

Research Report

Automated Scoring of Alzheimer's Disease Atrophy Scale with Subtype Classification Using Deep Learning-Based T1-Weighted Magnetic Resonance Image Segmentation

Yeong Sim Choe^a, Regina E.Y. Kim^a, Hye Weon Kim^a, JeeYoung Kim^{b,1}, Hyunji Lee^a,
Min Kyoung Lee^c, Minho Lee^a, Keun You Kim^d, Se-Hong Kim^e, Ji-hoon Kim^f, Jun-Young Lee^{d,g},
Eosu Kim^h, Donghyeon Kim^{a,*},² and Hyun Kook Lim^{i,*},²

^aResearch Institute, Neurophet Inc., Seoul, Republic of Korea

^bDepartment of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

^cDepartment of Radiology, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

^dDepartment of Psychiatry, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

^eDepartment of Family Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon, Republic of Korea

^fDepartment of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

^gDepartment of Psychiatry and Department of Medical Device Development, Seoul National University College of Medicine, Seoul, Republic of Korea

^hDepartment 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

ⁱDepartment of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

Received 7 August 2023

Accepted 27 March 2024

Published 17 May 2024

Abstract.

Background: Application of visual scoring scales for regional atrophy in Alzheimer's disease (AD) in clinical settings is limited by their high time cost and low intra/inter-rater agreement.

Objective: To provide automated atrophy scoring using objective volume driven from deep-learning segmentation methods for AD subtype classification using magnetic resonance imaging (MRI).

¹Deceased.

²These authors contributed equally to this work.

*Correspondence to: Donghyeon Kim, PhD, Research Institute, Neurophet Inc., 12F, 124, Teheran-ro, Gangnam-gu, Seoul, 06234, Republic of Korea. Tel.: +82 2 6954 7971; Fax: +82 2 6954 7972; E-mail: donghyeon.kim@neurophet.com. and

Hyun Kook Lim, MD, PhD, Professor, Department of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 10, 63-ro, Yeongdeungpo-gu, Seoul, 07345, Republic of Korea. Tel.: +82 2 23779 1048; Fax: +82 2 780 6577; E-mail: drblues@catholic.ac.kr.

Methods: We enrolled 3,959 participants (1,732 cognitively normal [CN], 1594 with mild cognitive impairment [MCI], and 633 with AD). The occupancy indices for each regional volume were calculated by dividing each volume by the size of the lateral and inferior ventricular volumes. MR images from 355 participants (119 CN, 119 MCI, and 117 AD) from three different centers were used for validation. Two neuroradiologists performed visual assessments of the medial temporal, posterior, and global cortical atrophy scores in the frontal lobe using T1-weighted MR images. Images were also analyzed using the deep learning-based segmentation software, Neurophet AQUA. Cutoff values for the three scores were determined using the data distribution according to age. The scoring results were compared for consistency and reliability.

Results: Four volumetric-driven scoring results showed a high correlation with the visual scoring results for AD, MCI, and CN. The overall agreement with human raters was weak-to-moderate for atrophy scoring in CN participants, and good-to-almost perfect in AD and MCI participants. AD subtyping by automated scores also showed usefulness as a research tool.

Conclusions: Determining AD subtypes using automated atrophy scoring for late-MCI and AD could be useful in clinical settings or multicenter studies with large datasets.

Keywords: Alzheimer's disease, atrophy, cognitive dysfunction, magnetic resonance imaging, visual scoring

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia and involves a decline in memory, cognition, language, and activities of daily living. With amyloid-targeting drugs currently on trial and expected to be on the market soon, the identification of AD biomarkers for adequate clinical intervention is of growing interest to clinicians, researchers, and public health service providers. To promptly introduce precision-medicine treatments, increased knowledge on the diverse subtypes of degenerative dementia, including AD, is necessary. Moreover, for the differential diagnosis of the degenerative dementia subtypes, the range and combination of atrophy findings across brain regions on brain magnetic resonance (MR) images are fundamental.

Brain magnetic resonance imaging (MRI) may provide essential information for diagnosing the etiology of dementia, also allowing to exclude secondary or reversible causes of dementia [1]. Thus, several brain structures are known to present atrophy earlier than others during the progression of AD, as widely reported in several MRI studies [2, 3]. Previous studies have also reported that hippocampal volume and cortical thinning in the entorhinal cortex are highly sensitive measures of structural change both in AD and mild cognitive impairment (MCI) [4, 5].

Brain atrophy scales, which are generally graded visually by clinicians and radiologists, have been proposed as an objective approach to quantify the degree of atrophy. Brain atrophy scales if they could reflect the actual clinical symptoms may be further utilized to identify AD subtypes. Common visual rating scales include the medial temporal atrophy (MTA) scale [6],

global cortical atrophy scale – frontal subscale (GCA-F) [7], and posterior atrophy (PA) scale [8]. Prior studies have also compared visual atrophy scores with measured brain volume, cortical thickness, memory performance, non-memory function, and disease progression [2, 9–17].

Automated atrophy scoring and MR image scoring for identifying AD subtypes have several advantages over the use of biomarkers extracted from the cerebrospinal fluid and positron emission tomography (PET) images. MRI is a non-invasive, more accessible approach than other diagnostic approaches. For example, cerebrospinal fluid biomarkers provide a way to obtain Amyloid/Tau/Neurodegenerative (A/T/N) biomarkers simultaneously; however, they require an invasive procedure accompanied by cerebrospinal fluid (CSF) lumbar tapping. The CSF lumbar tapping procedure presents potential risks, such as post-lumbar puncture headache, back discomfort or pain, bleeding, and brainstem herniation. PET images provide functional and early diagnostic measurements non-invasively, and with higher reliability; however, the disadvantages include the risks associated with radiation exposure and high costs [18]. In contrast, MR images can be obtained without radiation exposure and have a relatively lower cost than PET images [1].

Visual scoring, however, has low inter- and intra-rater reliability and is a time-consuming and labor-intensive task. Visual quantification of brain volumes and measurement of cortical thickness can differ, along with visual atrophy scale scoring, even between neuroradiologists and dementia specialists. These caveats can make it difficult for doctors to trust the atrophy scales scored by others, and can also be a barrier in integrating multi-rater or multi-center

datasets. Hence, many efforts are ongoing to convert conventional visual atrophy scoring into quantitative measuring methods to help reduce the workload and provide a higher degree of concordance. Furthermore, the utility of quantifiable and objective methods could also be expanded to differentiate the four AD subtypes of minimal atrophy, limbic-predominant, hippocampal-sparing, and typical AD [19, 20].

The purpose of this study was to introduce an automated rating approach for atrophy scales (MTA, GCA-F, and PA scales) using volumetric measurements. Integrable indices were used for quantification and automation. We then validated our automated rating approach by comparing them with visual ratings of MTA, GCA-F, and PA scales obtained from radiologists. In addition, a comparison between human raters and our method was provided, and appropriate cutoff values to differentiate between normal and abnormal cases were described for each region. Finally, we presented the implications of our method for use in AD subtype classification based only on volumetric atrophy information from MR images, without the need to integrate any other modality data or biomarkers.

METHODS

Study design overview

This study aimed to develop an automated atrophy scoring system based on volumetric measurements and utilized these scores for AD subtype classification. A total of 3,959 participants, including cognitively normal (CN), MCI, and AD subjects, were enrolled in the atrophy score development set. Regional volumes were quantified, and occupancy indices were calculated by normalizing each volume against the sum of inferior lateral ventricular volumes, which were then categorized into quintiles. To validate our automated score, MR images from 355 participants across three different centers were used. Visual assessments by neuroradiologists and automated atrophy scoring using Neurophet AQUA were compared for consistency and reliability, with cutoff values determined based on age-related data distributions and previous studies [21–23]. Subsequently, we compared the ability of identifying AD subtypes using the atrophy scores by using previous definitions [2] and comparing the neuroradiologists' score-based subtypes and the automated score-based subtypes. Our study flowchart is presented in Fig. 1.

Participants

In total, 3,959 participants were included in this study for the development and validation groups in seven different centers. The age range was between 20 and 110 years, and three diagnosis groups (CN, $n = 1,732$; MCI, $n = 1,594$; and AD, $n = 633$) were available. The validation dataset was also composed of elderly participants, including CN, MCI, and AD cases from three centers (The Catholic University of Korea Yeouido St. Mary's Hospital, SMG-SNU Boramae Medical Center, and The Yonsei University Severance Mental Health Hospital). For all centers, the diagnosis groups were identified by all qualified clinicians based on the National Institute on Aging Alzheimer's Association (NIA-AA) research criteria for the probable disease group [24] for AD. Participants diagnosed with MCI according to Peterson's criteria reported memory impairment (considered in the context of their age, education, and sex by clinicians) or were reported to exhibit memory impairments by a relative, with mostly intact function of daily living and Clinical Dementia Rating (CDR) of 0.5 [25–28]. The CN group had normal cognitive function without any significant impairment in memory function assessed by word list memory, word list recall, and word list recognition domains according to their age, sex, and education status, and a CDR score and Memory Box score of 0 [29].

This study has obtained Institutional Review Board (IRB) approval from each individual center. The development dataset was approved by Wonkwang University Hospital (approval no. 2019-09-002-002), the Catholic Aging Brain Imaging (CABI) database, which holds brain MRI scans of patients enrolled at the Catholic Brain Health Center, Yeouido St. Mary's Hospital, and Eunpyeong St. Mary's Hospital at the Catholic University of Korea (approval no. XC20RID10035V, PC20EISI0007, SC20RISI0198), and Seoul National University Hospital (approval no. D-2106-221-1233). The validation dataset was approved by the IRB of Yonsei University Health System (approval no. 1-2021-0025), Yeouido St. Mary's hospital (approval no. 2020-3348-0007), and SMG-SNU Boramae Medical Center (approval no. 30-2020-325).

MR images acquisition and image processing

All brain MR images were collected using 3T 3D T1-weighted images (T1-w) at individual centers. Detailed parameters are described in Supplemen-

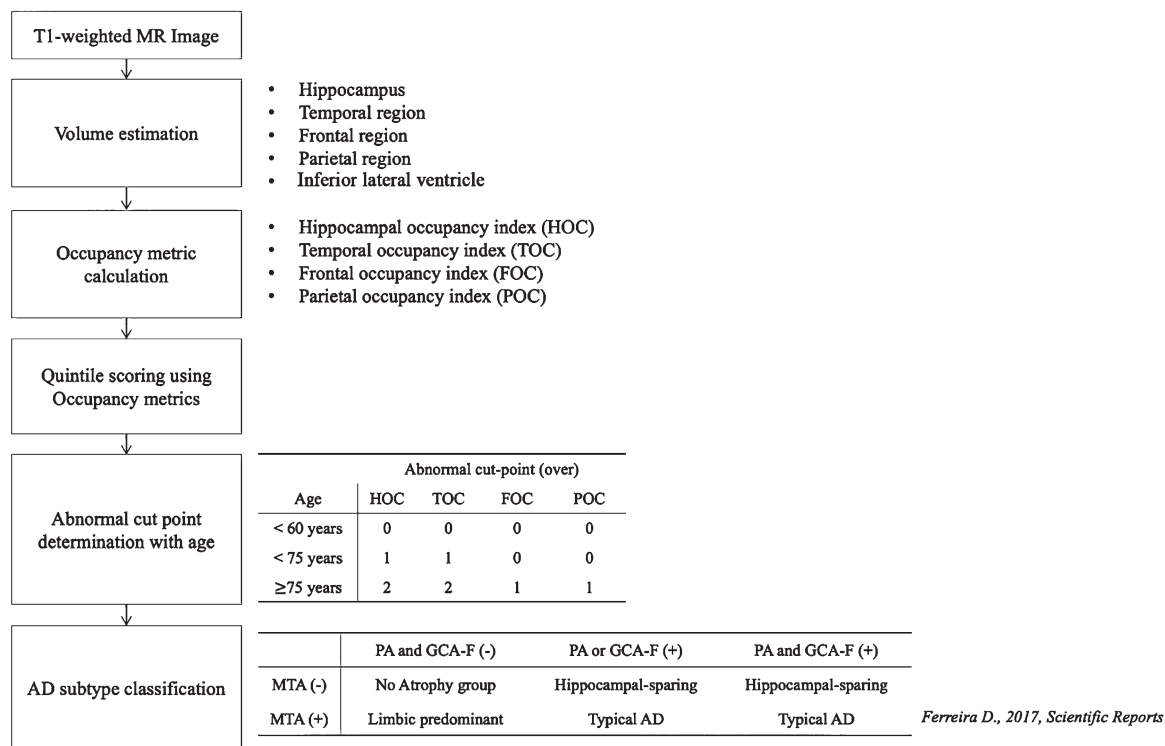


Fig. 1. Flowchart of study pipeline. AD, Alzheimer's diseases; MTA, medial temporal atrophy score; GCA-F, global cortical atrophy score of the frontal lobe; PA, posterior atrophy score; HOC, hippocampus occupancy index; TOC, medial temporal lobe occupancy index; FOC, frontal lobe occupancy index; POC, parietal lobe occupancy index; (-), normal; (+), abnormal.

tary Table 1. Brain volumes from the participants were estimated using T1-w MR images with a deep-learning-based MRI segmentation software, Neurophet AQUA (Neurophet Inc., Seoul, Republic of Korea) version 2.1.4. Neurophet AQUA is a brain MRI segmentation software based on the deep-learning algorithm, the Split-attention U-net (SAU-Net), as previously described [30]. Briefly, regions of interest (ROIs) were defined based on the Desikan–Killiany atlas as in FreeSurfer [31]. Neurophet AQUA was then trained with the ROI definitions reviewed and corrected by neuroradiologists. The validation of the software was conducted through comparison (dice overlap and the average symmetric surface distance) on 22 major ROIs (in the two hemispheres of cerebral gray matter, cerebral white matter, cerebellum, lateral ventricle, caudate, accumbens, putamen, amygdala, hippocampus, pallidum, and thalamus) and showed a higher performance than that of well-known software FastSurfer [32] or QuickNAT [33].

ROIs for volumetric estimation were based on the radiologists' visual assessment, referring to regions

decided based on discussion with psychologists. Individual lobe definitions were as follows: medial temporal lobe (entorhinal, parahippocampus, and hippocampus), parietal lobe (inferior and superior parietal, postcentral, precuneus, and supramarginal), and frontal lobe (caudal and rostral middle frontal, lateral and medial orbitofrontal, paracentral, pars opercularis, pars orbitalis, pars triangularis, precentral, superior frontal, and frontal pole) [2].

MR image visual assessment

Visual interpretation of the atrophy scale was conducted by two radiologists (J.Y.K. and M.Y.L.) in a blinded test without any patient information, including diagnostic information, age, sex, or other radiologists' rating results. The atrophy score was composed of three regional scores: MTA, PA, and GCA-F. MTA scale scores are based on the degree of atrophy in the medial temporal area, including the hippocampus, parahippocampal gyrus, and entorhinal cortex, divided by the size of inferior ventricle. The PA scores in the posterior atrophy involve

the posterior cingulate sulcus, precuneus, parieto-occipital sulcus, and overall parietal cortex. Finally, the GCA-F scores represent the frontal lobules surrounding the central sulcus, frontal bone, and fissure of Sylvius [34]. PA and GCA-F were composed of 0-to-3 scales and considered the widening of the sulci [35]. After the individual rating process was performed, both radiologists determined the final visual assessment result by discussion until an agreement was reached.

Visual atrophy scores were converted into a binary value considering age differences [21]. The MTA score ranges from 0 to 4, GCA-F scores from 0 to 3, and PA scores from 0 to 3. The definition of an abnormal MTA score varied according to age: abnormal MTA scores were 1–3, 2–3, and 3 for the under 70 years, 70–80 years, and ≥ 80 years age groups, respectively [21]. GCA-F and PA scores were defined as abnormal when the value was above zero (GCA-F or PA > 0) [34].

Development of automated atrophy scoring using population distribution and abnormality.

To estimate the automatic regional atrophy scores, we considered four occupancy indices [27]: the frontal occupancy index (FOC), parietal occupancy index (POC), temporal occupancy index (TOC), and hippocampal occupancy index (HOC); using the following equation:

$$\text{Occupancy index}_{\text{target region}} = \text{mean} \left(\frac{\text{Volume}_{\text{target region}}}{\text{Volume}_{\text{target region}} + \text{Volume}_{\text{Inferior lateral ventricle}}} \right).$$

TOC, POC, and FOC were originated from the HOC score, which is one of the features related with hippocampal atrophy measurement. Once the occupancy indices of the corresponding ROIs were calculated, they were further categorized into quintiles for automated regional atrophy scoring. These quintile cutoffs were used for automatic regional atrophy scoring from 0 to 4. All participants were included in this analysis.

The automatic regional atrophy scores were then further divided into normal and abnormal atrophy. The binary cutoff values according to age for automatic regional atrophy scores were determined after investigating the distribution of four occupancy indices (Fig. 2). The medial temporal regional age cutoff value was set at 0 for participants under 60 years of age, 1 for participants between 60 and 75 years of age, and 2 for those above 75 years of age. However, participants above 75 years

of age were assigned a lower cutoff value in the frontal and posterior lobes (0 for those under 75 years, and 1 for those above 75 years). These age-based cutoff values had been confirmed in previous studies [21–23].

Subtyping in Alzheimer's disease participants

In this study, AD cases were then divided into four types (typical AD, limbic-predominant, minimal atrophy, and hippocampal-sparing types) entirely based on automated atrophy scoring. Those AD subtyping using volumetric features from MR images which are independent from other clinical information or neuropsychological status were suggested previously [2]. In the previous study [2], the typical AD type was defined as participants with overall atrophy pattern in the parietal, temporal, and frontal cortices. The limbic-predominant type included those participants with medial temporal lobe atrophy only and normal posterior and frontal cortices. The minimal atrophy type included participants who did not show any atrophy pattern in the lobes related to AD. Within the minimal atrophy type, participants without atrophy in the medial temporal cortices were categorized into the hippocampal-sparing type. Following those previous definitions for the subtypes, we utilized our automated atrophy scoring of TOC, POC, and FOC, which correspond to MTA, PA, and GCA-F. For MTA, the TOC index was utilized instead of the HOC index because the MTA score was estimated using multiple correlations in the medial temporal, hippocampus, and inferior horn of the lateral ventricle [23]. In addition, we recruited patients with AD and MCI who showed a CDR sum of boxes (SOB) higher than 3.5 as a subtyping type.

Statistical analysis

We used a diverse statistical test to compare the differences between each group for numeric variables (an independent sample *t*-test, and one-way analysis of variance (ANOVA) and the qualitative variables (a chi-square test). After one-way ANOVA, the Bonferroni correction was adopted for *post-hoc* analysis [36] and a *p*-value < 0.001 was considered significant. Accuracy and Gwet's AC2 [37] were conducted to compare the visual ratings of the two radiologists (JY. K and M.Y.L.). Accuracy, sensitivity, and specificity are values that estimate the corrected sample number compared to the total sample number. Gwet's AC2, a second-order agreement coefficient with ordi-

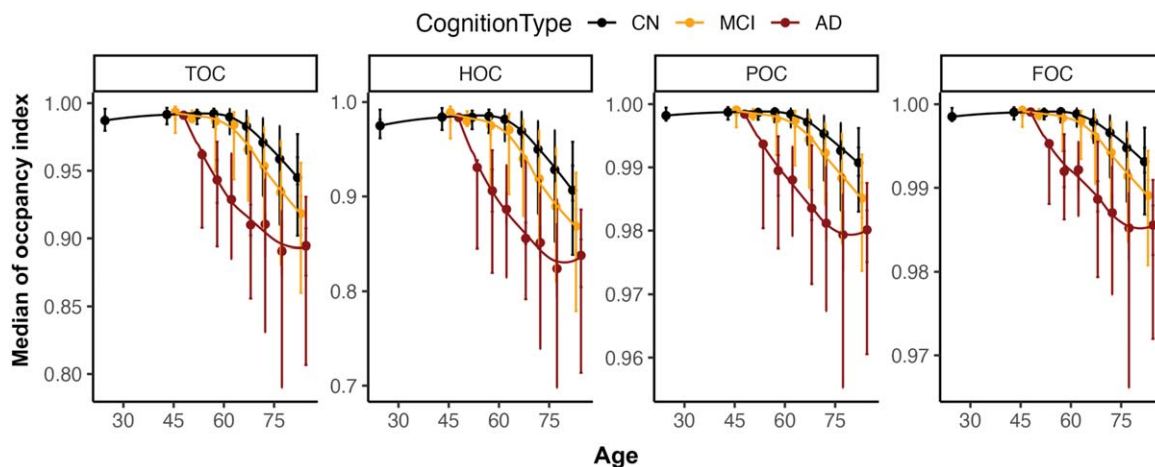


Fig. 2. Distribution overview for four occupancy indices (TOC, HOC, POC, and FOC) according to age. For the <30, 31–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, and 80–100 age groups, the median age and median occupancy indices were plotted with quintiles (Q1, Q2, Q3, and Q4 from the top to bottom as error bar). TOC, medial temporal lobe occupancy index; HOC, hippocampus occupancy index; POC, parietal lobe occupancy index; FOC, frontal lobe occupancy index; CN, Cognitive normal; MCI, mild cognitive impairment; AD, Alzheimer's disease.

nal weighting applied, was shown to have a higher inter-rater reliability coefficient and stable reliability coefficient because it was established to overcome the limitations of Cohen's kappa [38] when score properties were different between groups [39,40]. The probabilistic method for benchmarking, as suggested by Gwet. The minimal inter-rater agreement was 0.21–0.39 and 0.40–0.59, 0.60–0.79, 0.80–0.90, and >0.90 as minimal, weak, moderate, strong, and almost perfect [41, 42]. In addition, we used accuracy and Gwet's AC2 score as indicators of the agreement between the two raters and the performance score of the proposed atrophy scoring system compared to the visual atrophy scores. All statistical analyses were performed using R software version 3.6.3 and used the following R packages: caret [43], irr [44], and irrCAC [45] to compare the accuracy and inter-rater reliability.

RESULTS

Patients demographics

The demographics of participants in the development ($N=3,959$) and validation ($N=355$) datasets are presented in Table 1. Patients in both datasets showed similar mean age and sex distribution, while the development dataset showed a wider age range due to the inclusion of younger participants (<50 years, 206 participants, 1 AD, 8 MCI, and 197 CN) ($p=0.003$). There was no significant difference in the

Table 1
Demographics of the participants in the development and validation datasets

| | Development dataset | Validation dataset |
|------------------------|--------------------------|--------------------------|
| Number of participants | 3,959 | 355 |
| Age, y | 71.7 ± 12.3 | 73.1 ± 8.0 |
| Female, % (N) | 68.2% (2,702) | 69.3% (246) |
| Cognitive type | | |
| AD, % (N) | 16.0% (633) ^a | 33.0% (117) ^a |
| MCI, % (N) | 40.3% (1594) | 33.5% (119) |
| CN, % (N) | 43.7% (1732) | 33.5% (119) |

CN, cognitively normal group; AD, Alzheimer's disease; MCI, mild cognitive impairment. ^a $p < 0.001$ between CN and AD.

sex distribution between the development and validation datasets ($p=0.729$). However, the cognitive-type distribution showed a significant difference among AD and CN patients ($p < 0.001$), but not among MCI patients ($p=0.015$). As expected, there were significant differences in age according to the diagnosis group, both in the development dataset ($p=0.015$) and in the validation dataset ($p < 0.001$), with slightly younger ages registered in the CN and MCI groups.

Correspondence between atrophy scores across radiologists

The correspondence between the ratings of the two radiologists estimated using the validation dataset showed moderate similarity for all three atrophy scores in Table 4. The accuracy of the atrophy

scores between the two radiologists was 0.619, 0.627, and 0.547 for MTA, GCA-F, and PA, respectively. After dichotomizing the atrophy scores to normal and abnormal groups using age-adjusted cutoffs, the accuracy was improved to 0.834, 0.927, and 0.972, respectively. While sensitivity was similarly high across the three scores (0.791, 0.991, and 0.977), specificity showed quite a different pattern between MTA and PA or GCA-F (0.900, 0.207, 0.333, respectively). In addition, Gwet's AC2 was moderate in MTA (0.673) and almost perfect in GCA-F and PA (0.919 and 0.971).

Quantitative atrophy scoring using MR-driven measurements

The distributions of the four regional occupancy indices in our development dataset are shown in Fig. 2. As expected, all the occupancy indices decreased with age, since the volume of ROI also decreases with age. In particular, AD patients showed a faster decline in occupancy indices in TOC and HOC, including MTA-specific regions, than MCI or CN patients aged between 50 and 75 years. In contrast, POC and FOC regions showed a relatively moderate decline both with age and disease severity (Fig. 2). The four regional occupancy indices showed significant differences among AD, MCI, and CN patients ($p < 0.001$), and significantly decreased with age ($p < 0.001$). As shown in Table 2, the mean occupancy index significantly decreased from Q1 to Q5 ($p < 0.001$). The largest difference was observed for HOC (from 0.990 [Q1] to 0.742 [Q5]: -0.248), and the smallest difference was observed for FOC (from 0.999 [Q1] to 0.975 [Q5]: -0.024). Thresholds for each occupancy index were determined by including all CN participants and the age-specific cutoff values suggested in previous studies [21–23].

Application of quantitative atrophy scoring

To demonstrate the clinical application of our quantitative atrophy scoring system, the regional atrophy score and the corresponding age-weighted cutoff value were applied to our validation set. The trend of the occupancy index in the validation set was similar to that in the development set. The average of four indices was negatively correlated with the visual score, and the standard deviation increased with the visual score. In the case of low occupancy indices, we did not find any high atrophy scores determined by the radiologists. In contrast, in the case of low scores on the visual scale, a low occupancy index was included for PA and GCA-F (Fig. 3 and Supplementary Table 2). In general, the association of GCA-F with age was relatively gradual compared to that of MTA or PA.

After converting the atrophy score to a dichotomized abnormality score using age-adjusted cutoffs, most of AD and MCI cases were included in the abnormal group (Table 3). The correspondence between the radiologists' rating and the automated score improved for all four indices (Table 4). The highest accuracy and AC2 score corresponded to the score for the parietal lobe (accuracy/AC2: 0.876/0.853), while the lowest values were registered for the temporal lobe (accuracy/AC2: 0.721/0.483). We selected the TOC score from the medial temporal scoring index, and discarded the HOC score in our further analysis because of its lower accuracy, sensitivity, specificity, and Gwet's AC2 score.

Comparison of AD subtyping based on radiologists' score and volume-driven score

Atrophy scores derived from occupancy indices and from radiologists' rating were used to categorize AD subtypes, as summarized in Table 5 and Fig. 4. The minimal atrophy type was found in fewer

Table 2
Four regional occupancy indices summarized across quintiles for the development dataset including young and elderly cognitive normal

| | Q1 | Q2 | Q3 | Q4 | Q5 |
|-----|---------------|---------------|---------------|---------------|---------------|
| HOC | 0.990 ± 0.005 | 0.969 ± 0.008 | 0.929 ± 0.015 | 0.870 ± 0.019 | 0.742 ± 0.081 |
| TOC | 0.995 ± 0.003 | 0.983 ± 0.005 | 0.959 ± 0.009 | 0.921 ± 0.013 | 0.829 ± 0.064 |
| POC | 0.999 ± 0.000 | 0.997 ± 0.001 | 0.993 ± 0.002 | 0.986 ± 0.003 | 0.966 ± 0.016 |
| FOC | 0.999 ± 0.000 | 0.998 ± 0.001 | 0.995 ± 0.001 | 0.990 ± 0.002 | 0.975 ± 0.012 |

Q1–Q5, Quintile range 1–5; HOC, hippocampus occupancy index; TOC, medial temporal lobe occupancy index; FOC, frontal lobe occupancy index; POC, parietal lobe occupancy index.

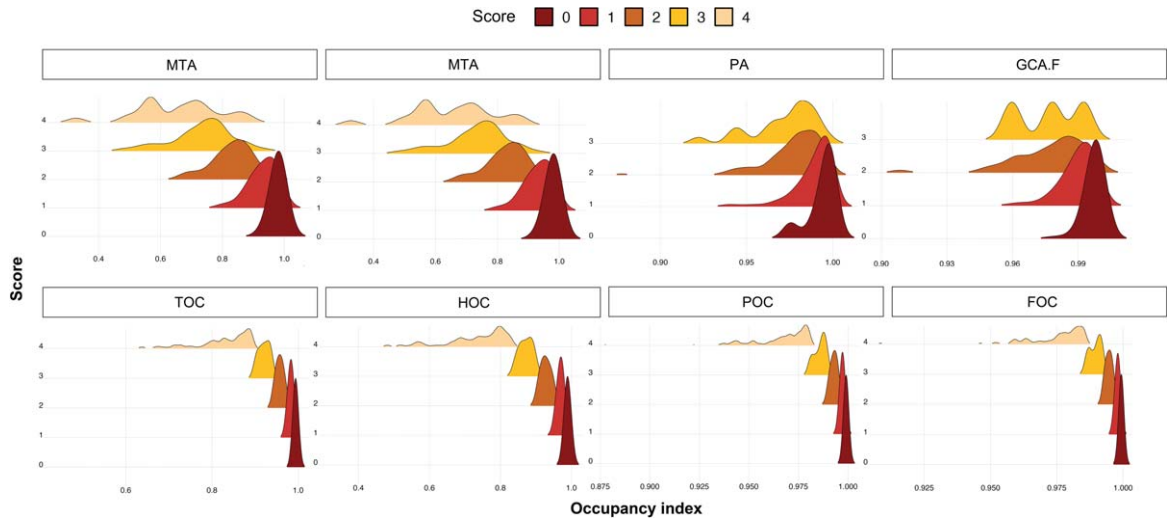


Fig. 3. Comparison of regional score data distribution between visual assessment (upper figures) and automated scores (bottom figures) across regional occupancy indices. MTA, medial temporal atrophy score; GCA-F, global cortical atrophy score of the frontal lobe; PA, posterior atrophy score; HOC, hippocampus occupancy index; TOC, medial temporal lobe occupancy index; FOC, frontal lobe occupancy index; POC, parietal lobe occupancy index.

Table 3
Demographics across the occupancy index abnormality groups

| | TOC ^a | | HOC ^a | |
|--------------------------------|------------------|------------|------------------|------------|
| | Normal | Abnormal | Normal | Abnormal |
| <i>N</i> | 118 | 237 | 116 | 239 |
| Age (y) | 70.5 ± 8.2 | 74.5 ± 7.6 | 70.1 ± 8.3 | 74.6 ± 7.5 |
| Cognition types % (<i>N</i>) | | | | |
| AD | 10.2 (12) | 44.3 (105) | 8.6 (10) | 44.8 (107) |
| MCI | 33.1 (39) | 33.8 (80) | 57.8 (67) | 33.5 (80) |
| CN | 56.8 (67) | 21.9 (52) | 33.6 (39) | 21.8 (52) |
| | POC ^b | | FOC ^b | |
| | Normal | Abnormal | Normal | Abnormal |
| <i>N</i> | 43 | 312 | 49 | 306 |
| Age (y) | 64.7 ± 7.5 | 74.3 ± 7.4 | 65.2 ± 7.5 | 74.4 ± 7.4 |
| Cognition types % (<i>N</i>) | | | | |
| AD | 4.7 (2) | 36.9 (115) | 6.1 (3) | 37.3 (114) |
| MCI | 23.3 (10) | 34.9 (109) | 26.5 (13) | 34.6 (106) |
| CN | 72.1 (31) | 28.2 (88) | 67.3 (33) | 28.1 (86) |

TOC, medial temporal lobe occupancy index; HOC, hippocampus occupancy index; POC, parietal lobe occupancy index; FOC, frontal lobe occupancy index; AD, Alzheimer's disease; MCI, mild cognitive impairment; CN, cognitively normal. ^aNormal: Q0, Abnormal: Q1–Q5 for participants under 60 years; Normal: Q0–Q1, Abnormal: Q2–Q5 for participants between 60 and 75 years; Normal: Q0–Q2, Abnormal: Q3–Q5 for participants aged 75 years or above. ^bNormal: Q0, Abnormal: Q1–Q5 for participants under 75 years; Normal: Q0–Q1, Abnormal: Q2–Q5 for participants over 75 years.

participants based on both the radiologists' and the quantitative scores, with one identification driven by the radiologists and two by the occupancy indices. Furthermore, patients with the minimal atrophy type were younger than those in the cognitively normal control group (59.0 ± 5.7 and 63.0 years). The hippocampal-sparing and typical AD types showed significant age differences, Mini-Mental State Examination scores, and CDR SOB compared with the

control group ($p < 0.001$) in data-driven subtyping results; however, age showed no significant difference in the radiologists-driven results. In the subtyping results from the radiologists, the hippocampal sparing type was older than the typical AD type ($p < 0.05$). However, there were no significant differences in sex, cognitive type, and apolipoprotein E $\epsilon 4$ allele carrier status between the visual and volume scores predicted as reflecting the same subtype.

Table 4

Comparison of sensitivity, specificity, accuracy, and Gwet's AC2 score: 1) between raters, and 2) between methods (occupancy index and visual assessment) after converting multiple scores to binary scores with disease group

| Visual rating | Cognitive type | | | | | | | | | | | | | | | |
|---------------|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | AD | | | | MCI | | | | CN | | | | Total | | | |
| | Sen | Spec | ACC | AC2 | Sen | Spec | ACC | AC2 | Sen | Spec | ACC | AC2 | Sen | Spec | ACC | AC2 |
| MTA | 0.812 | 1.000 | 0.863 | 0.751 | 0.808 | 0.891 | 0.840 | 0.688 | 0.737 | 0.855 | 0.798 | 0.600 | 0.791 | 0.900 | 0.834 | 0.673 |
| GCA-F | 1.000 | 1.000 | 1.000 | 1.000 | 0.991 | 0.300 | 0.933 | 0.924 | 0.980 | 0.059 | 0.850 | 0.820 | 0.991 | 0.207 | 0.927 | 0.919 |
| PA | - | - | 0.992 | - | 0.975 | 1.000 | 0.975 | 0.974 | 0.966 | 0.000 | 0.950 | 0.947 | 0.977 | 0.333 | 0.972 | 0.971 |

Occupancy index score vs. consensus visual rating

| | | | | | | | | | | | | | | | | |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| TOC vs. MTA | 0.938 | 0.194 | 0.709 | 0.569 | 0.877 | 0.652 | 0.790 | 0.612 | 0.629 | 0.772 | 0.698 | 0.396 | 0.829 | 0.583 | 0.732 | 0.503 |
| HOC vs. MTA | 0.951 | 0.167 | 0.709 | 0.575 | 0.863 | 0.630 | 0.773 | 0.581 | 0.613 | 0.754 | 0.681 | 0.363 | 0.824 | 0.561 | 0.721 | 0.483 |
| FOC vs. GCA-F | 0.991 | 0.333 | 0.957 | 0.954 | 0.960 | 0.450 | 0.874 | 0.834 | 0.825 | 0.487 | 0.714 | 0.506 | 0.934 | 0.462 | 0.848 | 0.792 |
| POC vs. PA | 0.991 | 0.500 | 0.983 | 0.982 | 0.938 | 0.500 | 0.916 | 0.904 | 0.755 | 0.444 | 0.731 | 0.627 | 0.896 | 0.471 | 0.876 | 0.853 |

AD, Alzheimer's disease group; MCI, mild cognitive impairment; CN, cognitive normal group; Sen, Sensitivity; Spec, Specificity; ACC, Accuracy; AC2, Gwet's AC2 score; MTA, medial temporal atrophy score; GCA-F, global cortical atrophy score of the frontal lobe; PA, posterior atrophy score; HOC, hippocampus occupancy index; TOC, medial temporal lobe occupancy index; FOC, frontal lobe occupancy index; POC, parietal lobe occupancy index.

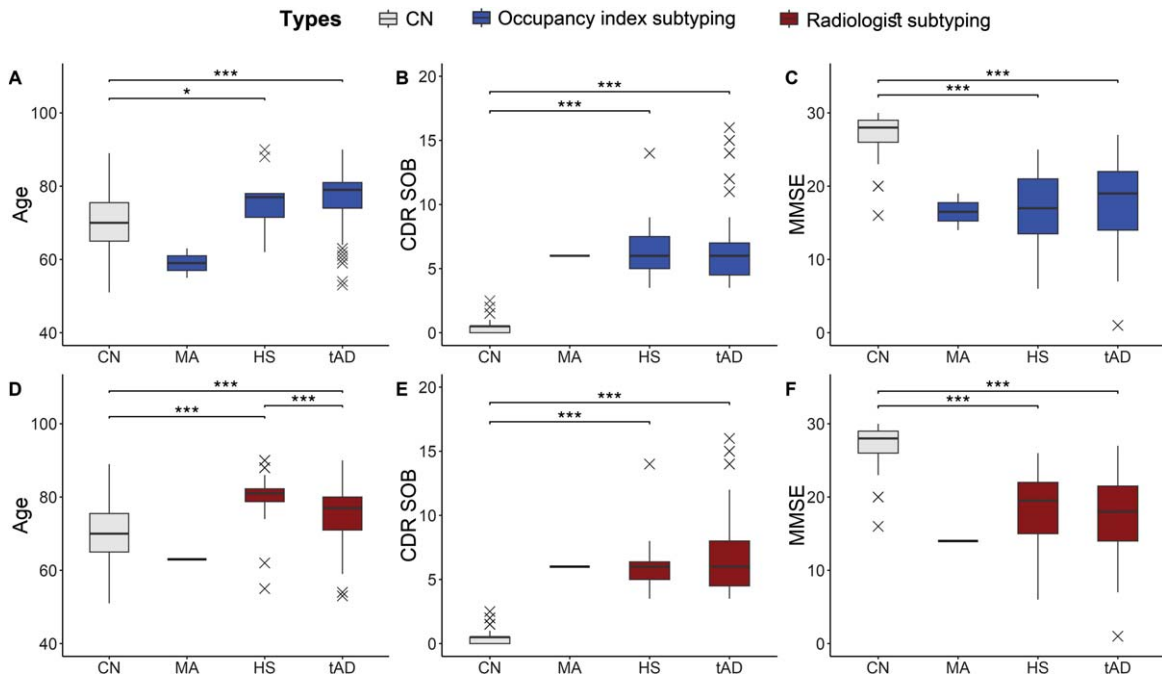


Fig. 4. Comparison of Alzheimer's disease (AD) subtype between visual assessment and automated scores. Age, CDR sum of boxes, and MMSE scores were compared between four AD subtypes based on the quantitative approach (A-C) and visual scoring approach (D-F). Occupancy and subtypes based on radiologists' assessment exhibit a similar trend for age (A, D), CDR SOB (B, E), and MMSE scores (C, F). The cognitively normal (CN) group is provided as a reference. ***significantly different with $p < 0.001$. *significantly different with $p < 0.05$. CN, Cognitive normal; MA, Minimal atrophy; HS, Hippocampal sparing; tAD, Typical Alzheimer's disease; CDR SOB, Clinical Dementia Rating Score Sum of Boxes.

DISCUSSION

We proposed an automated atrophy scoring scheme based on quantified volumetric information to provide objective and consistent scoring. In this study, our automated atrophy scoring scheme incorporating the inferior lateral ventricle size showed

moderate-to-good agreement with the radiologists' visual ratings and an enhanced correlation with age. Furthermore, we showed that the cutoff value for each atrophy score adjusted by age is better than that without adjustment, because natural aging accompanies brain atrophy to a certain degree. Our validation dataset demonstrated the usefulness of our automated

Table 5
Comparison of patients' demographic information according to occupancy index-based atrophy and radiologist visual scores

| | | CN | Subtyping by atrophy score | | |
|------------------------|---------------|-----------------|----------------------------|-------------------------|------------------------------|
| | | | Minimal Atrophy | Hippocampal sparing | Typical AD |
| <i>N</i> | <i>Auto.</i> | 119 | 2 | 11 | 108 |
| | <i>Expert</i> | | 1 | 36 | 84 |
| Age (y) | <i>Auto.</i> | 70.0 ± 7.5 | 59.0 ± 5.7 | 76.1 ± 8.4 ^a | 76.8 ± 7.5 ^a |
| | <i>Expert</i> | | 63.0 ± NA | 80.1 ± 6.7 | 75.1 ± 7.7 |
| Diagnosis (AD/MCI) | <i>Auto.</i> | – | 2/0 | 10/1 | 105/3 |
| | <i>Expert</i> | | 1/0 | 35/1 | 81/3 |
| Sex (M/F) (F %) | <i>Auto.</i> | 32/87 (73.1%) | 0/2 (100.0%) | 2/9 (81.8%) | 31/77 (71.3%) |
| | <i>Expert</i> | | 0/1 (100.0%) | 7/29 (80.6%) | 26/58 (69.0%) |
| APOE E4 (%) E4 : 0/1/2 | <i>Auto.</i> | 52/12/1 (20.0%) | 1/1/0 (50.0%) | 3/2/0 (40.0%) | 39/20/5 (39.0%) ^a |
| | <i>Expert</i> | | 0/1/0 (100.0%) | 10/3/1 (28.5%) | 33/19/4 (41.0%) |
| CDR SOB | <i>Auto.</i> | 0.5 ± 0.6 | 6.0 ± NA | 6.9 ± 3.5 ^a | 6.6 ± 2.9 ^a |
| | <i>Expert</i> | | 6.0 ± NA | 6.2 ± 2.1 ^a | 6.8 ± 3.2 ^a |
| MMSE | <i>Auto.</i> | 27.1 ± 2.4 | 16.5 ± 3.5 | 16.7 ± 6.1 ^a | 18.0 ± 5.1 ^a |
| | <i>Expert</i> | | 14.0 ± NA | 18.4 ± 4.8 ^a | 17.6 ± 5.3 ^a |
| HOC | <i>Auto.</i> | 0.919 ± 0.076 | 0.99 ± 0.003 | 0.932 ± 0.031 | 0.771 ± 0.117 |
| | <i>Expert</i> | | 0.988 ± NA | 0.849 ± 0.071 | 0.762 ± 0.130 |
| TOC | <i>Auto.</i> | 0.954 ± 0.046 | 0.995 ± 0.002 | 0.961 ± 0.019 | 0.852 ± 0.087 |
| | <i>Expert</i> | | 0.993 ± NA | 0.911 ± 0.044 | 0.843 ± 0.097 |
| FOC | <i>Auto.</i> | 0.995 ± 0.006 | 0.999 ± 0.0002 | 0.996 ± 0.002 | 0.982 ± 0.013 |
| | <i>Expert</i> | | 0.999 ± NA | 0.990 ± 0.006 | 0.980 ± 0.014 |
| POC | <i>Auto.</i> | 0.992 ± 0.008 | 0.999 ± 0.0003 | 0.994 ± 0.004 | 0.973 ± 0.018 |
| | <i>Expert</i> | | 0.999 ± NA | 0.985 ± 0.008 | 0.971 ± 0.020 |

CN, cognitively normal group; AD, Alzheimer's disease group; MCI, mild cognitive impairment; M/F, male/female; APOE E4, apolipoprotein E ε4 allele; CDR SOB, clinical dementia rating sum of boxes; MMSE, Mini-Mental State Examination; HOC, hippocampus occupancy index; TOC, medial temporal lobe occupancy index; FOC, frontal lobe occupancy index; POC, parietal lobe occupancy index. ^asignificant difference with CN ($p < 0.001$).

scoring approach in classifying AD subtypes based on MR-driven measurements. In addition, the results of this study demonstrated that the patients in the AD subtypes derived from our automated scoring method share similar demographic characteristics, and that our method shows reasonable classification results as compared to a visual scoring approach.

The highest atrophy scores determined by experts matched well the higher automated atrophy scores, which had lower volumes. The high scores provided by the radiologists consistently correlated with the automated scores, while the low visual scores showed some discrepancy due to the inclusion of high scores from the automated scoring. Although HOC presented slightly lower correspondence against visual scoring, both TOC and HOC, which are based on different ROIs, shared a similar classification pattern and could be used as MTA surrogates, as suggested by previous studies [2, 22].

In general, our automated atrophy scores corresponded well with the radiologists' scores. Moreover, our atrophy score was not significantly different from the radiologist-driven visual assessment score in each related region, that is, MTA, GCA-F, and PA. This finding supports the hypothesis that volume-driven

atrophy scores could provide objective surrogates for radiologists. This is consistent with previous studies reporting a strong correlation between visual scoring and brain measurements (of the cortical thickness and hippocampal volume) when classifying AD subtypes based on atrophy scores [2]. Typical AD showed broad atrophy patterns in the frontal, parietal, and temporal regions, including the hippocampus, while the hippocampal-sparing type showed a statistically significant difference only in hippocampal volume compared to the CN group [2]. Previous AD subtyping using the cortical thinning pattern [46] was also comparable to neuropsychological test results and clinical characteristics. Taken together, these results suggested that brain volume is highly related to cognition and clinical symptoms.

We also showed that an age-specific cutoff value, which was blinded to sex or any other clinical information, could be important in determining the degree of atrophy in the general population. This is because we utilized a large population dataset with ages ranging from 20 to 110 years. Furthermore, an age-specific cutoff value was also implemented using data including the heterogeneous cognitive states of CN, MCI, and AD. We believe that our proposed

automated atrophy scoring with the suggested age-specific cutoff value is easily generalizable to a larger population regardless of age, sex, and other clinical information.

Our results indicate that AD subtypes determined by volume-based atrophy scores showed trends similar to those of clinical AD categorized using visual atrophy rating scales. AD subtyping is mostly based on tauopathy and atrophy pattern observations with cognitive functional changes and clinical symptoms. In our study, participants classified in the typical AD type showed more severe atrophy in TOC and HOC than CN participants, compared with other subtypes. Both typical AD types, categorized either automatically or based on radiologists' assessment, had smaller volumes in the medial temporal lobe and hippocampal region than in the parietal or frontal lobe [47–49]. Furthermore, participants in the hippocampal-sparing type had a younger age distribution than typical AD participants [50]. This finding was similar to that for the clinical characteristics of traditional AD subtypes and suggests that contiguous AD subtypes related only to brain volume could be applicable without other clinical criteria.

Visual assessment may be labor-intensive with large inter- and intra-rater variations. In particular, the challenge of reliable visual scoring of minimal-to-moderate atrophy scores causes low inter-rater correspondence. In our dataset, the visual scoring results from two neuroradiologists showed moderate agreement. To incorporate more consistent visual scoring into our automated scoring system, we sought to derive consensus scoring results from the two radiologists. The final visual score was converted into a binary score. There is a high correlation between the concordance visual score and our data-driven atrophy score without any agreement. This implies that atrophy scoring with visual assessment requires a high ability from individual radiologists, and the automated volumetric score could support radiologists with adequate reliability.

Compared with PET, MRI has several advantages as it is inexpensive and does not require radiation exposure. In addition to tauopathy and neurodegeneration, a variety of biomarkers in AD are systematically connected, and studies of MR biomarkers need to be developed [51–55].

Limitations

One limitation of this study lies in the use of MR-based measurements for AD classification. MR-

based atrophy features could only be used for assessing the present situation or very-near future, but not for early detection or for prevention. The AD subtyping introduced in this study could be useful for categorizing participants with advanced disease status, such as late MCI and AD. In addition, our study had small sample sizes to validate the subtyping results from MR-based automated atrophy scoring. Our validation data only included three types (hippocampal-sparing, minimal atrophy, and typical AD) and did not include the limbic-predominant type, which was suggested to also exist by the MRI-based measurements [2, 46]. There were also only one or two participants with minimal atrophy at younger ages. Furthermore, disease progression or severity, two of the most important features in AD, could not be assessed due to the absence of longitudinal or amyloidosis and tauopathy datasets. Comparison with conventional biomarkers, namely CSF amyloid- β , p-tau, and t-tau levels, and amyloid, tau, and FDG PET, should be pursued in future studies. Age-specific cutoff values were comprehensively considered using large-scale data from previous studies [21, 22, 34, 56–58], which were heterogeneous. In previous studies, age thresholds for the MTA score were considered at 75 or 80 years [22, 56, 57]; however, Korean [21] and Chinese [58] studies described younger and more precise age ranges. In particular, the government's public health policy on dementia in Korea was systematically implemented in 2008 [59] based on the prominent public health care system [60] that could affect the careful observation of clinicians and the attention of elderly people. Of course, changes in brain volume could be affected by differences in race or environment, and future studies should compare more diverse populations. Finally, only two radiologists undertook the visual assessment, and there were several difficulties, including the high time cost, cumbersome visual assessment task in each region and across scales, and the uncertainty among radiologists.

CONCLUSION

This study suggested that our new automated atrophy scoring system could offer comparable results to those obtained from human raters with multiple and binary scores in the medial temporal, parietal, and frontal lobes. In the future, an AD subtype study could be conducted using a larger number of population datasets.

AUTHOR CONTRIBUTIONS

Yeong Sim Choe (Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft); Regina E.Y. Kim (Conceptualization; Formal analysis; Investigation; Methodology; Project administration; Validation; Visualization; Writing – original draft; Writing – review & editing); Hye Weon Kim (Investigation; Writing – original draft; Writing – review & editing); JeeYoung Kim (Conceptualization; Data curation; Methodology; Resources; Validation); Hyunji Lee (Formal analysis; Investigation; Methodology; Writing – original draft); Min Kyoung Lee (Data curation; Investigation; Resources); Minho Lee (Resources; Software); Keun You Kim (Data curation; Investigation; Resources); Se-Hong Kim (Resources); Ji-hoon Kim (Resources); Jun-Young Lee (Project administration; Resources); Eosu Kim (Project administration; Resources); Donghyeon Kim (Funding acquisition; Project administration; Software; Supervision); Hyun Kook Lim (Funding acquisition; Project administration; Resources; Supervision).

ACKNOWLEDGMENTS

We would like to thank our colleagues at Neurophet Inc. for their feedback and support throughout the research process.

FUNDING

This research was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) and the Korea Dementia Research Center (KDRC), funded by the Ministry of Health & Welfare and the Ministry of Science and ICT, Republic of Korea (grant number: HU20C0315).

CONFLICT OF INTEREST

This article utilized the segmentation tool, AQUA developed by Neurophet Research Institute, Inc., Seoul, South Korea. Y.S.C, R.E.K, H.W.K, H.L, M.L, D.K, and H.K.L are employees of Neurophet Inc. All other authors have no conflict of interest to report.

DATA AVAILABILITY

The data supporting the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/ADR-230105>.

REFERENCES

- [1] Lombardi G, Crescioli G, Cavado E, Lucenteforte E, Casazza G, Bellatorre AG, Lista C, Costantino G, Frisoni G, Virgili G, Filippini G (2020) Structural magnetic resonance imaging for the early diagnosis of dementia due to Alzheimer's disease in people with mild cognitive impairment. *Cochrane Database Syst Rev* **3**, CD009628.
- [2] Ferreira D, Verhagen C, Hernández-Cabrera JA, Cavallin L, Guo CJ, Ekman U, Muehlboeck JS, Simmons A, Barroso J, Wahlund LO, Westman E (2017) Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: Longitudinal trajectories and clinical applications. *Sci Rep* **7**, 46263.
- [3] Silhan D, Bartos A, Mrzilkova J, Pashkovska O, Ibrahim I, Tintera J (2020) The parietal atrophy score on brain magnetic resonance imaging is a reliable visual scale. *Curr Alzheimer Res* **17**, 534-539.
- [4] Villa C, Lavitrano M, Salvatore E, Combi R (2020) Molecular and imaging biomarkers in Alzheimer's disease: A focus on recent insights. *J Pers Med* **10**, 1-32.
- [5] Wolk DA, Das SR, Mueller SG, Weiner MW, Yushkevich PA (2017) Medial temporal lobe subregional morphometry using high resolution MRI in Alzheimer's disease. *Neurobiol Aging* **49**, 204-213.
- [6] Scheltens P, Pasquier F, Weerts JGE, Barkhof F, Leys D (1997) Qualitative assessment of cerebral atrophy on MRI. *Eur Neurol* **37**, 95-99.
- [7] Ferreira D, Cavallin L, Granberg T, Lindberg O, Aguilar C, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Soinen H, Lovestone S, Simmons A, Wahlund LO, Westman E (2016) Quantitative validation of a visual rating scale for frontal atrophy: associations with clinical status, APOE e4, CSF biomarkers and cognition. *Eur Radiol* **26**, 2597-2610.
- [8] Koedam ELGE, Lehmann M, van der Flier WM, Scheltens P, Pijnenburg YAL, Fox N, Barkhof F, Wattjes MP (2011) Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol* **21**, 2618-2625.
- [9] Whitwell JL, Dickson DW, Murray ME, Weigand SD, Tosakulwong N, Senjem ML, Knopman DS, Boeve BF, Parisi JE, Petersen RC, Jack CR, Josephs KA (2012) Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *Lancet Neurol* **11**, 868-877.
- [10] Whitwell JL, Jack CR, Przybelski SA, Parisi JE, Senjem ML, Boeve BF, Knopman DS, Petersen RC, Dickson DW, Josephs KA (2011) Temporoparietal atrophy: A marker of

- AD pathology independent of clinical diagnosis. *Neurobiol Aging* **32**, 1531-1541.
- [11] Scheltens NME, Galindo-Garre F, Pijnenburg YAL, van der Vlies AE, Smits LL, Koene T, Teunissen CE, Barkhof F, Wattjes MP, Scheltens P, van der Flier WM (2016) The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. *J Neurol Neurosurg Psychiatry* **87**, 235-243.
- [12] Shiino A, Watanabe T, Maeda K, Kotani E, Akiguchi I, Matsuda M (2006) Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease. *Neuroimage* **33**, 17-26.
- [13] Dubois B, Hospital D, Dubois B, Feldman HH, Jacova C, rey Cummings JL, DeKosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D, Gauthier S, Hampel H, Jicha GA, Meguro K, Pasquier F, Robert P, Rossor M, Salloway S, Sarazin M, de Souza LC, Stern Y, Visser PJ, Scheltens P (2010) Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol* **9**, 1118-1127.
- [14] Peter J, Abdulkadir A, Kaller C, Kümmerer D, Hüll M, Vach W, Klöppel S (2014) Subgroups of Alzheimer's disease: Stability of empirical clusters over time. *J Alzheimers Dis* **42**, 651-661.
- [15] Harper L, Fumagalli GG, Barkhof F, Scheltens P, O'Brien JT, Bouwman F, Burton EJ, Rohrer JD, Fox NC, Ridgway GR, Schott JM (2016) MRI visual rating scales in the diagnosis of dementia: Evaluation in 184 post-mortem confirmed cases. *Brain* **139**, 1211-1225.
- [16] Stopford CL, Snowden JS, Thompson JC, Neary D (2008) Variability in cognitive presentation of Alzheimer's disease. *Cortex* **44**, 185-195.
- [17] Davidson JE, Irizarry MC, Bray BC, Wetten S, Galwey N, Gibson R, Borrie M, Delisle R, Feldman HH, Hsiung GY, Fornazzari L, Gauthier S, Guzman D, Loy-English I, Keren R, Kertesz A, George-Hyslop PS, Wherrett J, Monesch AU (2010) An exploration of cognitive subgroups in Alzheimer's disease. *J Int Neuropsychol Soc* **16**, 233-243.
- [18] Chun MY, Kim GH, Park HK, Yang DW, Kim S, Choi SH, Jeong JH (2022) Predictive scale for amyloid pet positivity based on clinical and MRI variables in patients with amnesic mild cognitive impairment. *J Clin Med* **11**, 3433.
- [19] Murray E, Ross OA, Dickson DW, Dickson DW, Murray ME, Graff -Radford NR, Ross OA, Petersen RC, Duara R (2011) Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* **10**, 785-796.
- [20] Jellinger KA (2021) Pathobiological subtypes of Alzheimer disease. *Dement Geriatr Cogn Disord* **49**, 321-333.
- [21] Choi GS, Kim GH, Choi JH, Hwang J, Kwon E, Lee SA, Kong KA, Kang HJ, Yoon B, Kim BC, Yang DW, Na DL, Kim EJ, Na HR, Han HJ, Lee JH, Kim JH, Lee KY, Park KH, Park KW, Kim S, Han SH, Kim SY, Yoon SJ, Moon SY, Youn YC, Choi SH, Jeong JH (2018) Age-specific cutoff scores on a T1-weighted axial medial temporal-lobe atrophy visual rating scale in Alzheimer's disease using clinical research center for dementia of South Korea data. *J Clin Neurol* **14**, 27-282.
- [22] Velickaite V, Ferreira D, Lind L, Ahlström H, Kilander L, Westman E, Larsson EM (2020) Visual rating versus volumetry of regional brain atrophy and longitudinal changes over a 5-year period in an elderly population. *Brain Behav* **10**, e01662.
- [23] Pereira JB, Cavallin L, Spulber G, Aguilar C, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Spenger C, Aarsland D, Lovestone S, Simmons A, Wahlund LO, Westman E (2014) Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cutoffs. *J Intern Med* **275**, 317-330.
- [24] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease. *Neurology* **34**, 939.
- [25] Morris JC (1993) The clinical dementia rating (cdr): Current version and scoring rules. *Neurology* **43**, 2412-2414.
- [26] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment clinical characterization and outcome. *Arch Neurol* **56**, 304-308.
- [27] McEvoy LK, Brewer JB (2012) Biomarkers for the clinical evaluation of the cognitively impaired elderly: Amyloid is not enough. *Imaging Med* **4**, 343-357.
- [28] Kang DW, Wang SM, Um YH, Na HR, Kim NY, Lee CU, Lim HK (2021) Distinctive association of the functional connectivity of the posterior cingulate cortex on memory performances in early and late amnesic mild cognitive impairment patients. *Front Aging Neurosci* **13**, 696735.
- [29] Kang DW, Wang SM, Um YH, Kim NY, Lee CU, Lim HK (2022) associations between sub-threshold amyloid- β deposition, cortical volume, and cognitive function modulated by APOE $\epsilon 4$ carrier status in cognitively normal older adults. *J Alzheimers Dis* **89**, 1003-1016.
- [30] Lee M, Kim J, Kim REY, Kim HG, Oh SW, Lee MK, Wang SM, Kim NY, Kang DW, Rieu Z, Yong JH, Kim D, Lim HK (2020) Split-attention u-net: A fully convolutional network for robust multi-label segmentation from brain MRI. *Brain Sci* **10**, 1-22.
- [31] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968-980.
- [32] Henschel L, Conjeti S, Estrada S, Diers K, Fischl B, Reuter M (2020) FastSurfer - A fast and accurate deep learning based neuroimaging pipeline. *Neuroimage* **219**, 117012.
- [33] Roy AG, Conjeti S, Navab N, Wachinger C (2019) QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *Neuroimage* **186**, 713-727.
- [34] Ferreira D, Cavallin L, Larsson EM, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, Kloszewska I, Soininen H, Lovestone S, Simmons A, Wahlund LO, Westman E (2015) Practical cutoffs for visual rating scales of medial temporal, frontal and posterior atrophy in Alzheimer's disease and mild cognitive impairment. *J Intern Med* **278**, 277-290.
- [35] Zhu H, Lu H, Wang F, Liu S, Shi Z, Gan J, Du X, Yang Y, Li D, Wang L, Ji Y (2022) Characteristics of cortical atrophy and white matter lesions between dementia with lewy bodies and Alzheimer's disease: A case-control study. *Front Neurol* **12**, 779344.
- [36] Armstrong RA (2014) When to use the Bonferroni correction. *Ophthalmic Physiol Opt* **34**, 502-508.
- [37] Gwet KL (2014) Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters (4th edition). *Advanced Analytics, LLC*.
- [38] Fleiss JL (1971) Measuring nominal scale agreement among many raters. *Psychol Bull* **76**, 378-382.
- [39] Viswanathan M, Berkman ND (2012) Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* **65**, 163-178.
- [40] Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL (2013) A comparison of Cohen's Kappa and Gwet's AC1

- when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol* **13**, 61.
- [41] Landis JR, Koch GG (1977) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* **33**, 363-374.
- [42] McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochem Med* **22**, 276-282.
- [43] Kuhn M (2022) Caret: Classification and regression training. R package version 6.0-91.
- [44] Gamer M, Lemon J, Fellows I, Singh P (2019) irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1.
- [45] Gwet KL (2019) irrCAC: Computing chance-corrected agreement coefficients (CAC). R package version 1.0.
- [46] Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, Ye BS, Yoon CW, Kim HJ, Chin J, Park KH, Heilman KM, Na DL (2014) Anatomical heterogeneity of Alzheimer disease based on cortical thickness on MRIs. *Neurology* **83**, 1936-1944.
- [47] Henneman WJP, Sluimer JD, Barnes J, van der Flier WM, Sluimer IC, Fox NC, Scheltens FP, Vrenken H, Barkhof F (2009) Hippocampal atrophy rates in Alzheimer disease added value over whole brain volume measures. *Neurology* **72**, 999-1007.
- [48] Chan D, Fox NC, Scahill RI, Crum WR, Whitwell JL, Leschziner G, Rossor AM, Stevens JM, Cipolotti L, Rossor MN (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann Neurol* **49**, 433-442.
- [49] Ferreira D, Pereira JB, Volpe G, Westman E (2019) Subtypes of Alzheimer's disease display distinct network abnormalities extending beyond their pattern of brain atrophy. *Front Neurol* **10**, 524.
- [50] Ferreira D, Nordberg A, Westman E (2020) Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology* **94**, 436-448.
- [51] Emrani S, Lamar M, Price CC, Wasserman V, Matusz E, Au R, Swenson R, Nagele R, Heilman KM, Libon DJ (2020) Alzheimer's/vascular spectrum dementia: classification in addition to diagnosis. *J Alzheimers Dis* **73**, 63-71.
- [52] Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C (2020) Disentangling heterogeneity in alzheimer's disease and related dementias using data-driven methods. *Biol Psychiatry* **88**, 70-82.
- [53] Guerreiro RJ, Gustafson DR, Hardy J (2012) The genetic architecture of Alzheimer's disease: Beyond APP, PSEN5 and APOE. *Neurobiol Aging* **33**, 437-456.
- [54] Cole MA, Seabrook GR (2020) On the horizon—the value and promise of the global pipeline of Alzheimer's disease therapeutics. *Alzheimers Dement* **6**, e12009.
- [55] Lukiw WJ, Vergallo A, Lista S, Hampel H, Zhao Y (2020) Biomarkers for Alzheimer's disease (Ad) and the application of precision medicine. *J Pers Med* **10**, 1-12.
- [56] Velickaite V, Ferreira D, Cavallin L, Lind L, Ahlström H, Kilander L, Westman E, Larsson EM (2018) Medial temporal lobe atrophy ratings in a large 75-year-old population-based cohort: gender-corrected and education-corrected normative data. *Eur Radiol* **28**, 1739-1747.
- [57] Furtner J, Prayer D (2021) Neuroimaging in dementia. *Wiener Medizinische Wochenschrift* **171**, 274-281.
- [58] Wei M, Shi J, Ni J, Zhang X, Li T, Chen Z, Zhou M, Zhang L, Tan Z, Wang Y, Tian J (2019) A new age-related cutoff of medial temporal atrophy scale on MRI improving the diagnostic accuracy of neurodegeneration due to Alzheimer's disease in a Chinese population. *BMC Geriatr* **19**, 59.
- [59] Lee DW, Seong SJ (2018) Korean national dementia plans: From 1st to 3rd. *J Korean Med Assoc* **61**, 298-303.
- [60] OECD (2020) OECD Reviews of Public Health: Korea: A Healthier Tomorrow, *OECD Publishing, Paris*.