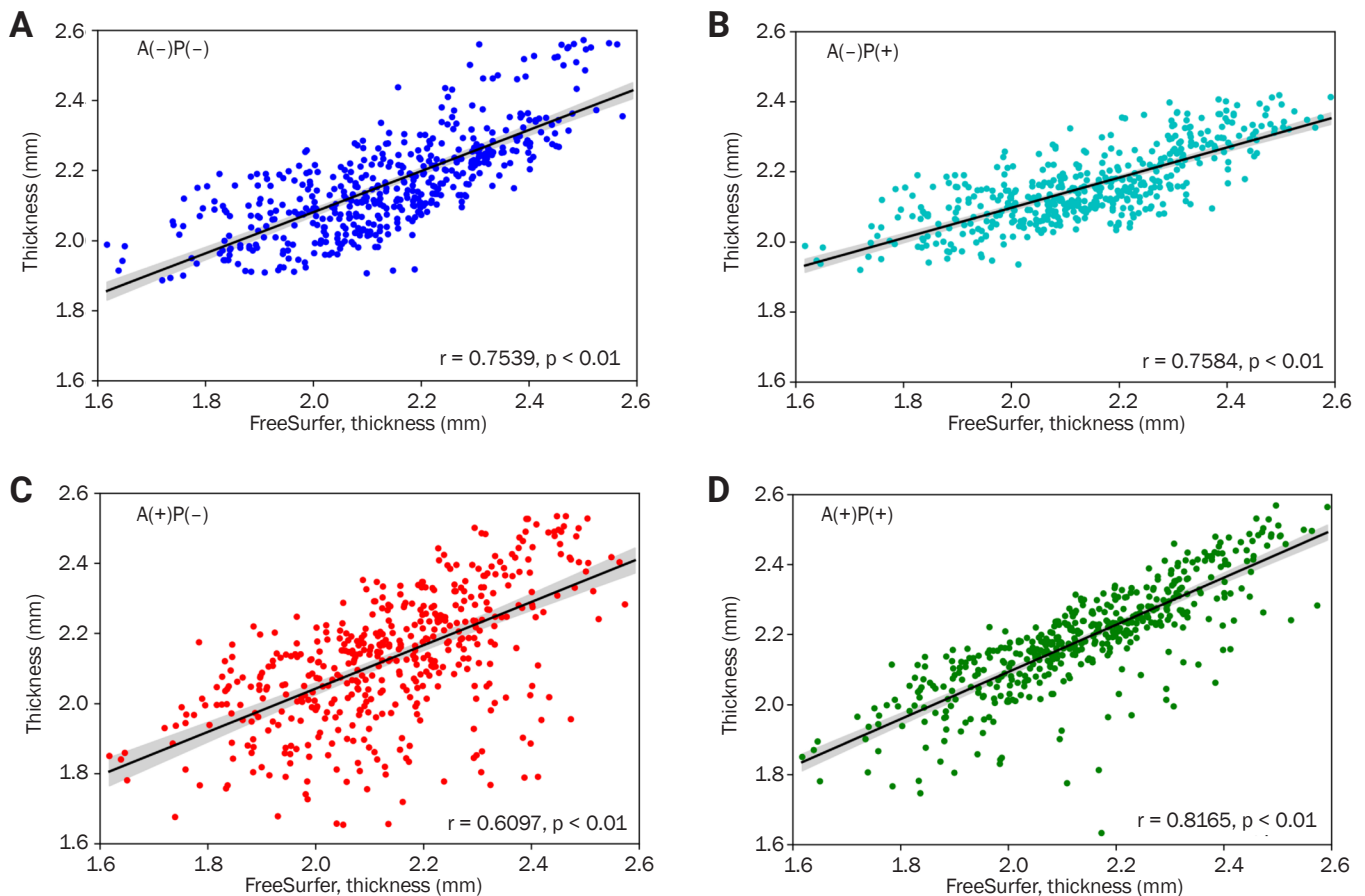
Cortical thickness was compared between Alzheimer disease and cognitively normal groups based on index methods. **(A)** Original image and no pre-processing [A(−) P(−)]. **(B)** Original image and pre-processing [A(−) P(+)]. **(C)** Augmentation and no pre-processing [A(+) P(−)]. **(D)** Augmentation and pre-processing of white matter [A(+) P(+)].
Freesurfer: version 6.0.0, available at http://surfer.nmr.mgh.harvard.edu.

tion (p < 0.001). A correlation coefficient of 0.7538 was obtained without augmentation and indicator processing (p < 0.001), and the correlation coefficient decreased to 0.6097 (p < 0.001) when only data augmentation was performed. When only indicator processing was performed without data augmentation, the value was 0.7584 (p < 0.001).

Discussion

In the present study, preprocessing using preindicators was verified to improve model performance when a deep learning model trained on data from a single site was applied to data from multiple centers. To validate our preprocessing approach, uniform subjects with normal cognitive functions and data from a single institution with limited inter-device vari-

ability for learning were used.

The performance of AI solutions in studying dementia has been extensively studied. Applications of AI in accurately diagnosing dementia and classifying it into subtypes have been compared with doctors' diagnoses [22-27]. Correlation analyses have been conducted in the segmentation area using programs such as arterial spin labeling, FreeSurfer [28], or manual segmentation [29]. The advancement of technology has increased analysis speed but generalizability can still be improved.

Data augmentation and preprocessing are widely used to improve the performance of medical image segmentation tasks. As the volume of training data decreases, this effect can be maximized [30]. In general, the state-of-the-art performance

Encephalitis

Jong-Hyeok Park et al. Performance of brain segmentation

of the top-ranked methods of tissue segmentation in the iSeg-2019 challenge showed a dice coefficient of 0.85 in GM segmentation [31]; similar DSC scores were obtained in our method (DSC, 0.8567).

To test the first hypothesis that specific image areas may have a higher SNR and better information, preprocessing was performed with different preindicators, and the model learned from data of CN individuals at a single site. The test index as an indicator was brain WM and the control was the entire brain. When the brain WM region was used as a preindicator instead of the entire brain, the performance significantly increased. Thus, WM can alleviate some differences due to multiple devices.

The second hypothesis that characteristics of the external data can be converted to fit those of the training data to improve the performance of the AI model was also tested. Preprocessing utilizing the training data with or without data augmentation improved the model performance. Comparison with the DSCs of preprocessed and non-processed training data indicated the performance improved in areas with a lower SD than the training data. The performance in areas with higher SD was lower. These issues can be resolved with data augmentation; without preprocessing of the preindicator method, better performance can be obtained when the SD of the WM of the test set is higher than the training data.

Data normalization based on the learning condition may provide improved results in a condition-dependent manner. In addition, if a task matches the characteristics of the training data, the effect is based on SD data of intensity rather than the average intensity (Supplementary Figure 3). The SD indicates the variability of the amplitude and frequency and is less influenced by the shift effect [32]. Therefore, these features can affect the segmentation performance rather than average intensity.

In contrast, a notable trend was observed when the cortical thickness correlation was studied. The difference between the AD and CN groups was significant, however, the distribution of cortical thickness was wide when data augmentation was performed. Conversely, pretreatment narrowed the distribution of cortical thickness. When data augmentation and WM-based preprocessing were used simultaneously, the highest correlation value of the entire GM thickness between FreeSurfer and our model was obtained (Figure 4). The change in correlation due to preprocessing and data augmentation was

mainly caused by segmentation because the cortical thickness measurement in the same segmentation was 0.9623 (Supplementary Figure 2). This showed that when the corresponding index was used to distinguish between the AD and CN groups, stable differences could be obtained only with data augmentation and preprocessing. Analysis of cortical thickness between the AD and CN groups showed that both the t-value and Cohen's distance were highest when data augmentation and preprocessing were performed.

In the present study, a preprocessing method based on a preindicator to achieve better performance of AI algorithms was described. This technique, in addition to data augmentation, is useful for medical data analyses where obtaining sufficient data is difficult. This method showed a significant increase in performance under specific settings, thus, insights underpinning performance improvement using data augmentation is possible. Although the preprocessing method using only mean and SD was used, subsequently, other effects using different techniques should be evaluated. Because the mean and SD are important features of MRI intensity, they play an important role in GM segmentation. However, in GM parcellation, any influence of mean or SD of intensity after preprocessing was not found probably because parcellated areas had similar intensity. Therefore, in the future, whether the preindicator method will be useful in tasks other than intensity-based segmentation can be determined. Furthermore, these analyses can promote stable measurements of brain cortical thickness and further the development of advanced methodology for generalizing brain segmentation software.

Conflicts of Interest

Park JH, Kim D, Lee MJ, and Kang S are employees of the JLK. All were blinded to group allocation. Kang SJ, Yoon DH, Lee SK, and Park KI have nothing to disclose. Additional information for correspondence and requests for data should be addressed to Park KI.

Author Contributions

Conceptualization: Park KI, Kang SJ; Formal analysis: Park JH, Park KI, Kim DM, Lee M, Kang S; Visualization: Park JH, Kim DM, Lee M, Kang S; Investigation: Park KI, Park JH, Lee SK; Resources: Park KI, Kang SJ, Yoon DH; Supervision: Park KI, Yoon DH; Writing–original draft: Park JH; Writing–review and editing: all authors

Jong-Hyeok Park et al. Performance of brain segmentation

encephalitis

Acknowledgements

Supplementary Materials

Supplementary Figures 1 to 4 can be found via https://doi.org/10.47936.2022.00108.

References

1. Alosco ML, Gunstad J, Xu X, et al. The impact of hypertension on cerebral perfusion and cortical thickness in older adults. J Am Soc Hypertens 2014;8:561–570.

2. Brundel M, van den Heuvel M, de Bresser J, Kappelle LJ, Biessels GJ; Utrecht Diabetic Encephalopathy Study Group. Cerebral cortical thickness in patients with type 2 diabetes. J Neurol Sci 2010;299:126–130.

3. Ronan L, Alexander-Bloch AF, Wagstyl K, et al. Obesity associated with increased brain age from midlife. Neurobiol Aging 2016;47:63–70.

4. Fleischman DA, Arfanakis K, Kelly JF, et al. Regional brain cortical thinning and systemic inflammation in older persons without dementia. J Am Geriatr Soc 2010;58:1823–1825.

5. Khedher L, Ramírez J, Górriz JM, Brahim A, Segovia F; Alzheimer's Disease Neuroimaging Initiative. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. Neurocomputing 2015;151(Part 1):139–150.

6. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical image computing and computer-assisted intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer; Cham: 2015. p. 234–241.

7. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. Presented at 2016 Fourth International Conference on 3D vision (3DV); October 25-28, 2016. Piscataway, NJ: IEEE; 2016. p. 565–571.

8. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, editors. Medical image computing and computer-assisted intervention – MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer; Cham: p. 424–432.

9. Kamnitsas K, Ledig C, Newcombe VF, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 2017;36:61–78.

10. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal 2017;35:18–31.

11. Ledig C, Schuh A, Guerrero R, Heckemann RA, Rueckert D. Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. Sci Rep 2018;8:11258.

12. Domingos P. A few useful things to know about machine learning. Commun ACM 2012;55:78–87.

13. Leung KK, Clarkson MJ, Bartlett JW, et al. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. Neuroimage 2010;50:516–523.

14. Lemieux L, Wieshmann UC, Moran NF, Fish DR, Shorvon SD. The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization. Med Image Anal 1998;2:227–242.

15. Ghanei A, Soltanian-Zadeh H, Jacobs MA, Patel S. Boundary-based warping of brain MR images. J Magn Reson Imaging 2000;12:417–429.

16. Ashburner J, Friston KJ. Voxel-based morphometry: the methods.

Encephalitis

Jong-Hyeok Park et al. Performance of brain segmentation

Neuroimage 2000;11(6 Pt 1):805–821.

17. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 2016;35:1240–1251.

18. Roth HR, Lu L, Farag A, et al. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical image computing and computer-assisted intervention -- MICCAI 2015. Lecture Notes in Computer Science, vol 9349. Springer; Cham: 2015. p. 556–564.

19. Zhao A, Balakrishnan G, Durand F, Guttag JV, Dalca AV. Data augmentation using learned transformations for one-shot medical image segmentation. Presented at 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 15-20, 2019. Piscataway, NJ: IEEE; 2019. p. 8535–8545.

20. Fischl B. FreeSurfer. Neuroimage 2012;62:774–781.

21. Aganj I, Sapiro G, Parikshak N, Madsen SK, Thompson PM. Measurement of cortical thickness from MRI by minimum line integrals on soft-classified tissue. Hum Brain Mapp 2009;30:3188–3199.

22. Alkabawi EM, Hilal AR, Basir OA. Computer-aided classification of multi-types of dementia via convolutional neural networks. Presented at 2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA); May 7-10, 2017. Piscataway, NJ: IEEE; 2017. p. 45–50.

23. Akhila JA, Markose C, Aneesh RP. Feature extraction and classification of dementia with neural network. Presented at 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT); July 6-7, 2017. Piscataway, NJ: IEEE; 2017. p. 1446–1450.

24. Dolph CV, Alam M, Shboul Z, Samad MD, Iftekharuddin KM. Deep learning of texture and structural features for multiclass Alzheimer's disease classification. Presented at 2017 International Joint Conference on Neural Networks (IJCNN); May 14-19, 2017. Piscataway, NJ:

IEEE; 2017. p. 2259–2266.

25. Faturrahman M, Wasito I, Hanifah N, Mufidah R. Structural MRI classification for Alzheimer's disease detection using deep belief network. Presented at 2017 11th International Conference on Information & Communication Technology and System (ICTS); October 31, 2017. Piscataway, NJ: IEEE; 2017. p. 37–42.

26. Suk HI, Lee SW, Shen D; Alzheimer's Disease Neuroimaging Initiative. Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal 2017;37:101–113.

27. Islam J, Zhang Y. A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI Data. In: Zeng Y, He Y, Kotaleski JH, , editors. International Conference on Brain Informatics-BI 2017. Lecture Notes in Computer Science, vol 10654. Springer; Cham: 2017. p. 213–222.

28. Rebsamen M, Rummel C, Reyes M, Wiest R, McKinley R. Direct cortical thickness estimation using deep learning-based anatomy segmentation and cortex parcellation. Hum Brain Mapp 2020;41:4804–4814.

29. Kumar P, Nagar P, Arora C, Gupta A. U-segnet: fully convolutional neural network based automated brain tissue segmentation tool. Presented at 2018 25th IEEE International Conference on Image Processing (ICIP); October 7-10, 2018. Piscataway, NJ: IEEE; 2018. p. 3503–3507.

30. Thyreau B, Taki Y. Learning a cortical parcellation of the brain robust to the MRI segmentation with convolutional neural networks. Med Image Anal 2020;61:101639.

31. Sun Y, Gao K, Wu Z, et al. Multi-site infant brain segmentation algorithms: the iSeg-2019 challenge. IEEE Trans Med Imaging 2021;40:1363–1376.

32. Azami H, Sanei S, Mohammadi K. A novel signal segmentation method based on standard deviation and variable threshold. Int J Comput Appl 2011;34:27–34.