ORIGINAL ARTICLE
https://doi.org/10.30773/pi.2020.0304

Print ISSN 1738-3684 / On-line ISSN 1976-3026
OPEN ACCESS

Development of Random Forest Algorithm Based Prediction Model of Alzheimer's Disease Using Neurodegeneration Pattern

JeeYoung Kim[1]*, Minho Lee[2]*, Min Kyoung Lee[3], Sheng-Min Wang[4], Nak-Young Kim[4], Dong Woo Kang[5], Yoo Hyun Um[6], Hae-Ran Na[4], Young Sup Woo[4], Chang Uk Lee[5], Won-Myong Bahk[4], Donghyeon Kim[2] ✉, and Hyun Kook Lim[4] ✉

[1]Department of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
[2]Research Institute, NEUROPHET Inc., Seoul, Korea
[3]Department of Radiology, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea
[4]Department of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea
[5]Department of Psychiatry, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea
[6]Department of Psychiatry, St. Vincent's Hospital Seoul, College of Medicine, The Catholic University of Korea, Suwon, Korea

Objective Alzheimer's disease (AD) is the most common type of dementia and the prevalence rapidly increased as the elderly population increased worldwide. In the contemporary model of AD, it is regarded as a disease continuum involving preclinical stage to severe dementia. For accurate diagnosis and disease monitoring, objective index reflecting structural change of brain is needed to correctly assess a patient's severity of neurodegeneration independent from the patient's clinical symptoms. The main aim of this paper is to develop a random forest (RF) algorithm-based prediction model of AD using structural magnetic resonance imaging (MRI).

Methods We evaluated diagnostic accuracy and performance of our RF based prediction model using newly developed brain segmentation method compared with the Freesurfer's which is a commonly used segmentation software.

Results Our RF model showed high diagnostic accuracy for differentiating healthy controls from AD and mild cognitive impairment (MCI) using structural MRI, patient characteristics, and cognitive function (HC vs. AD 93.5%, AUC 0.99; HC vs. MCI 80.8%, AUC 0.88). Moreover, segmentation processing time of our algorithm (<5 minutes) was much shorter than of Freesurfer's (6–8 hours).

Conclusion Our RF model might be an effective automatic brain segmentation tool which can be easily applied in real clinical practice.

Psychiatry Investig 2021;18(1):69-79

Key Words Random forest, Alzheimer's disease, Mild cognitive impairment, Segmentation, MRI.

INTRODUCTION

Dementia is a clinical syndrome characterized by chronic and progressive cognitive decline, behavior disturbance, and loss of daily functioning.[1] The prevalence of dementia increased rapidly with the rise of aging population. Globally, around 50 million people have dementia, and nearly 10 million new cases are confirmed every year. More than 70 causes of dementia are known, and Alzheimer's disease (AD) is the most common type of dementia accounting more for than 50−70% of all cases.[2]

AD can be definitively diagnosed only after death, by linking clinical measures with histopathological evidence of amyloid plaques and hyperphosphorylated tau tangles in postmortem brain.[3] Thus, diagnosis of AD still resides primarily on clinical decision, which is based on evaluating a patient's cognitive dysfunction, behavior and psychological symptom, and functional impairment.[4] However, studies showed that deposition of cerebral β-amyloid (Aβ), hyperphosphorylation of tau protein, and neurodegeneration of cerebral cortex begin decades before the onset of clinical symptoms of dementia.[5] Recent advance in diagnostic technologies have enabled us to

Received: August 14, 2020 **Revised:** October 10, 2020
Accepted: October 25, 2020

✉ **Correspondence:** Hyun Kook Lim, MD, PhD
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: +82-2-3779-1048, **Fax:** +82-2-780-6577, **E-mail:** drblues@catholic.ac.kr

✉ **Correspondence:** Donghyeon Kim, PhD
Chief Technology Officer (CTO), Research Institute, NEUROPHET Inc., 175 Yeoksam-ro, Gangnam-gu, Seoul 06247, Republic of Korea
Tel: +82-70-5223-4414, **Fax:** +82-2-6954-7972
E-mail: donghyeon.kim@neurophet.com

*These authors contributed equally to this work.

assess or measure such AD pathologies using in vivo biomarkers, which could play an important role as diagnostic tools of AD by corresponding with post-mortem histopathological findings.[6] In these regards, the National Institute of Aging and Alzheimer's Disease Association (NIA-AA) recently proposed a newer biomarker-based definition of AD and staging of the disease labeled as ATN system.[7] For a more accurate characterization of AD trajectory, the ATN staging system is designed with quantification of three core biomarkers including Aβ deposition (A), pathologic tau (T), and neurodegeneration (N). Since, cerebral accumulation of pathologic proteins may result in neurodegenerative change before clinical symptom presentations, several neurodegeneration biomarkers using structural MRI such as lowered hippocampal formation and medial temporal lobe volume, and decreased cortical thickness are increasingly suggested to be effective in classifying patient with AD trajectory.[7-11] In addition, recent development of neuroimaging initiatives, which contain larger number of databases freely accessible to researchers and clinicians, enabled the development of and urged for the application of automated whole brain degenerative pattern recognition for early detection of AD.[12]

Machine learning (ML) is a technique which is useful in recognizing and extracting meaningful patterns from medical images.[13] Previous ML studies using structural MRI showed that it can efficiently classify subjects within AD continuum including cognitive normal adults, mild cognitive impairment (MCI), and AD.[14-19] Previous ML studies in neuroimaging data mostly relied on a single classifier such as support vector machine and linear discriminant analysis.[20,21] However, recently developed ensemble ML algorithms such as Random Forest (RF) showed better performance than single ML classifier algorithms in diagnosis of various neurological diseases.[22] In addition, RF is known to have additional advantages when compared to other ML methods.[23] First, it has less risk of overfitting. Second, it is considered to be more stable even in the presence of outliers and in the very high dimensional parameter spaces. An intrinsic feature selection step, which gives important values to each feature to reduce the variables space, is another important characteristic of RF.[22,23] Previous studies using structural MRI from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database showed that ML with RF model showed mean accuracies for the binary classifications of AD vs. healthy controls (HC) and MCI vs. HC were 90.3% and 81.3%, respectively.[22,24] However, the structural images used in the study were collected from 1.5T MRI machine. Thus, replication of RF algorithm to a more variety of structural images acquired from 3T MRI is needed to improve its diagnostic validity and test-retest reliability. Lastly, most of the previous studies used brain segmentation algorithm from the Freesurfer program which requires 6–10 hours per subject. This long segmentation time hindered application of RF based classifying algorithm for AD to real clinical settings.[23]

In this study, we first developed a fast processing automated segmentation method based on deep learning algorithm and then compared accuracy of our segmentation results with that of the segmentation done using Freesurfer in subjects with AD, MCI, and HC. In addition, we compared diagnostic accuracy of RF based classification model of AD using our newly developed brain segmentation methods with that of Freesurfer's.

METHODS

Subjects

A total 647 subjects (AD=100, MCI=86, and NC=461) were included in this study. They were recruited from volunteers of the Catholic Aging Brain Imaging (CABI) database, which holds brain MRI scans of outpatients at the Catholic Brain Health Center, Yeouido St Mary's Hospital, the Catholic University of Korea from 2017 to 2019. Each participant provided written informed consent. Study procedures were approved by an Institutional Review Board (IRB Number: SC18RNDI0070) of Yeouido St Mary's Hospital, the Catholic University of Korea. The inclusion criteria of the NC group were as follows: 1) subjects aged ≥60 years; 2) Mini-Mental Status Examination score ≥27; and 3) Overall clinical Dementia Rating score=0. Following subjects meeting Petersen's criteria of MCI were included: 1) objective cognitive impairment for age, education, and gender; 2) memory complaint, preferably provided by an informant; 3) essentially conserved general cognitive function; 4) largely intact activities of daily living; and 5) not demented.[25] All MCI patients had an overall CDR score of 0.5. Objective cognitive impairment was defined as a performance score of 1.5 standard deviation (SD) below the each age-, education- and gender-specific normative means in at least one of the nine cognitive tests included in the Korean version of Consortium to Establish a Registry for Alzheimer's disease (CERAD-K) neuropsychological battery.[26] In addition, the inclusion criteria of the AD group were as follows: 1) subjects aged ≥60 years; 2) the National Institute of Neurological and Communication Disorders and Stroke/Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) diagnosis of probable AD; and 3) overall CDR score of more than 1.0.[7] Subjects with any neurological, psychiatric and unstable medical conditions were excluded. Table 1 shows the baseline demographic data for the three groups.

Sub-sampling

Number of subjects included among three groups (NC: 461, MCI: 86, AD: 100) were not equally balanced. Thus, when performing 3-fold cross validation, among the data number

Table 1. Demographic and clinical characteristics of the study participants

	NC (N=461)	MCI (N=86)	AD (N=100)	p value
Age (years±SD)	70.36±8.66	78.29±6.53	80.02±8.10	<0.001
Education (years±SD)	11.05±4.92	8.98±5.34	8.75±5.41	<0.001
Gender (M:F)	147:314	15:71	29:71	-
CDR (SD)	0.21±0.33	0.51±0.06	1.40±0.58	<0.001
CERAD-K battery (SD)				
VF	14.40±4.32	9.67±3.13	6.15±4.29	<0.001
BNT	11.97±2.25	9.29±2.88	6.06±3.66	<0.001
MMSE	27.21±2.39	22.86±3.47	15.57±5.20	<0.001
WLM	18.08±4.01	12.65±3.47	6.84±3.90	<0.001
CP	10.44±1.15	9.43±1.85	7.61±2.90	<0.001
WLR	5.58±2.23	2.16±1.70	0.63±0.92	<0.001
WLRc	8.84±1.58	6.45±2.68	2.76±2.54	<0.001
CR	6.49±3.25	2.06±1.98	0.84±1.38	<0.001

SD: standard deviation, NS: not significant, CDR: Clinical Dementia Rating, CERAD-K: the Korean version of Consortium to Establish a Registry for Alzheimer's Disease, VF: verbal fluency, BNT: 15-item Boston Naming Test, MMSE: Mini Mental Status Examination, WLM: word list memory, CP: constructional praxis, WLR: word list recall, WLRc: word list recognition, CR: constructional recall

of binary classes to classify the validation set, we first divided the group containing the smallest sample size by 3 (and the number of samples for the comparison groups was matched accordingly [i.e for MCI: 86 vs. AD: 100; 86/3=rounded up to 28 yielding MCI (28) vs. AD (28), HC (29) vs. MCI (29), and HC (33) vs. AD (33)].

MR image acquisition

Imaging data were collected at the Department of Radiology, Yeouido St Mary's Hospital, The Catholic University of Korea, using a 3T Siemens MAGETOM Skyra machine and eight channel Siemens head coil (Siemens Medical Solutions, Erlangen, Germany). The parameters used for the T1-weighted volumetric magnetization-prepared rapid gradient echo scan sequences were TE=2.6 ms, TR=1,940 ms, inversion time= 979 ms, FOV=230 mm, matrix=256×256, and voxel size= 1.0×1.0×1.0 mm^3.

Preprocessing and features extraction

T1-weighted MRIs were processed for an automated classification of HC, MCI and AD. MRIs were preprocessed by the Freesurfer (version 6) and U-Net++ deep learning-based segmentation processing.[27] We used the Freesurfer and our deep learning method for extracting numerical data into a table format (Figure 1). The detailed algorithm for our deep learning-based segmentation methods are described in the supplemental materials. The set of 106 layer sub-volume-based features used for the training procedure are also described in the Table 2.

Problem formulation

Our classification model was based on the RF method and its operational capabilities. In terms of the model, we performed feature importance using the Gini impurity index, which is a type of feature importance measurement mainly used in RF. The background information of the concerned methodologies and description of each classification model that we employed are described in detail elsewhere.[22,28]

Random forest

RF is a popular machine learning approach used in regression, classification and other tasks.[28] The method involves construction of the decision trees, and randomness is utilized in the following ways (Figure 2). First, respective decision tree is built using a different bootstrap sample. Second, during the building of respective decision tree, node split involves the random selection of a subset of variables based on which the best split is determined and used. For the prediction of unknown cases, the decisions of the constructed trees are aggregated by utilizing majority voting for classification and averaging for regression tasks. Operational feature of RF is its natural ability to supply a ranking of the importance of variables in a regression or classification task. This can be achieved in two ways. The first one is based on statistical permutation tests, while the second, which is used in this study, is based on Gini impurity index. The Gini impurity is calculated at all node split during the building of a decision tree in an RF model and is employed for measuring the quality of the split in terms of dividing the samples of the different classes in the considered node. For a variable, the Gini impurity index is calculated as in the follow-

Table 2. Summary of brain sub-volumes

Left	Right	Center
Left-Cerebral-White-Matter	Right-Cerebral-White-Matter	Posterior-Corpus-Callosum
Left-Lateral-Ventricle	Right-Lateral-Ventricle	Mid-Posterior-Corpus-Callosum
Left-Inferior-Lateral-Ventricle	Right-Inferior-Lateral-Ventricle	Central-Corpus-Callosum
Left-Cerebellum-White-Matter	Right-Cerebellum-White-Matter	Mid-Anterior-Corpus-Callosum
Left-Cerebellum-Cortex	Right-Cerebellum-Cortex	Anterior-Corpus-Callosum
Left-Thalamus	Right-Thalamus	3rd-Ventricle
Left-Caudate	Right-Caudate	4th-Ventricle
Left-Putamen	Right-Putamen	
Left-Pallidum	Right-Pallidum	
Left-Hippocampus	Right-Hippocampus	
Left-Amygdala	Right-Amygdala	
Left-Accumbens-area	Right-Accumbens-area	
Left-VentralDC	Right-VentralDC	
Left-choroid-plexus	Right-choroid-plexus	
Left WM-hypointensities	Right WM-hypointensities	
ctx-Left-bankssts	ctx-Right-bankssts	
ctx-Left-caudal-anterior-cingulate	ctx-Right-caudal-anterior-cingulate	
ctx-Left-caudal-middle-frontal	ctx-Right-caudal-middle-frontal	
ctx-Left-cuneus	ctx-Right-cuneus	
ctx-Left-entorhinal	ctx-Right-entorhinal	
ctx-Left-fusiform	ctx-Right-fusiform	
ctx-Left-inferior-parietal	ctx-Right-inferior-parietal	
ctx-Left-inferior-temporal	ctx-Right-inferior-temporal	
ctx-Left-isthmus-cingulate	ctx-Right-isthmus-cingulate	
ctx-Left-lateral-occipital	ctx-Right-lateral-occipital	
ctx-Left-lateral-orbito-frontal	ctx-Right-lateral-orbito-frontal	
ctx-Left-lingual	ctx-Right-lingual	
ctx-Left-medial-orbito-frontal	ctx-Right-medial-orbito-frontal	
ctx-Left-middle-temporal	ctx-Right-middle-temporal	
ctx-Left-parahippocampal	ctx-Right-parahippocampal	
ctx-Left-paracentral	ctx-Right-paracentral	
ctx-Left-pars-opercularis	ctx-Right-pars-opercularis	
ctx-Left-pars-orbitalis	ctx-Right-pars-orbitalis	
ctx-Left-pars-triangularis	ctx-Right-pars-triangularis	
ctx-Left-pericalcarine	ctx-Right-pericalcarine	
ctx-Left-postcentral	ctx-Right-postcentral	
ctx-Left-posterior-cingulate	ctx-Right-posterior-cingulate	
ctx-Left-precentral	ctx-Right-precentral	
ctx-Left-precuneus	ctx-Right-precuneus	
ctx-Left-rostral-anterior-cingulate	ctx-Right-rostral-anterior-cingulate	
ctx-Left-rostral-middle-frontal	ctx-Right-rostral-middle-frontal	
ctx-Left-superior-frontal	ctx-Right-superior-frontal	
ctx-Left-superior-parietal	ctx-Right-superior-parietal	
ctx-Left-superior-temporal	ctx-Right-superior-temporal	
ctx-Left-supramarginal	ctx-Right-supramarginal	
ctx-Left-frontal-pole	ctx-Right-frontal-pole	
ctx-Left-temporal-pole	ctx-Right-temporal-pole	
ctx-Left-transverse-temporal	ctx-Right-transverse-temporal	
ctx-Left-insula	ctx-Right-insula	

ctx: cortex

ing equation:

$$I_G(n)=1-\sum_{i=1}^{J}(p_i)^2$$

Where the node n is 1 minus the sum over all the classes J (for a binary classification task this is two) of the fraction of examples in each class p_i squared. For a given node split, the values of the Gini impurity index for the two resulting nodes are less than the value for the parent node. If we sum the Gini impurity decreases for 3 each variable in a dataset over all trees in a RF model, we get the corresponding Gini importance value for each variable, which be used for the feature selection as a result.

Modes description

In terms of the features of the classification model, it involved the training of a RF classifier on the whole feature set, as well as feature selection by means of the Gini importance measure, which provided the final feature subset that was utilized for retraining the RF model. Finally, for the prediction of unknown data based on the outputs of the RF model. In the case of ties, the class with the highest probability estimate was selected as the final prediction. RF model is developed using the RandomForestClassifier of Python-based scikit-learn library package.[29]

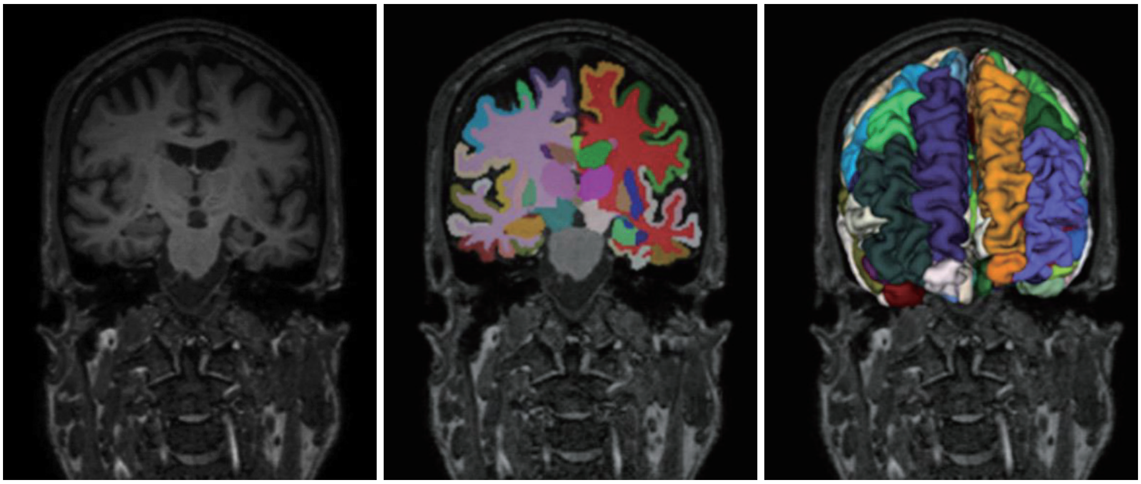

Figure 1. 106-layer brain segmentation results.

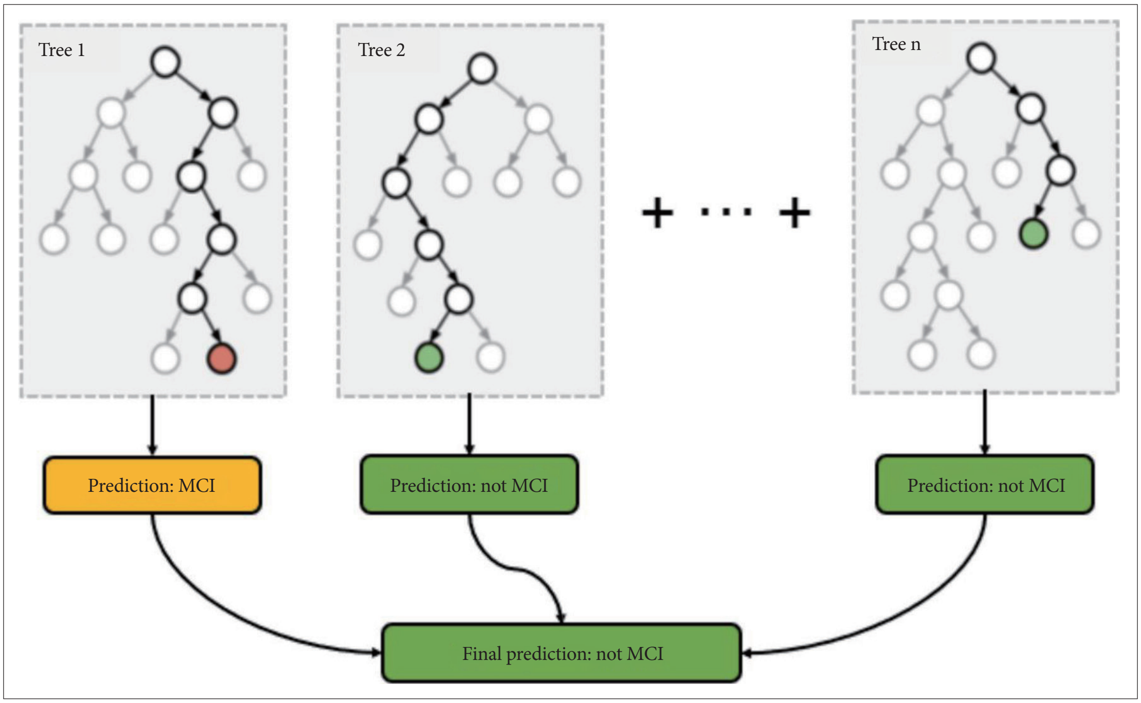

Figure 2. A flowchart that describes the RF classifier mode. MCI: mild cognitive impairment.

RESULTS

Baseline demographic and clinical data

Table 1 shows the baseline demographic data for the three groups. All variables were normally distributed. There were no significant differences in age, gender and education among the groups other than the neuropsychological data.

Performance of the classification

With regard to the RF parameters that we utilized, the number of trees for each RF model was set empirically, while for each RF model and for each node split during the growing of a tree, the number k of the subset of variables used to decide the best split was set based on repeated 3 fold cross-validation that was performed using the Python package.[29] Thus, the following parameter values were used for the RF model.

RF Model: Number of trees=1,000, k=3

Regarding the threshold values for the Gini importance measure during the feature selection task in the RF model, the following values were used: 0.5 for RF Model. We evaluated the following step; HC vs. MCI, HC vs. A, DMCI vs. AD

We used two types of features (with the use of the Gini importance measure) for RF model. One included cortical and subcortical volume information and basic patient characteristics including age and gender (sub-features), while another included cognitive functioning in addition to cerebral volume information and basic patient characteristics (all-features). The performance of binary classification is shown in Table 3.

The performance for diagnosis prediction is shown in Table 3 and Figure 3. The matrix summarizes a total of 180 test samples in the cross-validation set with fold-3. Our segmentation method showed average accuracy for predicting diagnoses of HC vs. MCI was 71.5% (AUC=0.81) in sub-features is and 80.8% (AUC=0.88) in all-features, for diagnoses of HC vs. AD the accuracy was 84.4% (AUC=0.92) in sub-features and 93.5% (AUC=0.99) in all-features, and for diagnoses of MCI vs. AD the accuracy was 64.5% (AUC=0.71) in sub-features and 80.8% (AUC=0.88) in all-features. Diagnostic accuracy using Freeusurfer for HC vs. MCI diagnoses was 72.0% (AUC=0.81) in sub-features and 80.3% (AUC=0.89) in all-features, for HC vs. AD it was 82.4% (AUC=0.911) in sub features, and for MCI vs. AD it was 63.9% (AUC=0.70) in sub-features and 79.1% (AUC=0.87) in all-features. The classifier identified HC vs. AD diagnoses accurately, but relatively higher inaccuracy was noted for MCI versus AD diagnoses. Thus, the AUC value is calculated to micro-average for 3 cross validation set. Figure 4 shows the confusion matrix of third fold validation set (maximum accuracy).

Feature importance

In the feature importance, the feature values of the cognitive test were ranked high for all-features. In terms of the features of cortical and subcortical volume information, inferior lateral ventricle, lateral ventricle, hippocampus, amygdala, lingual, inferior parietal, fusiform were ranked high. The performance of feature importance is shown in Figure 5.

DISCUSSION

In the contemporary model of AD, it is regarded as a disease continuum involving preclinical stage to severe dementia rather than as disease with three or four distinct entities.[7,30,31] Thus, quantification index which could reflect structural change of brain in a continuous measurement is needed to correctly assess a patient's severity of neurodegeneration. Cortical volume or thickness measurement of structural MRI has been commonly used as an objective indicator of disease progression in AD research.[32-34] In terms of technological evolution, automatic segmentation methods such as Freesurfer and Statistical Parametric Mapping (SPM) became widely available. Howev-

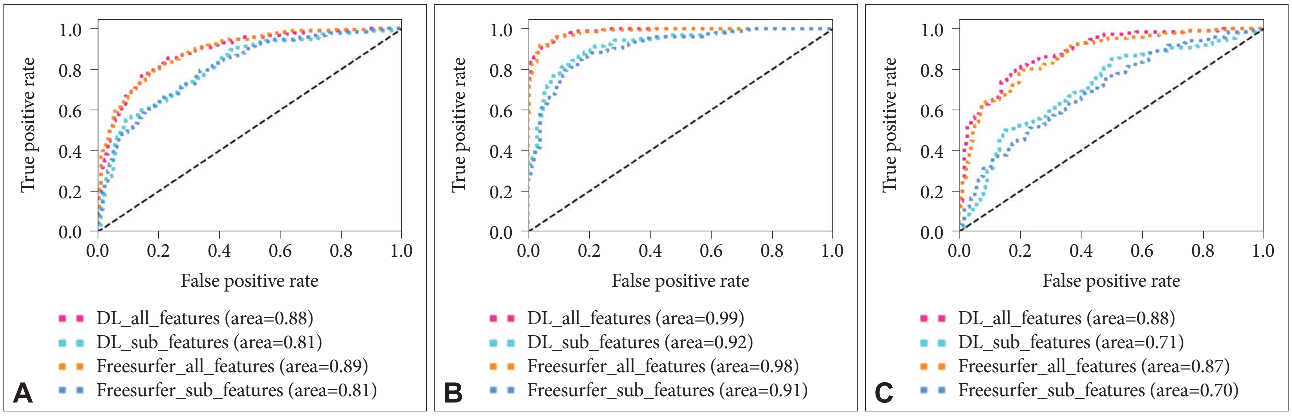

Figure 3. A: HC vs MCI. B: HC vs AD. C: MCI vs AD. HC: healthy controls, MCI: mild cognitive impairment, AD: Alzheimer's disease.

Table 3. The performance of binary classification

SW	Feature	Fold 1	Fold 2	Fold 3	Average
HC vs. AD					
Neuro-phet	Sub features				
	Acc'	0.788	0.833	0.912	0.844
	Prec'	0.792	0.849	0.913	0.851
	Sens'	0.849	0.939	0.941	0.910
	Spec'	0.727	0.727	0.882	0.779
	All features				
	Acc'	0.909	0.924	0.971	0.935
	Prec'	0.911	0.928	0.972	0.937
	Sens'	0.939	0.970	1.000	0.970
	Spec'	0.879	0.879	0.941	0.900
Free-surfer	Sub features				
	Acc'	0.758	0.833	0.882	0.824
	Prec'	0.766	0.849	0.895	0.837
	Sens'	0.849	0.939	0.971	0.920
	Spec'	0.667	0.727	0.794	0.729
	All features				
	Acc'	0.894	0.894	0.971	0.919
	Prec'	0.897	0903	0972	0.924
	Sens'	0.939	0.970	1.000	0.941
	Spec'	0.849	0.818	0.941	0.869
HC vs. MCI					
Neuro-phet	Sub features				
	Acc'	0.759	0.696	0.690	0.715
	Prec'	0.764	0.703	0.698	0.722
	Sens'	0.828	0.786	0.793	0.802
	Spec'	0.690	0.607	0.586	0.628
	All features				
	Acc'	0.879	0.839	0.707	0.808
	Prec'	0.883	0.850	0.720	0.818
	Sens'	0.931	0.929	0.828	0.896
	Spec'	0.828	0.750	0.586	0.721
Free-surfer	Sub features				
	Acc'	0.741	0.661	0.759	0.720
	Prec'	0.749	0.671	0.780	0.733
	Sens'	0.828	0.786	0.897	0.837
	Spec'	0.655	0.536	0.621	0.604
	All features				
	Acc'	0.828	0.839	0.741	0.803
	Prec'	0.834	0.850	0.767	0.817
	Sens'	0.897	0.929	0.897	0.907
	Spec'	0.759	0.750	0.586	0.698
MCI vs. AD					
Neuro-phet	Sub features				
	Acc'	0.589	0.655	0.690	0.645
	Prec'	0.589	0.655	0.693	0.646
	Sens'	0.571	0.655	0.621	0.616
	Spec'	0.607	0.655	0.759	0.674
	All features				
	Acc'	0.839	0.776	0.810	0.808
	Prec'	0.843	0.779	0.820	0.814
	Sens'	0.893	0.828	0.724	0.815
	Spec'	0.786	0.724	0.897	0.802
Free-surfer	Sub features				
	Acc'	0.589	0.672	0.655	0.639
	Prec'	0.590	0.674	0.656	0.640
	Sens'	0.536	0.724	0.621	0.627
	Spec'	0.643	0.621	0.690	0.651
	All features				
	Acc'	0.821	0.759	0.793	0.791
	Prec'	0.823	0.760	0.795	0.792
	Sens'	0.857	0.793	0.759	0.803
	Spec'	0.786	0.724	0.828	0.779

Sub-features: patient information and sub-volume feature information extracted from deep learning-based segmentation method. Inc., All features: sub-volume, patient and cognitive test feature information. Acc': accuracy, Prec': precision, Sens': sensitivity, Spec': specificity

er, the role of automatic segmentation-based diagnosis assisting algorithms in clinical practice was limited because such automatic segmentation methods required a long processing time. Low inter-method reproducibility was another important limitation hindering previous methods to be widely applied in clinical practice.[35]

In this study, we developed a quick cerebral cortical and subcortical volume analysis method using deep learning-based algorithm. We also evaluated newly developed algorithm's segmentation performance by comparing it with segmentation conducted with Freesurfer. Compared with Freesurfer, segmentation processing time of our algorithm was much shorter (less than 5 minutes for our method and around 6−8 hours for Freesurfer). Despite much faster time for segmentation, there was a high correlation between our methods and Freesurfer's with the average Dice coefficients of 106 labelling area of 0.840±0.083 (Supplementary Materials in the online-only Data Supplement). Thus, our method might be an effective automatic brain segmentation tool which can be easily applied in real

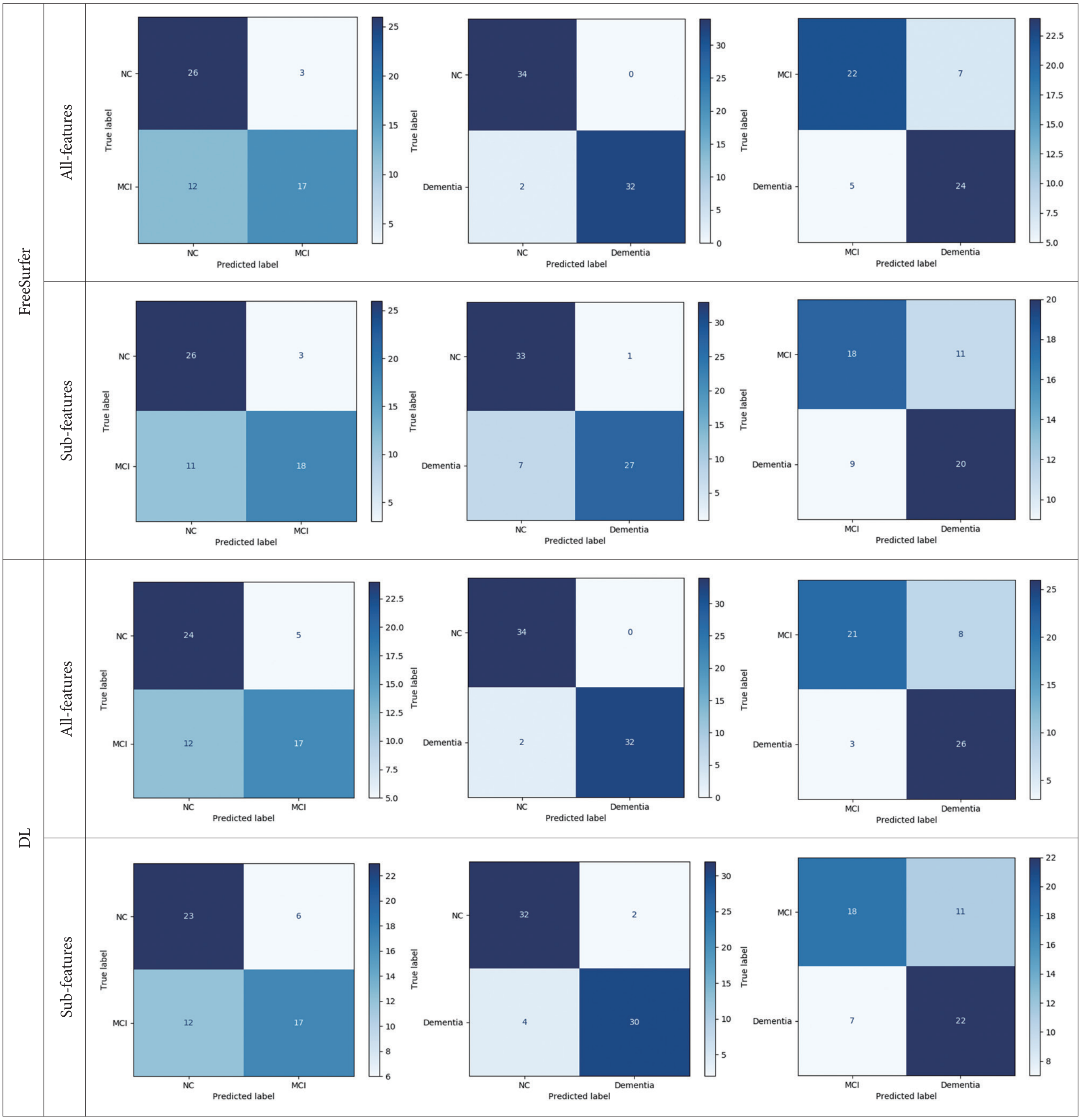

Figure 4. The confusion matrix of third fold validation set. DL: deep learning-based segmentation, FreeSurfer: FreeSurfer-based segmentation, All features: sub-volume, patient and cognitive test feature information, Sub features: patient information and sub-volume feature information extracted from deep learning-based segmentation method.

clinical practice.

ML can identify distinctive images and clinical features, which is considered as a promising technique for differential diagnosis of AD.[15] Thus, increasing studies in the field of neuroimaging have focused on the use of advanced ML algorithm in assisting differential diagnosis among patients with AD, MCI, and normal cognition.[16,18,23,36-38] However, besides longer process time and low inter-method reproducibility, low reliability of previous ML models was another important limitation. To fill in this important gap, we developed a more advanced RF model for prediction of AD which primarily used cortical and subcortical volumes as important predictive factors and also used additional information by combining cognitive function and patient characteristics. This model showed stable performance in testing process and acceptable accuracy for prediction of AD. Besides being faster than Freesurfer, our model

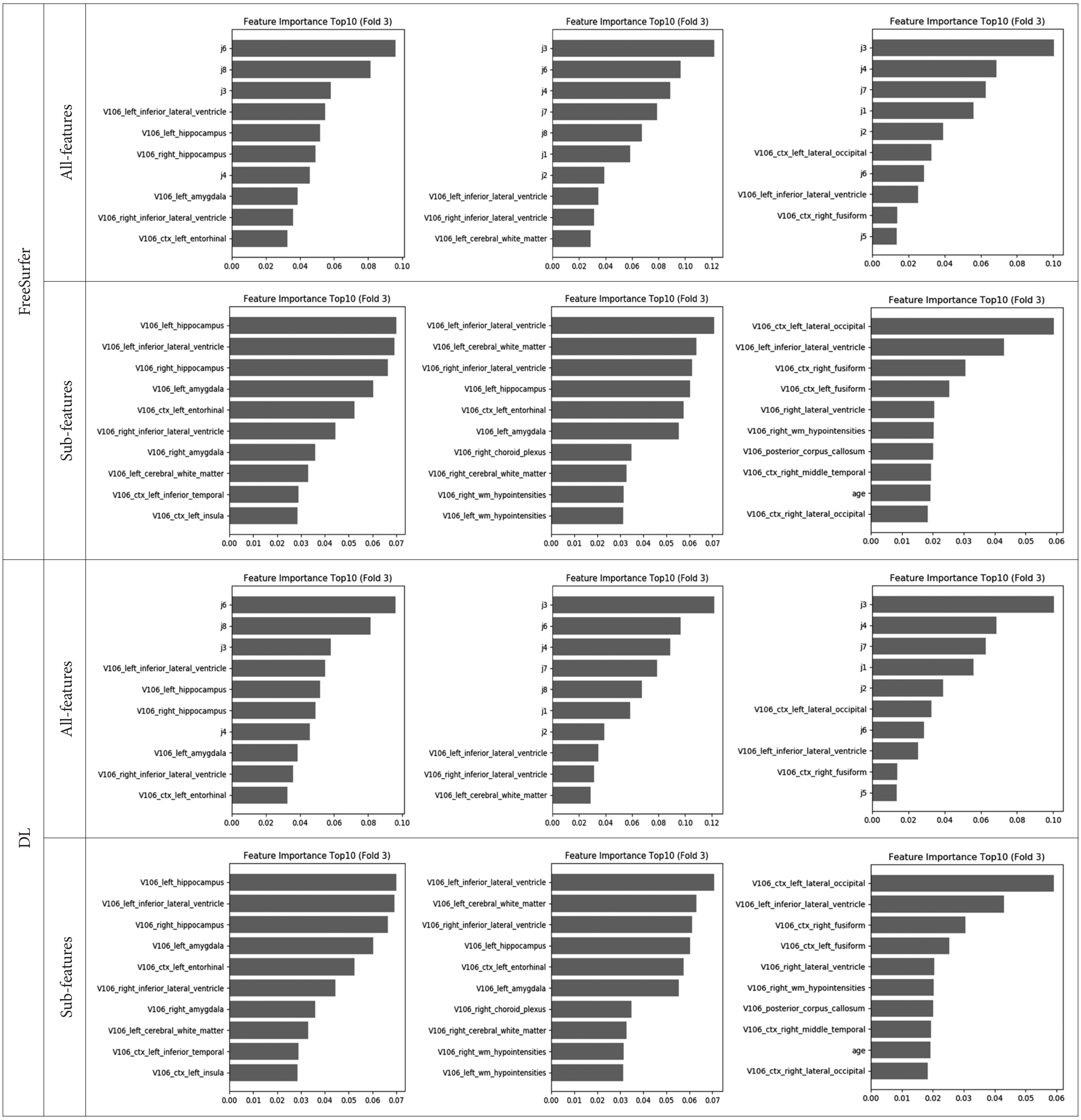

Figure 5. Feature importance of third fold validation set. Left side is HC vs. MCI, Center is HC vs. AD, and right side is MCI vs. AD. All features: sub-volume, patient and cognitive test feature information, Sub features: patient information and sub-volume feature information extracted from deep learning-based segmentation method, HC: healthy controls, MCI: mild cognitive impairment, AD: Alzheimer's disease.

also showed high accuracy in differentiating HC from AD with 93.5% (AUC 0.99), which was higher than that of the previous model which used Freesurfer 91.9% (AUC 0.98). In addition, the accuracy of our model for differentiating MCI from HC and MCI from AD were 80.8% and 80.8%, respectively, which was similar to a previous RF model showing accuracy of 81.3% (AUC) in differentiating MCI from HC using the cortical thickness and subcortical volumes and MMSE total scores.[24] In another study, accuracy of differentiating MCI from AD was in ML mode was 60−90%.[17] In line with previous researches, the accuracy was higher in differentiating AD from HC than in differentiating MCI from HC or MCI from AD.[17,24] MCI is an intermediate stage between the expected cognitive decline of normal aging or HC and the more serious decline of dementia, so patients with MCI inevitably contains overlapping clinical and neurodegenerative features with both AD and

HC.[39] Thus, both our and previous model had a more difficulty in classifying this grey area, MCI, than more distinctive stage including AD and HC. Nevertheless, the accuracy of our RF model for differentiating MCI from HC or MCI from AD was a relatively higher than that of previous literature reporting 60−90%.[18,38,40,41] Moreover, the diagnostic accuracy of our RF model was higher when the model used all-features including cortical and subcortical cortical volume, patient characteristics, and cognitive information rather than using sub-features which did not include cognitive function. Thus, both multi parameters and our RF model have contributed to the improvement in accuracy.[34]

The feature importance of our RF model for differentiating HC from AD was similar with previously known cortical atrophy patterns of AD, such as inferior lateral ventricle (temporal horn of lateral ventricle), hippocampus, and amygdala, but some regions were such as cerebral white matter, fusiform gyrus, insular cortex were considered important only in our model.[42] The difference could be due to the RF model itself or different patient characteristics, and further studies containing larger sample size are needed to clarify this issue. Additional issues should also be resolved before a computer-based brain imaging diagnostics or ML-based diagnostics can be readily applied in clinical practice. First, the stable diagnosis results should be produced despite variety of the images were collected from different MRI scanner, magnetic field strength and image resolution, and pulse sequence. Second, reliable results should be provided under various ages, gender, and race. Therefore, more studies containing diverse age, sex, and race collected from diverse MRI scanners including classical 1.5T MRI and more recent 3T MRIs are needed to enhance clinically utility of automatic segmentation-based diagnosis assisting algorithms.

There are several limitations in this study. First, our study only involved single center data with small sample size. Second, we did not have pathological diagnosis and amyloid PET for inclusion of AD subjects, although we carefully included patients with probable AD and possible AD based on the NINCDS-ADRDA Alzheimer's Criteria.[2,42] Third, we used only cortical and subcortical volume features from structural MRI and did not include other important MRI modalities reflecting white matter hyperintensity or cerebral bleeding.

Our results had relatively lower accuracy for MCI vs. AD. A larger dataset is required for the study to improve the corresponding measures' accuracy on those specific populations of interests; the inclusion of larger pool into the segmentation development cycle may help to improve segmentation performance. We attempted to address the above-mentioned lower accuracy issue, we incorporated demographic information, cognitive test results, and specific ROI volumes into our model. We expect a better performance of MCI vs. AD diagnosis with advanced imaging techniques, such as deep learning-based methods utilizing multi-modal imaging features and feature fusion in the future.

In conclusion, we showed that our RF model showed acceptable clinical feasibility and accuracy for differentiating HC from MCI and AD using structural MRI, patient information and cognitive function. This RF model not only may help clinicians to predict patients with AD continuum but may also aid to recognize patient having higher risk of AD in routine clinical practices.

Supplementary Materials

The online-only Data Supplement is available with this article at https://doi.org/10.30773/pi.2020.0304.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea goverment (MSIT) (No.2019R1A2C2009100), Research and Business Development Program through the Korea Institute for Advancement of Technology (KIAT) funded by the Ministry of Trade, Industry and Energy (MOTIE) (grant number: P0009507) and the "HPC Support" Project, supported by the Ministry of Science and ICT and National IT Industry Promotion Agency (NIPA).

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Hyun Kook Lim, JeeYoung Kim, Donghyeon Kim. Data curation: Min Kyoung Lee, Sheng-Min Wang, Nak-Young Kim, Dong Woo Kang, Yoo Hyun Um, Hae-Ran Na, Young Sup Woo, Chang Uk Lee, Won-Myong Bahk. Formal analysis: JeeYoung Kim, Minho Lee. Funding acquisition: Donghyeon Kim, Hyun Kook Lim. Investigation: JeeYonng Kim, Minho Lee, Min Kyoung Lee. Methodology: Minho Lee, Donghyeon Kim. Project administration: Donghyeon Kim, Hyun Kook Lim. Resources: Donghyeon Kim, Hyun Kook Lim. Software: Minho Lee. Supervision: Donghyeon Kim, Hyun Kook Lim, Sheng-Min Wang. Validation: Minho Lee. Visualization: Minho Lee. Writing—original draft: JeeYoung Kim, Minho Lee, Donghyeon Kim, Hyun Kook Lim. Writing—review & editing: JeeYoung Kim, Minho Lee, Sheng-Min Wang, Donghyeon Kim, Hyun Kook Lim.

ORCID iDs

JeeYoung Kim	https://orcid.org/0000-0002-2812-8159
Minho Lee	https://orcid.org/0000-0002-4821-0221
Min Kyoung Lee	https://orcid.org/0000-0003-3172-3159
Sheng-Min Wang	https://orcid.org/0000-0003-2521-1413
Nak-Young Kim	https://orcid.org/0000-0003-0116-6283
Dong Woo Kang	https://orcid.org/0000-0003-3289-075X
Yoo Hyun Um	https://orcid.org/0000-0002-3403-4140
Hae-Ran Na	https://orcid.org/0000-0002-7960-8603
Young Sup Woo	https://orcid.org/0000-0002-0961-838X
Chang Uk Lee	https://orcid.org/0000-0001-6398-7330
Won-Myong Bahk	https://orcid.org/0000-0002-0156-2510
Donghyeon Kim	https://orcid.org/0000-0003-0047-0259
Hyun Kook Lim	https://orcid.org/0000-0001-8742-3409

REFERENCES

1. Oboudiyat C, Glazer H, Seifan A, Greer C, Isaacson RS. Alzheimer's

disease. Semin Neurol 2013;33:313-329.

2. Lane CA, Hardy J, Schott JM. Alzheimer's disease. Eur J Neurol 2018; 25:59-70.
3. DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. Mol Neurodegener 2019;14:32.
4. Robillard A. Clinical diagnosis of dementia. Alzheimers Dement 2007; 3:292-298.
5. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280-292.
6. Marquez F, Yassa MA. Neuroimaging biomarkers for Alzheimer's disease. Mol Neurodegener 2019;14:21.
7. Jack Jr CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dementia 2018;14:535-562.
8. de Flores R, La Joie R, Chételat G. Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. Neuroscience 2015; 309:29-50.
9. Eskildsen SF, Coupé P, Fonov VS, Pruessner JC, Collins DL; Alzheimer's Disease Neuroimaging Initiative. Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. Neurobiol Aging 2015;36(Suppl 1):S23-S31.
10. Park M, Moon WJ. Structural MR imaging in the diagnosis of Alzheimer's disease and other neurodegenerative dementia: current imaging approach and future perspectives. Korean J Radiol 2016;17:827-845.
11. Moon SW, Lee B, Choi YC. Changes in the hippocampal volume and shape in early-onset mild cognitive impairment. Psychiatry Investig 2018;15:531-537.
12. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimers Dement 2012;8:S1-S68.
13. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics 2017;37:505-515.
14. Bratić B, Kurbalija V, Ivanović M, Oder I, Bosnić Z. Machine learning for predicting cognitive diseases: methods, data sources and risk factors. J Med Syst 2018;42:243.
15. Bryan RN. Machine Learning Applied to Alzheimer Disease. Chicago: Radiological Society of North America; 2016.
16. Kim J, Lee B. Automated discrimination of dementia spectrum disorders using extreme learning machine and structural t1 mri features. Annu Int Conf IEEE Eng Med Biol Soc 2017;2017:1990-1993.
17. Mirzaei G, Adeli A, Adeli H. Imaging and machine learning techniques for diagnosis of Alzheimer's disease. Rev Neurosci 2016;27:857-870.
18. Pellegrini E, Ballerini L, Hernandez MDCV, Chappell FM, González-Castro V, Anblagan D, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimers Dement (Amst) 2018;10:519-535.
19. Salvatore C, Battista P, Castiglioni I. Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. Curr Alzheimer Res 2016;13:509-533.
20. Sorensen L, Nielsen M; Alzheimer's Disease Neuroimaging Initiative. Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination. J Neurosci Methods 2018;302:66-74.
21. Zhao M, Chan RH, Tang P, Chow TW, Wong SW. Trace ratio linear discriminant analysis for medical diagnosis: a case study of dementia. IEEE Signal Process Lett 2013;20:431-434.
22. Dimitriadis SI, Liparas D, Tsolaki MN; Alzheimer's Disease Neuroimaging Initiative. Random forest feature selection, fusion and ensemble strategy: combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and alzheimer's disease patients: from the alzheimer's disease neuroimaging initiative (ADNI) database. J Neurosci Methods 2018;302:14-23.
23. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci 2017;9:329.
24. Lebedeva AK, Westman E, Borza T, Beyer MK, Engedal K, Aarsland D, et al. MRI-based classification models in prediction of mild cognitive impairment and dementia in late-life depression. Front Aging Neurosci 2017;9:13.
25. Kang DW, Choi WH, Jung WS, Um YH, Lee CU, Lim HK. Impact of amyloid burden on regional functional synchronization in the cognitively normal older adults. Sci Rep 2017;7:1-9.
26. Hahn C, Lee CU, Won WY, Joo SH, Lim HK. Thalamic shape and cognitive performance in amnestic mild cognitive impairment. Psychiatry Investig 2016;13:504-510.
27. Fischl B. FreeSurfer. NeuroImage 2012;62:774-781.
28. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18-22.
29. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238.
30. Choi JB, Cho KJ, Kim JC, Kim CH, Chung YA, Jeong HS, et al. The effect of daily low dose tadalafil on cerebral perfusion and cognition in patients with erectile dysfunction and mild cognitive impairment. Clin Psychopharmacol Neurosci 2019;17:432-437.
31. Han JW, Kim TH, Kwak KP, Kim K, Kim BJ, Kim SG, et al. Overview of the Korean longitudinal study on cognitive aging and dementia. Psychiatry Investig 2018;15:767-774.
32. Mak E, Su L, Williams GB, Watson R, Firbank MJ, Blamire AM, et al. Progressive cortical thinning and subcortical atrophy in dementia with Lewy bodies and Alzheimer's disease. Neurobiol Aging 2015;36:1743-1750.
33. Fan LY, Tzen KY, Chen YF, Chen TF, Lai YM, Yen RF, et al. The relation between brain amyloid deposition, cortical atrophy, and plasma biomarkers in amnesic mild cognitive impairment and Alzheimer's disease. Front Aging Neurosci 2018;10:175.
34. Jung WS, Um YH, Kang DW, Lee CU, Woo YS, Bahk WM, et al. Diagnostic validity of an automated probabilistic tractography in amnestic mild cognitive impairment. Clin Psychopharmacol Neurosci 2018;16: 144-152.
35. Palumbo L, Bosco P, Fantacci M, Ferrari E, Oliva P, Spera G, et al. Evaluation of the intra-and inter-method agreement of brain MRI segmentation software packages: a comparison between SPM12 and Free-Surfer v6. 0. Physica Medica 2019;64:261-272.
36. Ahmed MR, Zhang Y, Feng Z, Lo B, Inan OT, Liao H. Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. IEEE Rev Biomed Eng 2018;12:19-33.
37. Álvarez JD, Matias-Guiu JA, Cabrera-Martín MN, Risco-Martín JL, Ayala JL. An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders. BMC Bioinformatics 2019;20:491.
38. Lee JS, Kim C, Shin JH, Cho H, Shin DS, Kim N, et al. Machine learning-based individual assessment of cortical atrophy pattern in Alzheimer's disease spectrum: development of the classifier and longitudinal evaluation. Sci Rep 2018;8:4161.
39. Portet F, Ousset PJ, Visser PJ, Frisoni GB, Nobili F, Scheltens P, et al. Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. J Neurol Neurosurg Psychiatry 2006;77:714-718.
40. Cure S, Abrams K, Belger M, Happich M. Systematic literature review and meta-analysis of diagnostic test accuracy in Alzheimer's disease and other dementia using autopsy as standard of truth. J Alzheimers Dis 2014;42:169-182.
41. Moscoso A, Silva-Rodríguez J, Aldrey JM, Cortés J, Fernández-Ferreiro A, Gómez-Lado N, et al. Prediction of Alzheimer's disease dementia with MRI beyond the short-term: Implications for the design of predictive models. NeuroImage Clin 2019;23:101837.
42. 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.

SUPPLEMENTARY MATERIALS

Deep learning-based segmentation method

We performed the Desikan-Killiany atlas-based freesurfer segmentation on 388 patients of Yeouido St. Mary's Hospital dataset as well as public datasets such as HCP, ADNI, PPMI, AIBL, and IXI, and two expert performed manual correction to produced fine-turned ground truth dataset. In addition, hypo-intensity region was added. The dataset was separated into three sets: training, validation, and testing. We first randomly shuffled the dataset and separated 49 patients for testing. The remaining patient's data were used for training and validation (9.5:0.5). The training data was constructed by extracting the three-dimension patch image using uniform sampling (96×96×96) for the individual ground truth data (Figure 1).

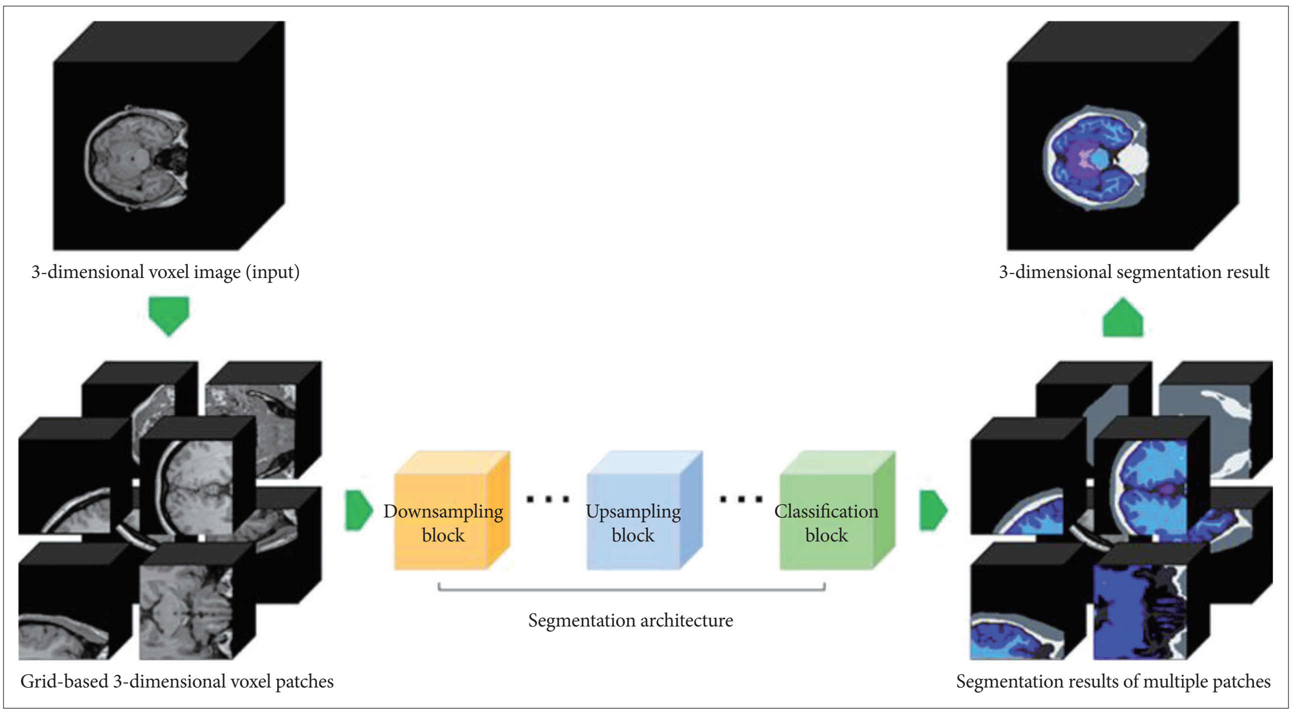
3-dimensional voxel image (input)
3-dimensional segmentation result
Downsampling block
Upsampling block
Classification block
Segmentation architecture
Grid-based 3-dimensional voxel patches
Segmentation results of multiple patches

Figure 1. Three-dimension patch-based training.

We improved the UNet++ deep learning architecture with a three-dimension methodology to train about 104 labels. This algorithm has a convolutional layer in the skip path, which bridges the semantic gap between the encoder and decoder characteristic maps. There is a dense skip-connection in the skip path, which improves the gradient flow, has a deep supervision, which enables model pruning, improves performance, or at worst compares to using only one lossy layer. Performance can be achieved (Figure 2).

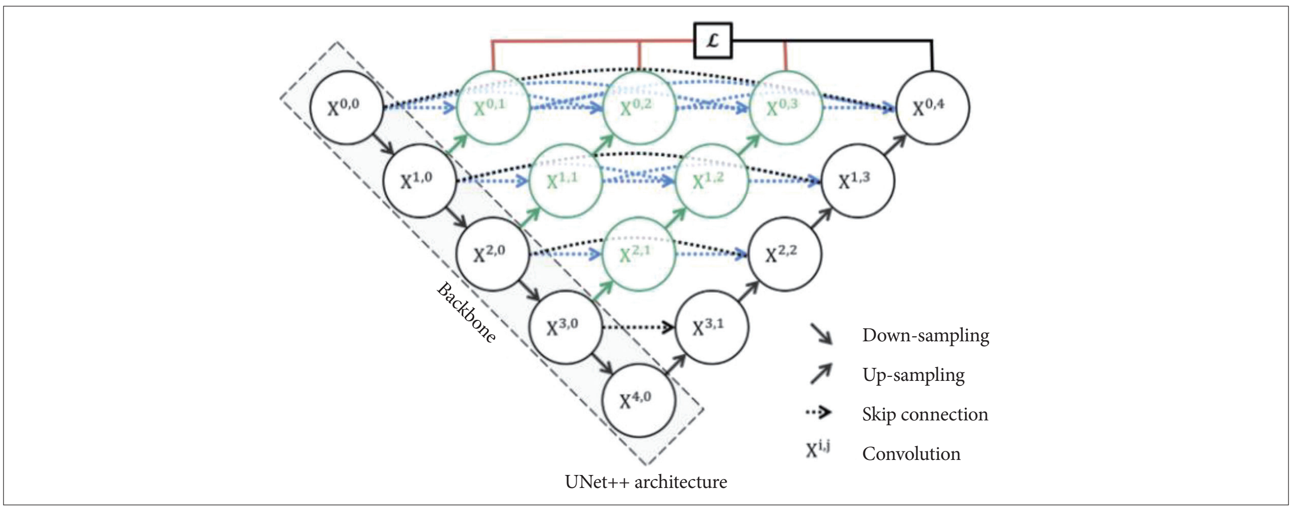

Figure 2. U-Net++ deep learning architecture.

Because the voxel by voxel segmentation learning method is used, the CrossEntropy loss function is used, and the learning rate for Adam optimizer is 0.0001. The total number of iterations is 300,000. Segmentation results are obtained by merging inference data using a three-dimension patch sliding aggregator. Figure 3 shows the segmentation result of brain sub-volumes.

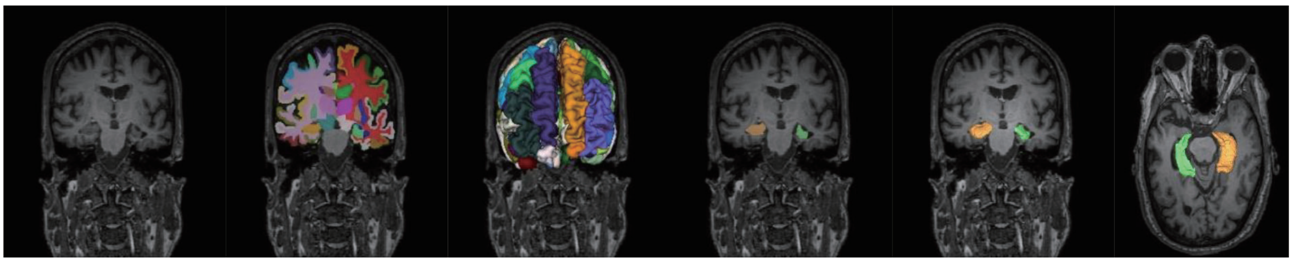

Figure 3. Deep learning-based segmentation result.

We perform the Dice overlap with the extra-validation set and, the average Dice coefficient is 0.840±0.083. Table 1 shows the whole Dice coefficients with the test set.

Table 1. Dice overlap result with the test set (49 case)

Case	Average dice 106 labels
0	0.706±0.262
1	0.791±0.186
2	0.866±0.085
3	0.873±0.067
4	0.857±0.068
5	0.877±0.066
6	0.822±0.100
7	0.857±0.070
8	0.856±0.072
9	0.765±0.113
10	0.768±0.099
11	0.766±0.108
12	0.779±0.088
13	0.758±0.106
14	0.829±0.098
15	0.886±0.052
16	0.874±0.053
17	0.886±0.057
18	0.853±0.084
19	0.873±0.065
20	0.876±0.063
21	0.875±0.068
22	0.876±0.061
23	0.877±0.060
24	0.872±0.061
25	0.881±0.058
26	0.863±0.102
27	0.871±0.071
28	0.883±0.058
29	0.84±0.075
30	0.861±0.064
31	0.825±0.067
32	0.837±0.097
33	0.851±0.080
34	0.839±0.097
35	0.822±0.103
36	0.867±0.065
37	0.865±0.060
38	0.794±0.101
39	0.758±0.109
40	0.864±0.060
41	0.789±0.089
42	0.878±0.059
43	0.883±0.060
44	0.752±0.102
45	0.855±0.092
46	0.859±0.069
47	0.863±0.064
48	0.854±0.059
Mean±std	0.840±0.083