Dement Neurocogn Disord. 2025 Apr;24(2):115-125
https://doi.org/10.12779/dnd.2025.24.2.115
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Original Article

Check for updates

Quantifying Brain Atrophy Using a CSF-Focused Segmentation Approach

Kyoung Yoon Lim,[1] **Seongbeom Park**,[1] **Duk L. Na**,[1] **Sang Won Seo**,[2,3,4,5] **Min Young Chun**,[6,7] **Kichang Kwak**,[1] **on behalf of the K-ROAD study and ADNI**

[1]BeauBrain Healthcare, Inc., Seoul, Korea
[2]Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
[3]Alzheimer's Disease Convergence Research Center, Samsung Medical Center, Seoul, Korea
[4]Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Korea
[5]Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Korea
[6]Department of Neurology, Yonsei University College of Medicine, Seoul, Korea
[7]Depeartment of Neurology, Yongin Severance Hospital, Yonsei University Health System, Yongin, Korea

OPEN ACCESS

Received: Feb 2, 2025
Revised: Mar 10, 2025
Accepted: Mar 19, 2025
Published online: Apr 9, 2025

Correspondence to
Min Young Chun
Department of Neurology, Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin 16995, Korea.
Email: myc5198@gmail.com

Kichang Kwak
BeauBrain Healthcare, Inc., 314 Hakdong-ro, Gangnam-gu, Seoul 06098, Korea.
Email: kichang.kwak@beaubrain.bio

ORCID iDs
Kyoung Yoon Lim
https://orcid.org/0009-0006-5239-6424
Seongbeom Park
https://orcid.org/0000-0002-0759-6826
Duk L. Na
https://orcid.org/0000-0002-0098-7592
Sang Won Seo
https://orcid.org/0000-0002-8747-0122
Min Young Chun
https://orcid.org/0000-0003-3731-6132

ABSTRACT

Background and Purpose: Brain atrophy, characterized by sulcal widening and ventricular enlargement, is a hallmark of neurodegenerative diseases such as Alzheimer's disease. Visual assessments are subjective and variable, while automated methods struggle with subtle intensity differences and standardization, highlighting limitations in both approaches. This study aimed to develop and evaluate a novel method focusing on cerebrospinal fluid (CSF) regions by assessing segmentation accuracy, detecting stage-specific atrophy patterns, and testing generalizability to unstandardized datasets.
Methods: We utilized T1-weighted magnetic resonance imaging data from 3,315 participants from Samsung Medical Center and 1,439 participants from other hospitals. Segmentation accuracy was evaluated using the Dice similarity coefficient (DSC), and W-scores were calculated for each region of interest (ROI) to assess stage-specific atrophy patterns.
Results: The segmentation demonstrated high accuracy, with average DSC values exceeding 0.9 for ventricular and hippocampal regions and above 0.8 for cortical regions. Significant differences in W-scores were observed across cognitive stages (cognitively unimpaired, mild cognitive impairment, dementia of Alzheimer's type) for all ROIs (all, $p<0.05$). Similar trends were observed in the images from other hospitals, confirming the algorithm's generalizability to datasets without prior standardization.
Conclusions: This study demonstrates the robustness and clinical applicability of a novel CSF-focused segmentation method for assessing brain atrophy. The method provides a scalable and objective framework for evaluating structural changes across cognitive stages and holds potential for broader application in neurodegenerative disease research and clinical practice.

Keywords: Alzheimer Disease; Neurodegenerative Diseases; Atrophy; Cerebrospinal Fluid; Deep Learning; Magnetic Resonance Imaging

Kichang Kwak
https://orcid.org/0000-0003-4542-0001

Conflict of Interest
Lim KY, Park S, Na DL, and Kwak K were employed by BeauBrain Healthcare, Inc. The remaining authors have no potential conflicts of interest to disclose.

Author Contributions
Conceptualization: Lim KY, Park S, Na DL, Seo SW, Chun MY, Kwak K; Data curation: Lim KY, Park S, Na DL, Seo SW, Chun MY, Kwak K; Formal analysis: Lim KY, Park S, Kwak K; Investigation: Lim KY, Park S, Na DL, Seo SW, Chun MY, Kwak K; Methodology: Lim KY, Park S, Na DL, Seo SW, Chun MY, Kwak K; Software: Lim KY, Park S, Kwak K; Validation: Lim KY, Park S, Kwak K; Writing - original draft: Lim KY, Park S, Na DL, Seo SW, Chun MY, Kwak K; Writing - review & editing: Lim KY, Park S, Na DL, Seo SW, Chun MY, Kwak K.

INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia and is characterized by progressive cognitive decline and structural changes in the brain.[1] One of the hallmark features of AD is brain atrophy, which results from the loss of neurons and synapses,[2] particularly in regions such as the medial temporal lobe and association cortices.[2] This neuronal loss leads to gyral shrinkage and sulcal widening, structural changes that are closely linked to disease progression and severity.[3] Quantifying brain atrophy has therefore become a critical aspect of diagnosing and monitoring AD, as it provides valuable insights into the extent of neurodegeneration and its clinical implications.[4]

Clinicians have traditionally evaluated brain atrophy by visually inspecting neuroimaging for evidence of sulcal widening and ventricular enlargements.[3,5,6] While this approach aligns with the structural changes observed in AD, it is inherently subjective and prone to inter-rater and intra-rater variability, leading to inconsistencies in assessment.[7] To address these challenges, visual assessment protocols were developed to standardize evaluations of brain atrophy.[5,8] However, reliability issues persist, particularly when evaluating subtle changes or when expertise levels vary among raters.[9] These limitations underscore the need for objective and reproducible methods to assess brain atrophy more consistently across clinical and research settings.

Existing automated methods, such as cortical thickness measurements[10,12] and voxel-based morphometry,[13] aim to provide objective assessments by segmenting the boundary between gray matter and white matter. However, these methods often face challenges due to the subtle intensity differences at this interface, resulting in variability across imaging protocols and vendors.[14] Such methods also require extensive preprocessing and standardization, which can be time-consuming and limit their utility in real-world clinical environments. Moreover, these approaches may not perform well on lower-resolution imaging modalities, such as 2-dimensional (2D) T1 magnetic resonance imaging (MRI) or computed tomography (CT), further complicating their applicability in routine clinical practice.

In this study, we developed a novel method focusing on extracerebral cerebrospinal fluid (CSF) regions that reflect sulcal widening and ventricular volume. Specifically, we aimed to 1) evaluate the segmentation accuracy across all regions of interest (ROIs) by comparing it to silver standard, 2) assess the algorithm's ability to detect stage-specific atrophy patterns across cognitive stages (cognitively unimpaired [CU], mild cognitive impairment [MCI], and dementia of Alzheimer's type [DAT]), and 3) determine its generalizability by testing its performance on datasets without prior standardization. This method offers a practical and scalable solution for assessing brain atrophy, with potential applications in both clinical and research settings to improve the diagnosis and monitoring of neurodegenerative diseases.

METHODS

Participants

A total of 4,754 participants were recruited from the Korea-Registries to Overcome dementia and Accelerate Dementia Research (K-ROAD) project.[15] K-ROAD is a nationwide multicenter cohort study conducted in South Korea, involving Samsung Medical Center (SMC) and 16 other hospitals, designed to advance the understanding and diagnosis of AD and related dementias. It is a member of the worldwide Alzheimer's Disease Neuroimaging Initiative (ADNI).[16]

Among the hospitals utilizing diverse protocols (e.g., vendor, scanner type, voxel size), SMC, which provided the largest number of participants' magnetic resonance (MR) images, along with the other hospitals, were separated into independent cohorts. The participants of both cohorts comprised individuals with CU, MCI, and DAT. CU individuals demonstrated no subjective cognitive complaints or functional impairments, with cognitive performance confirmed to be within normal limits through detailed neuropsychological assessments. MCI was diagnosed based on the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria,[17] characterized by measurable cognitive decline in one or more domains without significant interference in daily functional activities. DAT was diagnosed following the NIA-AA guidelines, requiring evidence of significant cognitive decline, including memory impairment, that interfered with independence in daily life and was consistent with an Alzheimer's disease etiology. The segmentation model was trained using MR images from 1,889 subjects, supplemented by public dataset, to delineate CSF in 14 brain regions. The W-score model, developed to define normal CSF volumes and quantify atrophy, used 3D T1 images from 936 cognitively normal subjects, independent of the 3,315 recruits. Finally, to analyze the brain atrophy patterns across each group of participants within 2 independent cohorts, we included 3,315 participants from SMC (Cohort 1) and 1,439 participants from other hospitals (Cohort 2).

We excluded participants who had any of the following conditions: 1) white matter hyperintensities due to radiation injury, multiple sclerosis, vasculitis, leukodystrophy or metabolic disorders; 2) traumatic brain injury; 3) territorial infarction; 4) brain tumor; and 5) rapidly progressive dementia.

The study protocol received approval from the Institutional Review Board (IRB) of SMC (IRB No. 2021-02-135). Written informed consent was obtained from each participant and all procedures were conducted in accordance with the approved guidelines.

Acquisition of 3D T1 images

To acquire 3D T1 turbo field-echo MRI scans from all participants at SMC, a 3.0 T MRI scanner (Philips 3.0 T Achieva; Philips Healthcare, Andover, MA, USA) was used with following parameters: sagittal slice thickness of 1.0 mm with 50% overlap, repletion time of 9.9 ms, echo time of 4.6 ms, and matrix size of 240×240 pixels reconstructed to 480×480 over a field of view of 240 mm. The characteristics of MRI for all hospitals are shown in **Supplementary Table 1**.

Segmentation methods for regions of CSF

Overview of proposed pipeline is illustrated in **Fig. 1**. The pipeline for modeling the algorithm to compute the CSF volume and the W-scores for each ROI volume consists of a 2-step process. In the first step, a segmentation model is trained to compute each ROI volume. A 1 mm isotropic voxel resampled 3D MR images and corresponding label images were derived from the original 3D T1-weighted raw image using the SynthSeg[18] function in FreeSurfer version 7.4.2. The CSF ROIs were then extracted from these label maps to create silver standard. The preprocessed images along with the silver standard were fed into a deep learning-based segmentation model for training.

A segmentation model was developed to delineate 14 ROIs using convolutional neural network-based deep learning approach. The defined ROIs included the CSF adjacent to the gray matter within the lobe regions, specifically, the left and right of frontal (L_Frontal, R_Frontal), occipital (L_Occipital, R_Occipital), parietal (L_Parietal, R_Parietal), and

https://doi.org/10.12779/dnd.2025.24.2.115

DND Dementia and Neurocognitive Disorder

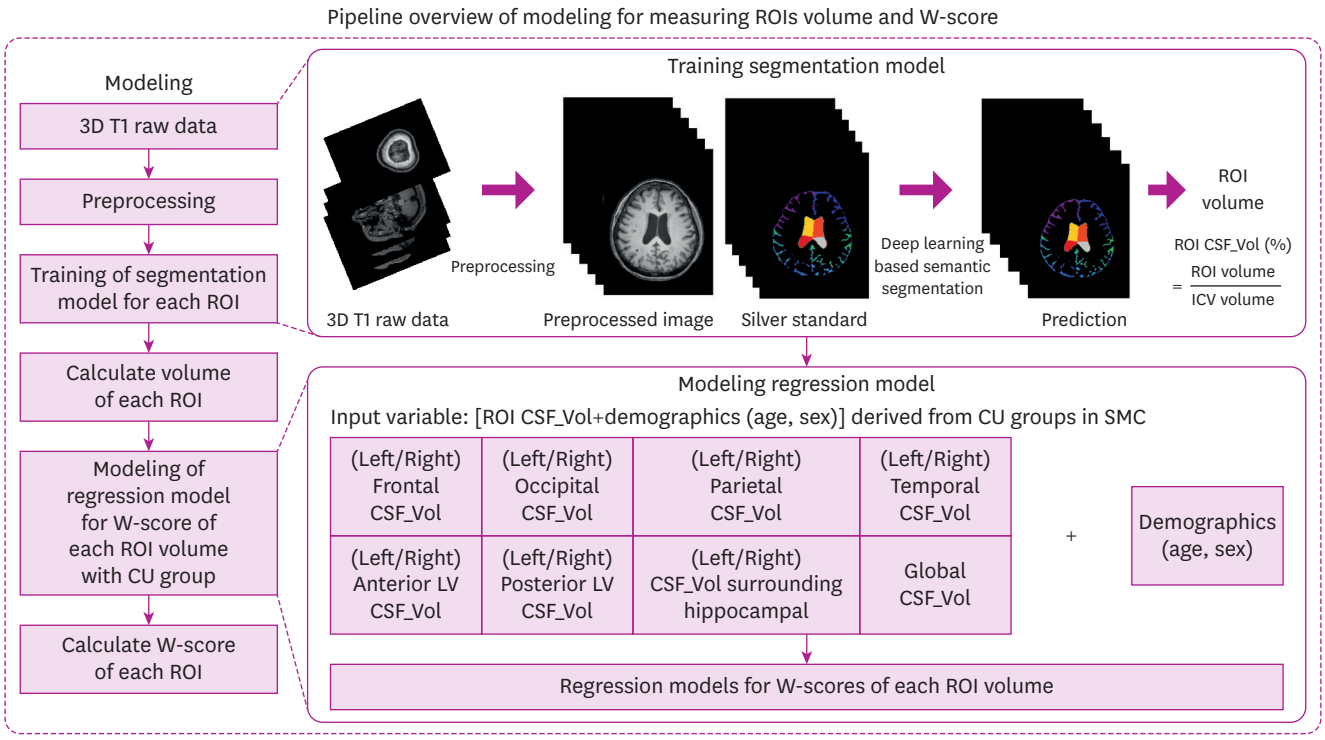

Fig. 1. Overview of the proposed pipeline in this study. The figure illustrates the algorithmic modeling process for measuring the volume and W-scores of each ROI. The pipeline consists of 2 main steps: segmentation, which calculates the volume, and regression, which calculates the W-scores for each ROI. In the segmentation step, the model is trained using preprocessed 3D T1-weighted images and silver standard for 14 ROIs and intracranial volume. The regression model is designed to calculate the W-scores using the volume obtained from the segmentation step and the demographic information of the CU group. The global volume is defined as the sum of the CSF_Vols of the 14 ROIs.
ROI: region of interest, 3D: 3-dimensional, CU: cognitively unimpaired, CSF_Vol: cerebrospinal fluid volume, LV: lateral ventricle.

temporal (L_Temporal, R_Temporal) lobes. Additionally, ventricular spaces were considered, including the left and right of anterior lateral ventricle (LV) (L_Anterior_LV, R_Anterior_LV), posterior LV (L_Posterior_LV, R_Posterior_LV), and CSF space surrounding the left and right of hippocampal regions (L_Hippocampal, R_Hippocampal).

MR images used for training and testing were obtained from the Cohort 1, ADNI, International Consortium for Brain Mapping,[19] Information eXtraction from Images, and Open Access Series of Imaging Studies.[20] Further details are provided in **Supplementary Table 2**. We used 2D nnUNet[21] as a segmentation model and performed 5-fold cross validation. The preprocessed images were further processed using the default preprocessor of nnUNet and were trained according to the following implementation details: leaky rectified linear unit as the activation function, loss function combining Dice loss and cross-entropy loss, stochastic gradient descent as the optimizer, a learning rate of 1e-2, a weight decay of 3e-5, 200 epochs, a batch size of 64. The evaluation metric was Dice similarity coefficient (DSC) which calculated overlap between silver standard and prediction as follows:

$$\mathrm{DSC} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

where TP, FP, and FN represent true positives, false positives, and false negatives respectively. The volumes of each ROI were calculated by summing the voxel values within the predicted ROIs, with each voxel representing the presence or absence of a target region.

https://doi.org/10.12779/dnd.2025.24.2.115

Regression for W-score of each region of CSF

W-score[22] is a statistical metric that can be adjusted for specific covariates. In this study, the W-score was derived from the CSF volume predicted from a multiple linear regression model with age and sex as covariates. The regression model, referred to as W-score model, was developed for each ROI volume, using age and sex as covariates for 936 subjects in CU group from SMC in the second step (**Fig. 1**). The W-score is calculated as follows:

$$W - score_{ROI} = \frac{E_{W\_ROI}(A,S) - V_{ROI}}{\sigma_{W\_ROI}}$$

where $E_{W\_ROI}(A,S)$ is expected CSF_Vol of ROI in W-score model with Age (A) and Sex (S), V_{ROI} is the participant's CSF_Vol of ROI derived from segmentation model, and σ_{W_ROI} is the standard deviation of the residuals in W-score model.

Statistical analysis

All statistical analyses were performed using R software, version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org). χ^2 test was applied to evaluate differences in categorical variables. Group differences in continuous variable in continuous variables were also analyzed with an analysis of variance (ANOVA). ANOVA was performed, followed by post-hoc tests when relevant. The distribution pattern of the W-scores for each ROI across the groups was visualized and analyzed using boxplots. Subsequently, Cohen's d was calculated to assess the effect size of differences between the groups for each ROI.

RESULTS

Participant characteristics

The demographic and clinical characteristics of participants in the Cohort 1 and Cohort 2 are presented in **Table 1**. In the Cohort 1 (n=3,315), the mean age was 71.7±7.8 years, with 58.6% female participants. The average years of education were 11.8±4.8, and the mean Mini-Mental State Examination (MMSE) score was 24.7±4.9. The mean intracranial volume (ICV) was 1,483.7±133.6 mL. In the Cohort 2 (n=1,439), participants had a mean age of 72.3±7.9 years, with 64.1% being female. The average years of education were 9.3±4.8, and the mean MMSE score was 24.0±4.7. The mean ICV was 1,475.6±164.2 mL.

Table 1. Demographics of participants in 2 cohorts

Characteristics	Samsung Medical Center (Cohort 1)				Other hospitals (Cohort 2)			
	Total	CU	MCI	DAT	Total	CU	MCI	DAT
No. of subjects	3,315	900	1,559	856	1,439	414	821	204
Age (yr)	71.7±7.8	71.2±7.2	72.0±7.5	71.8±8.8	72.3±7.9	68.7±7.7	73.7*±7.4	73.9*±8.4
Sex, female	1,942 (58.6%)	557 (61.9%)	868 (55.7%)	517 (60.4%)	922 (64.1%)	267 (64.5%)	533 (64.9%)	122 (59.8%)
Education (yr)	11.8±4.8	11.8±4.7	12.1±4.7	11.2*†±4.9	9.3±4.8	10.4±4.6	8.9*±4.8	8.8*±5.1
MMSE score	24.7±4.9	28.1*±2.0	25.7*±3.2	19.4*†±5.1	24.0±4.7	27.2±2.4	23.8*±4.1	18.9*†±5.3
ICV (mL)	1,483.7±133.6	1,484.2±130.9	1,492.0±136.0	1,467.8*†±130.6	1,475.6±164.2	1,482.2±130.7	1,469.7±143.8	1,486.2±269.3

The sex value expressed as number of female (%) and the other value expressed as mean ± standard deviation. The distributions of age, education, MMSE, and ICV were analyzed using analysis of variance and Tukey's post hoc test for statistical significance between groups. The sex distribution was tested for significance using a χ^2 test.
CU: cognitively unimpaired, MCI: mild cognitive impairment, DAT: dementia of Alzheimer's type, MMSE: Mini-Mental State Examination, ICV: intracranial volume.
*Significant difference with CU group, †Significant difference with MCI group.

https://doi.org/10.12779/dnd.2025.24.2.115

DND Dementia and Neurocognitive Disorder

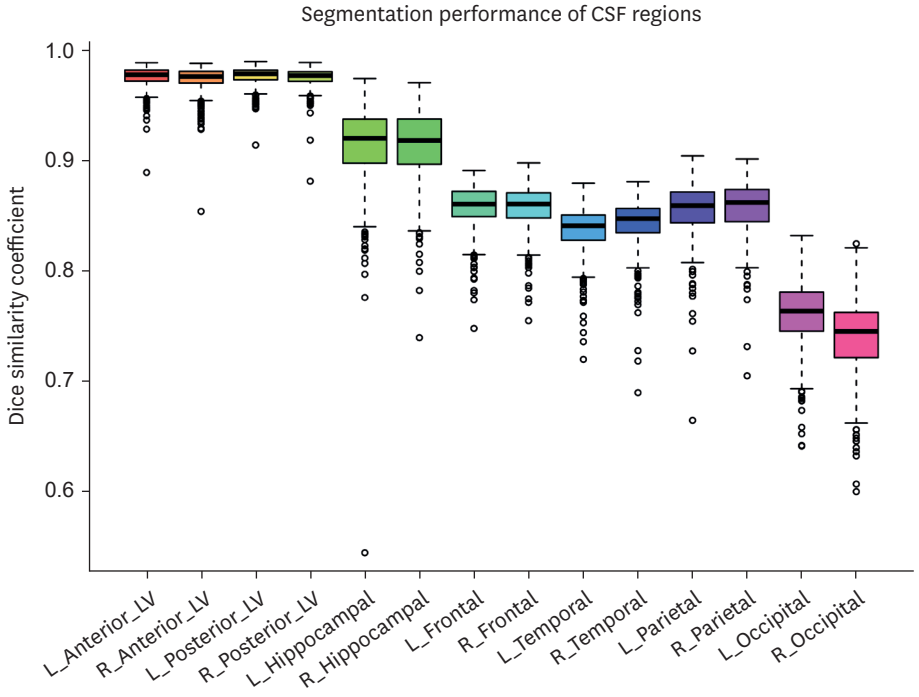

Fig. 2. Segmentation performance of proposed model across ROIs. The figure shows the segmentation performance of the proposed model, evaluated using the DSC by comparing model outputs to silver standard. ROI: region of interest, DSC: Dice similarity coefficient, CSF: cerebrospinal fluid, L_Anterior_LV: left anterior lateral ventricle, R_Anterior_LV: right anterior lateral ventricle, L_Posterior_LV: left posterior lateral ventricle, R_Posterior_LV: right posterior lateral ventricle, L_Hippocampal: left hippocampal, R_Hippocampla: right hippocampal, L_Frontal: left frontal, R_Frontal: right frontal, L_Temporal: left temporal, R_Temporal: right temporal, L_Parietal: left parietal, R_Parietal: right parietal, L_Occipital: left occipital, R_Occipital: right occipital.

Segmentation performance of each ROI

The segmentation performance for 14 ROIs and ICV was evaluated using the average DSC from 5-fold cross-validation (**Fig. 2**). The ROIs surrounding the LVs achieved the highest DSC values, with 0.976 for L_Anterior_LV, 0.974 for R_Anterior_LV, 0.977 for L_Posterior_LV, and 0.975 for R_Posterior_LV. The hippocampal regions also showed high DSCs of 0.914 for both L_Hippocampal and R_Hippocampal. Among the cortical ROIs, the frontal, parietal, and temporal regions had DSCs above 0.8, including 0.859 for L_Frontal, 0.858 for R_Frontal, 0.856 for L_Parietal, 0.858 for R_Parietal, 0.838 for L_Temporal, and 0.844 for R_Temporal. In contrast, the occipital regions displayed slightly lower performance, with DSCs of 0.761 for L_Occipital and 0.740 for R_Occipital. Representative examples of the segmentation results can be found in **Supplementary Fig. 1**.

Regional variations in W-scores among CU, MCI, and DAT

Fig. 3 shows the distribution and statistical significance of W-scores for each ROI across the CU, MCI, and DAT groups in the Cohort 1 (**Fig. 3A**) and Cohort 2 (**Fig. 3B**). In both cohorts, all ROIs showed significant differences in W-scores between CU and both MCI and DAT groups (all, $p<0.05$), indicating consistent patterns of differentiation across the 2 cohorts. Significant differences in W-scores between MCI and DAT groups were observed across all ROIs in the Cohort 1 ($p<0.05$). In the Cohort 2, similar differences were found, except in the left frontal, parietal and occipital ROIs.

https://doi.org/10.12779/dnd.2025.24.2.115

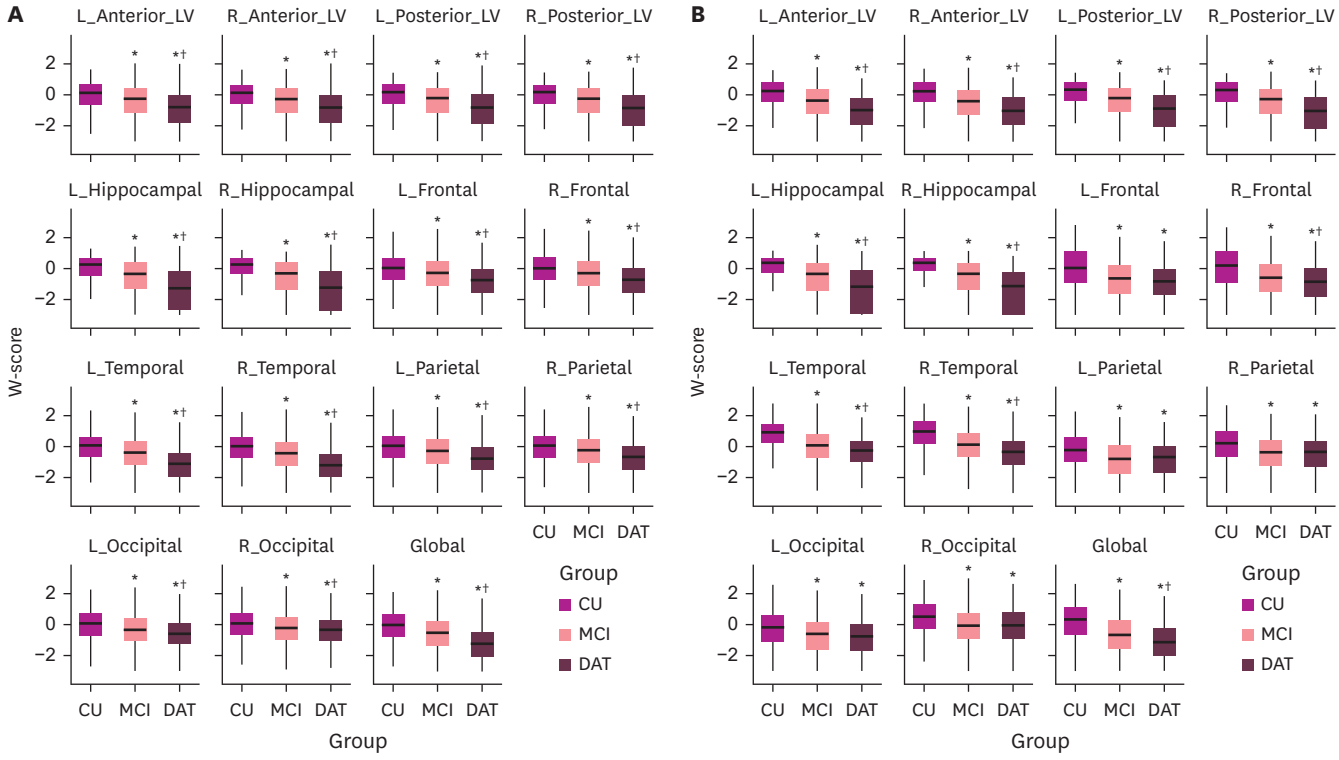

Fig. 3. Box plots of W-scores for CSF volume of each ROI in (A) Cohort 1 and (B) Cohort 2. For visual convenience, W-scores less than −3 are treated as −3, and those greater than 3 are treated as 3, but this adjustment applies only in the boxplots.
CU: cognitively unimpaired, MCI: mild cognitive impairment, DAT: dementia of Alzheimer's type, CSF: cerebrospinal fluid, L_Anterior_LV: left anterior lateral ventricle, R_Anterior_LV: right anterior lateral ventricle, L_Posterior_LV: left posterior lateral ventricle, R_Posterior_LV: right posterior lateral ventricle, L_Hippocampal: left hippocampal, R_Hippocampla: right hippocampal, L_Frontal: left frontal, R_Frontal: right frontal, L_Temporal: left temporal, R_Temporal: right temporal, L_Parietal: left parietal, R_Parietal: right parietal, L_Occipital: left occipital, R_Occipital: right occipital, Global: sum of all cerebrospinal fluid regions of interest.
*Asterisk indicates significant difference with cognitively unimpaired group and †Dagger indicates significant difference with mild cognitive impairment group.

The effect size of W-scores for each ROI, based on three pairwise comparisons, are presented in **Fig. 4** for both the Cohort 1 (**Fig. 4A**) and Cohort 2 (**Fig. 4B**). LVs, hippocampal, and temporal ROIs showed large effect size differentiating CU and DAT groups in both cohorts. Across the 2 cohorts, the effect sizes in the LV and hippocampal ROIs showed similar effect size differentiating the CU-MCI and MCI-DAT groups, whereas the effect sizes in the lobe region ROIs displayed divergent trends.

DISCUSSION

In this study, we developed an algorithm based on T1 MRI images to calculate CSF volumes and W-scores for assessing brain atrophy. Our major findings were as follows. First, the segmentation performance, evaluated through DSC, demonstrated high agreement with silver standard, indicating robust segmentation accuracy. Second, significant differences in W-scores were observed across cognitive stages (CU, MCI, and DAT) for all ROIs, reflecting the algorithm's ability to detect stage-specific atrophy patterns. Finally, similar trends were observed in the images from the Cohort 2, confirming the generalizability of our findings. Taken together, our findings suggest that this algorithm provides a reliable framework for quantifying brain atrophy and detecting stage-specific atrophy patterns across cognitive stages. Clinically, this approach has the potential to enhance the understanding and monitoring of cognitive impairment progression.

https://doi.org/10.12779/dnd.2025.24.2.115

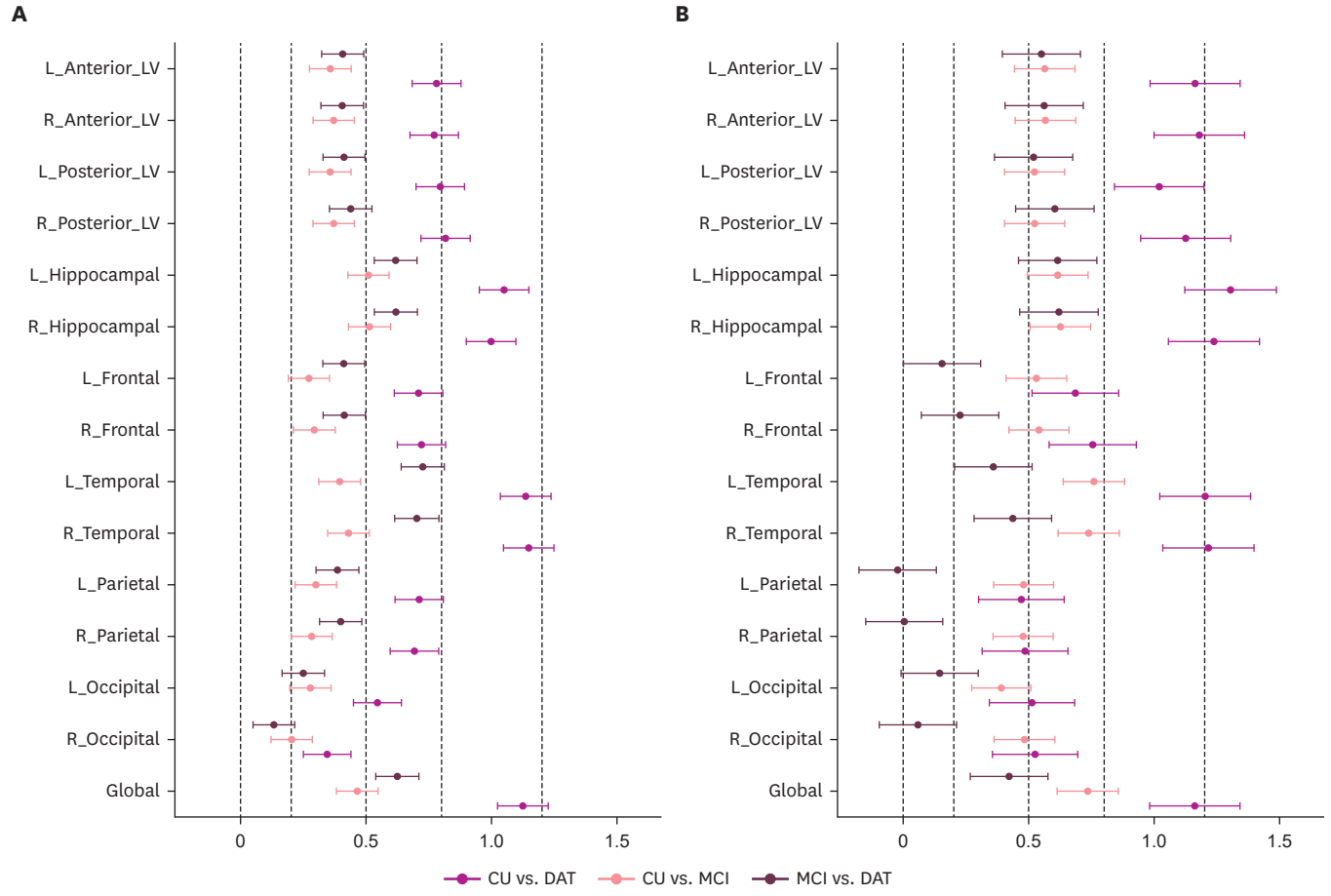

Fig. 4. Forest plots presenting the effect size on W-scores of the CSF volume of each ROI through 3 pairwise comparisons in (A) Cohort 1 and (B) Cohort 2. The dots in plot represent Cohen's d values, and the values on both sides represent 95% confidence intervals. Cohen's d values less than 0.2 are considered "very small," less than 0.5 "small," less than 0.8 "medium," less than 1.2 "large," and greater than or equal to 1.2 "very large."
CSF: cerebrospinal fluid, ROI: region of interest, CU: cognitively unimpaired, MCI: mild cognitive impairment, DAT: dementia of Alzheimer's type, L_Anterior_LV: left anterior lateral ventricle, R_Anterior_LV: right anterior lateral ventricle, L_Posterior_LV: left posterior lateral ventricle, R_Posterior_LV: right posterior lateral ventricle, L_Hippocampal: left hippocampal, R_Hippocampla: right hippocampal, L_Frontal: left frontal, R_Frontal: right frontal, L_Temporal: left temporal, R_Temporal: right temporal, L_Parietal: left parietal, R_Parietal: right parietal, L_Occipital: left occipital, R_Occipital: right occipital, Global: sum of all cerebrospinal fluid regions of interest.

Our first major finding was that the segmentation performance, evaluated through DSC, demonstrated high agreement with silver standard, highlighting the robustness of the algorithm. The highest DSC values were observed in the ventricular and hippocampal regions, exceeding 0.9, followed by cortical regions, including the frontal, temporal, and parietal lobes, which achieved DSC values above 0.8. Unlike conventional segmentation approaches[10-13] that rely on the gray matter-white matter boundary, which often shows subtle intensity differences, our method focuses on the CSF-gray matter and CSF-white matter boundaries, where intensity differences are more pronounced. This methodological advantage enhances segmentation accuracy, particularly in regions with complex geometries, such as the ventricles and hippocampal areas.

An important strength of our approach lies in its ability to provide a quantitative framework for assessing ROIs[23] routinely evaluated by clinicians in cognitive impairment. Traditional visual assessment, while commonly used, is inherently subjective and prone to inter-rater and intra-rater variability.[7] By offering precise and reproducible measurements of CSF volume

https://doi.org/10.12779/dnd.2025.24.2.115

and W-scores, our method overcomes these limitations, providing objective data critical for accurate and consistent evaluation of brain atrophy.

Our second major finding was that significant differences in W-scores were observed across cognitive stages (CU, MCI, and DAT) for all ROIs, reflecting the algorithm's ability to detect stage-specific atrophy patterns. Temporal, hippocampal, and ventricular regions exhibited the largest differences, with W-scores progressively decreasing from CU to DAT, indicating increasing CSF volume associated with advanced atrophy. These results confirm that our approach, while novel, performs comparably to existing methods[24-26] in identifying stage-specific atrophy patterns. Furthermore, by focusing on a comprehensive set of ROIs that cover both cortical and subcortical regions, our method provides a broader view of brain atrophy progression compared to prior studies[27-30] that often emphasize isolated regions such as the hippocampus or LVs. This capability ensures sensitivity across the spectrum of cognitive impairment, offering a robust tool for detecting and quantifying structural brain changes. The alignment of these results with clinical markers such as ventricular enlargement and hippocampal atrophy underscores the clinical relevance of our algorithm and its potential to support decision-making in both research and practice.

Our final major finding was that similar trends were observed in the images from the Cohort 2, confirming the generalizability of our findings. Significant differences in W-scores across cognitive stages (CU, MCI, and DAT) were consistently observed for key ROIs including temporal, hippocampal, and ventricular regions in the Cohort 2, mirroring the results from the Cohort 1. This consistency highlights the robustness of the algorithm across diverse datasets. A key reason for focusing on ROIs with pronounced intensity differences, such as the CSF-gray matter and CSF-white matter boundaries, was to ensure that the algorithm performs reliably even for imaging datasets without extensive preprocessing or standardization. Unlike traditional approaches[10,13] that often rely on tightly controlled imaging conditions, our method was designed to be adaptable to diverse datasets. The observed consistency between the Cohort 1 and Cohort 2 suggests that the algorithm's reliance on clearly defined intensity contrasts helps make it adaptable across various clinical and research settings, even when image acquisition protocols differ. By focusing on intensity differences that tend to remain stable across imaging modalities and datasets, this algorithm aims to improve reproducibility and standardization in heterogeneous populations. This adaptability makes its potential utility in real-world settings and suggests that it could be useful in broader clinical workflows, including those incorporating CT or 2D T1 MRI data.

Our study's methodological strength lies in its ability to leverage intensity differences at CSF boundaries to achieve robust and reproducible segmentation and quantification across diverse cohorts and imaging conditions. However, this study has several limitations. First, the analysis was based solely on 3D T1 MRI data, and the performance of the algorithm on other modalities, such as CT or 2D T1 MRI, remains untested. Second, while the images from the Cohort 2 demonstrated similar trends, variability in imaging protocols and demographic characteristics could influence generalizability. Notably, significant differences between MCI and DAT were not observed in the left frontal, parietal and occipital ROIs in the Cohort 2, possibly due to differences in image quality or acquisition parameters. Further investigation is needed to determine whether this reflects methodological variability or biological differences. Third, the algorithm primarily focuses on CSF segmentation, which may limit its application in evaluating other brain regions, such as deep gray matter structures. Finally, the cross-sectional nature of the study precludes any conclusions about the algorithm's ability

https://doi.org/10.12779/dnd.2025.24.2.115

to track longitudinal changes in brain atrophy. Nonetheless, our findings provide valuable insights into brain atrophy patterns across cognitive stages and demonstrate the potential of CSF-based segmentation methods may contribute to both research and clinical workflows. While the algorithm's adaptability across diverse cohorts, additional validation is required to confirm its broader applicability, particularly in routine clinical practice and lower-resolution imaging modalities.

In summary, our study developed a robust algorithm leveraging CSF boundaries to achieve precise segmentation and quantification of brain atrophy, demonstrating strong performance across diverse cohorts. Significant differences in W-scores were consistently observed across cognitive stages, validating the algorithm's ability to detect stage-specific atrophy patterns. These findings highlight the potential of this method to enhance clinical and research workflows, offering a scalable tool for broader applications in neurodegenerative disease assessment.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
MRI information for all cohorts

Supplementary Table 2
Demographic information of dataset for segmentation model

Supplementary Fig. 1
Segmentation results for the 14 cerebrospinal fluid space across cognitive stages.

REFERENCES

1. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease. Neurology 1984;34:939-944. PUBMED | CROSSREF
2. Tzioras M, McGeachan RI, Durrant CS, Spires-Jones TL. Synaptic degeneration in Alzheimer disease. Nat Rev Neurol 2023;19:19-38. PUBMED | CROSSREF
3. Im K, Lee JM, Seo SW, Hyung Kim S, Kim SI, Na DL. Sulcal morphology changes and their relationship with cortical thickness and gyral white matter volume in mild cognitive impairment and Alzheimer's disease. Neuroimage 2008;43:103-113. PUBMED | CROSSREF
4. Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 2010;6:67-77. PUBMED | CROSSREF
5. Ferreira D, Verhagen C, Hernández-Cabrera JA, Cavallin L, Guo CJ, Ekman U, et al. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Sci Rep 2017;7:46263. PUBMED | CROSSREF
6. Fumagalli GG, Basilico P, Arighi A, Mercurio M, Scarioni M, Carandini T, et al. Parieto-occipital sulcus widening differentiates posterior cortical atrophy from typical Alzheimer disease. Neuroimage Clin 2020;28:102453. PUBMED | CROSSREF
7. Pasquier F, Leys D, Weerts JGE, Mounier-Vehier F, Barkhof F, Scheltens P. Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. Eur Neurol 1996;36:268-272. PUBMED | CROSSREF
8. Harper L, Barkhof F, Fox NC, Schott JM. Using visual rating to diagnose dementia: a critical evaluation of MRI atrophy scales. J Neurol Neurosurg Psychiatry 2015;86:1225-1233. PUBMED | CROSSREF
9. Cavallin L, Løken K, Engedal K, Oksengård AR, Wahlund LO, Bronge L, et al. Overtime reliability of medial temporal lobe atrophy rating in a clinical setting. Acta Radiol 2012;53:318-323. PUBMED | CROSSREF

10. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci U S A 2000;97:11050-11055. PUBMED | CROSSREF
11. Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. Neuroimage 2014;99:166-179. PUBMED | CROSSREF
12. Dahnke R, Yotter RA, Gaser C. Cortical thickness and central surface estimation. Neuroimage 2013;65:336-348. PUBMED | CROSSREF
13. Ashburner J, Friston KJ. Voxel-based morphometry--the methods. Neuroimage 2000;11:805-821. PUBMED | CROSSREF
14. Leung KK, Clarkson MJ, Bartlett JW, Clegg S, Jack CR Jr, Weiner MW, 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. PUBMED | CROSSREF
15. Jang H, Shin D, Kim Y, Kim KW, Lee J, Kim JP, et al. Korea-Registries to Overcome Dementia and Accelerate Dementia Research (K-ROAD): a cohort for dementia research and ethnic-specific insights. Dement Neurocogn Disord 2024;23:212-223. PUBMED | CROSSREF
16. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 2010;74:201-209. PUBMED | CROSSREF
17. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263-269. PUBMED | CROSSREF
18. Billot B, Greve DN, Puonti O, Thielscher A, Van Leemput K, Fischl B, et al. SynthSeg: segmentation of brain MRI scans of any contrast and resolution without retraining. Med Image Anal 2023;86:102789. PUBMED | CROSSREF
19. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 2001;356:1293-1322. PUBMED | CROSSREF
20. LaMontagne PJ, Benzinger TL, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medRxiv. Dec 15, 2019. https://doi.org/10.1101/2019.12.13.19014902. CROSSREF
21. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-211. PUBMED | CROSSREF
22. La Joie R, Perrotin A, Barré L, Hommet C, Mézenge F, Ibazizene M, et al. Region-specific hierarchy between atrophy, hypometabolism, and β-amyloid (Aβ) load in Alzheimer's disease dementia. J Neurosci 2012;32:16265-16273. PUBMED | CROSSREF
23. Chandra A, Dervenoulas G, Politis M; Alzheimer's Disease Neuroimaging Initiative. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. J Neurol 2019;266:1293-1302. PUBMED | CROSSREF
24. Fan Y, Batmanghelich N, Clark CM, Davatzikos C; Alzheimer's Disease Neuroimaging Initiative. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 2008;39:1731-1743. PUBMED | CROSSREF
25. Ma X, Li Z, Jing B, Liu H, Li D, Li H, et al. Identify the atrophy of Alzheimer's disease, mild cognitive impairment and normal aging using morphometric MRI analysis. Front Aging Neurosci 2016;8:243. PUBMED | CROSSREF
26. Apostolova LG, Green AE, Babakchanian S, Hwang KS, Chou YY, Toga AW, et al. Hippocampal atrophy and ventricular enlargement in normal aging, mild cognitive impairment (MCI), and Alzheimer disease. Alzheimer Dis Assoc Disord 2012;26:17-27. PUBMED | CROSSREF
27. Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry 1992;55:967-972. PUBMED | CROSSREF
28. Koedam ELGE, Lehmann M, van der Flier WM, Scheltens P, Pijnenburg YAL, Fox N, et al. Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol 2011;21:2618-2625. PUBMED | CROSSREF
29. Ferreira D, Cavallin L, Granberg T, Lindberg O, Aguilar C, Mecocci P, et al. Quantitative validation of a visual rating scale for frontal atrophy: associations with clinical status, APOE e4, CSF biomarkers and cognition. Eur Radiol 2016;26:2597-2610. PUBMED | CROSSREF
30. O'Donovan J, Watson R, Colloby SJ, Firbank MJ, Burton EJ, Barber R, et al. Does posterior cortical atrophy on MRI discriminate between Alzheimer's disease, dementia with Lewy bodies, and normal aging? Int Psychogeriatr 2013;25:111-119. PUBMED | CROSSREF