NeuroImage: Clinical 23 (2019) 101811

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease

Jun Pyo Kim[a,1], Jeonghun Kim[b,1], Yu Hyun Park[a], Seong Beom Park[a], Jin San Lee[d], Sole Yoo[e], Eun-Joo Kim[f], Hee Jin Kim[a], Duk L. Na[a], Jesse A. Brown[g], Samuel N. Lockhart[h], Sang Won Seo[a,i,j,k,l,**,2], Joon-Kyung Seong[b,c,*,2]

[a] Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[b] Department of Bio-convergence Engineering, Korea University, Seoul, Republic of Korea
[c] School of Biomedical Engineering, Korea University, Seoul, Republic of Korea
[d] Department of Neurology, Kyunghee University Medical Center, Seoul, Republic of Korea
[e] Department of Cognitive Science, Yonsei University, Seoul, Republic of Korea
[f] Department of Neurology, Busan National University Hospital, Busan, Republic of Korea
[g] Department of Neurology, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA
[h] Department of Internal Medicine, Section of Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA
[i] Neuroscience Center, Samsung Medical Center, Seoul, Republic of Korea
[j] Samsung Alzheimer Research Center, Samsung Medical Center, Seoul, Republic of Korea
[k] Center for Clinical Epidemiology, Samsung Medical Center, Seoul, Republic of Korea
[l] Department of Clinical Research Design and Evaluation, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Frontotemporal dementia
Classification model
Machine learning

ABSTRACT

Background: In a clinical setting, an individual subject classification model rather than a group analysis would be more informative. Specifically, the subtlety of cortical atrophy in some frontotemporal dementia (FTD) patients and overlapping patterns of atrophy among three FTD clinical syndromes including behavioral variant FTD (bvFTD), non-fluent/agrammatic variant primary progressive aphasia (nfvPPA), and semantic variant PPA (svPPA) give rise to the need for classification models at the individual level. In this study, we aimed to classify each individual subject into one of the diagnostic categories in a hierarchical manner by employing a machine learning-based classification method.

Methods: We recruited 143 patients with FTD, 50 patients with Alzheimer's disease (AD) dementia, and 146 cognitively normal subjects. All subjects underwent a three-dimensional volumetric brain magnetic resonance imaging (MRI) scan, and cortical thickness was measured using FreeSurfer. We applied the Laplace Beltrami operator to reduce noise in the cortical thickness data and to reduce the dimension of the feature vector. Classifiers were constructed by applying both principal component analysis and linear discriminant analysis to the cortical thickness data. For the hierarchical classification, we trained four classifiers using different pairs of groups: Step 1 - CN vs. FTD + AD, Step 2 - FTD vs. AD, Step 3 - bvFTD vs. PPA, Step 4 - svPPA vs. nfvPPA. To evaluate the classification performance for each step, we used a10-fold cross-validation approach, performed 1000 times for reliability.

Results: The classification accuracy of the entire hierarchical classification tree was 75.8%, which was higher than that of the non-hierarchical classifier (73.0%). The classification accuracies of steps 1–4 were 86.1%, 90.8%, 86.9%, and 92.1%, respectively. Changes in the right frontotemporal area were critical for discriminating behavioral variant FTD from PPA. The left frontal lobe discriminated nfvPPA from svPPA, while the bilateral anterior temporal regions were critical for identifying svPPA.

Conclusions: In the present study, our automated classifier successfully classified FTD clinical subtypes with good to excellent accuracy. Our classifier may help clinicians diagnose FTD subtypes with subtle cortical atrophy and facilitate appropriate specific interventions.

* Correspondence to: J. K. Seong, School of Biomedical Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, Republic of Korea.
** Correspondence to: S. W. Seo, Department of Neurology, Samsung Medical Center, Sungkyunkwan University, School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Republic of Korea.
E-mail addresses: sw72.seo@samsung.com (S.W. Seo), jkseong@korea.ac.kr (J.-K. Seong).
[1] The authors contributed equally to this work.
[2] The authors contributed equally to this work.

https://doi.org/10.1016/j.nicl.2019.101811
Received 9 December 2018; Received in revised form 30 March 2019; Accepted 1 April 2019
Available online 03 April 2019

1. Background

Frontotemporal dementia (FTD) is one of the leading causes of early-onset degenerative dementia (Vieira et al., 2013). The clinically defined syndromes within the FTD spectrum include three variants: the behavioral variant FTD (bvFTD), which is associated with early behavioral and executive deficits; semantic variant primary progressive aphasia (svPPA), which is associated with semantic anomia and impaired comprehension; and non-fluent/agrammatic variant primary progressive aphasia (nfvPPA), which is a progressive disorder of speech, grammar and word output.(Bang et al., 2015)

Interpretation of magnetic resonance imaging (MRI) scans largely relies on the intuition and experience of clinicians, though MRI scans help clinicians diagnose FTD as an auxiliary tool. With the rapid development of neuroimaging analysis, we can automatically analyze cortical atrophy. In this regard, a previous study suggested that FTD syndromes are characterized by cortical atrophy in the frontal, anterior temporal and frontoinsular regions (Rosen et al., 2002). The relative involvement patterns of frontotemporal structures in FTD also vary among clinical syndromes. That is, patients with bvFTD had cortical atrophy in the anterior cingulate and frontal insular cortices, most prominently early in the course of the disease (Rosen et al., 2002; Davies et al., 2009). Conversely, studies in patients with nfvPPA indicate cortical atrophy in the left frontal area, especially the inferior frontal gyrus, pars triangularis, Rolandic operculum and precentral gyrus with left predominance. (Pereira et al., 2009; Gorno-Tempini et al., 2006) Semantic variant PPA is known to have the most distinct atrophic pattern among FTD clinical syndromes, which is most prominent in the left anterior temporal lobe.(Davies et al., 2009; Hodges and Patterson, 2007; Brambati et al., 2015)

In clinical settings, an individual subject classification model rather than a group analysis would be more informative. The subtlety of cortical atrophy in early-stage FTD patients and overlapping characteristics of atrophy patterns among the FTD clinical syndromes, Alzheimer's disease (AD) and normal aging give rise to the need for an automated image analysis procedure which can be used at the individual level. Specifically, considering that different forms of dementia correlate with different underlying neuropathologies,(Seelaar et al., 2011) distinguishing between different causes of dementia will become more important with the emergence of targeted therapies.

In this study, we therefore aimed to classify each individual subject into one of the diagnostic categories in a hierarchical manner by employing a machine learning-based classification method using surface-based cortical thickness data. The hierarchical scheme of our classification algorithm was designed based on the clinical decision process. In clinical practice, after noticing that a patient has abnormal findings that cannot be explained by normal aging, a clinician has to rule out AD first since it is the most common cause of degenerative dementia. If the patient has behavioral or language problems that suggest FTD, the clinician will determine which clinical syndrome it is. To emulate this process, first, we discriminated the dementia group (FTD and AD groups combined) from the cognitively normal (CN) group. Subsequently, the dementia group was classified into FTD and AD groups. Afterwards, subjects from the FTD group were classified into bvFTD and PPA groups. Finally, the PPA group was further classified into nfvPPA and svPPA groups. Our algorithm used a Laplace Beltrami operator to reduce noise, followed by linear discriminant analysis (LDA) in combination with principal component analysis (PCA).

2. Methods

2.1. Participants

We consecutively recruited 143 patients with FTD who visited the dementia clinic of Samsung Medical Center (Seoul, Korea) from September 2007 to March 2017. All FTD patients who were enrolled in this study met the diagnostic criteria for FTD clinical subtypes proposed by Rascovsky et al. (Rascovsky et al., 2011) (for bvFTD) and Gorno-Tempini et al.(Gorno-Tempini et al., 2011) (for nfvPPA and svPPA).

All patients were evaluated by comprehensive interviews, neurological examinations, and neuropsychological assessment. In brief, caregivers were interviewed in depth by neurologists and neuropsychologists. Blood tests to exclude secondary causes of dementia included a complete blood count, blood chemistry tests, vitamin B12/folate, syphilis serology, and thyroid function tests. Conventional brain MRI scans confirmed the absence of structural lesions such as tumors, traumatic brain injuries, hydrocephalus, and severe white matter hyperintensities. Thirty-four out of 143 FTD patients underwent ^{18}F-florbetaben or ^{18}F-flutemetamol amyloid positron emission tomography (PET) scanning and four of them had significant amyloid deposition. A committee that included 5–10 dementia specialists held a quarterly meeting to review the clinical histories and brain imaging results of all cases enrolled in this study, and to reach a consensus regarding clinical diagnosis.

We also recruited 50 age-matched AD dementia patients and 146 CN subjects from an in-house registry of individuals who underwent amyloid PET scanning (^{18}F-florbetaben or ^{18}F-flutemetamol) from August 2015 to July 2017 and performed the same clinical assessments and imaging studies. All AD dementia patients met the criteria for probable AD dementia with evidence of the AD pathophysiological process proposed by the National Institute on Aging-Alzheimer's Association (McKhann et al., 2011) based on clinical assessments and Aβ positivity shown in the amyloid PET. The CN group consisted of cognitively normal subjects without amyloid deposition on ^{18}F-florbetaben PET.

2.2. Ethics statement

The institutional review boards at all participating centers approved this study, and informed consent was obtained from the patients and caregivers.

2.3. PET image acquisition and analysis

We used ^{18}F-florbetaben PET or ^{18}F-flutemetamol PET to detect amyloid in the brain. PET images were dichotomized as either amyloid positive or negative using visual reads. We defined ^{18}F-florbetaben PET as positive when a score of 2 or 3 was assigned during visual assessment on the brain Aß plaque load (BAPL) scoring system.(Barthel et al., 2011) Visual interpretation of ^{18}F-flutemetamol PET images relied upon a systematic review of five brain regions (frontal, parietal, posterior cingulate and precuneus, striatum and lateral temporal lobes). If any one of the brain regions systematically reviewed for ^{18}F-flutemetamol PET was positive in either hemisphere, the scan was considered positive.(Farrar, 2017) In this study, PET images were used to confirm the diagnostic label of subjects in AD and CN groups.

2.4. MR image acquisition

All subjects underwent a three-dimensional (3D) volumetric brain MRI scan. An Achieva 3.0-Tesla MRI scanner (Philips, Best, the Netherlands) was used to acquire 3D T1 Turbo Field Echo (TFE) MRI data using the following imaging parameters: sagittal slice thickness, 1.0 mm with 50% overlap; no gap; repetition time of 9.9 ms; echo time of 4.6 ms; flip angle of 8°; and matrix size of 240 × 240 pixels reconstructed to 480 × 480 over a field of view of 240 mm.

2.5. Image preprocessing

For automated surface modeling and measurement of each subject's cortical thickness, we applied the FreeSurfer (version 5.1) pipeline to the T1 weighted MR image (http://surfer.nmr.mgh.harvard.edu/). Fig. 1A shows an overview of the proposed method. The first step segments the T1 weighted image based on signal intensity, which

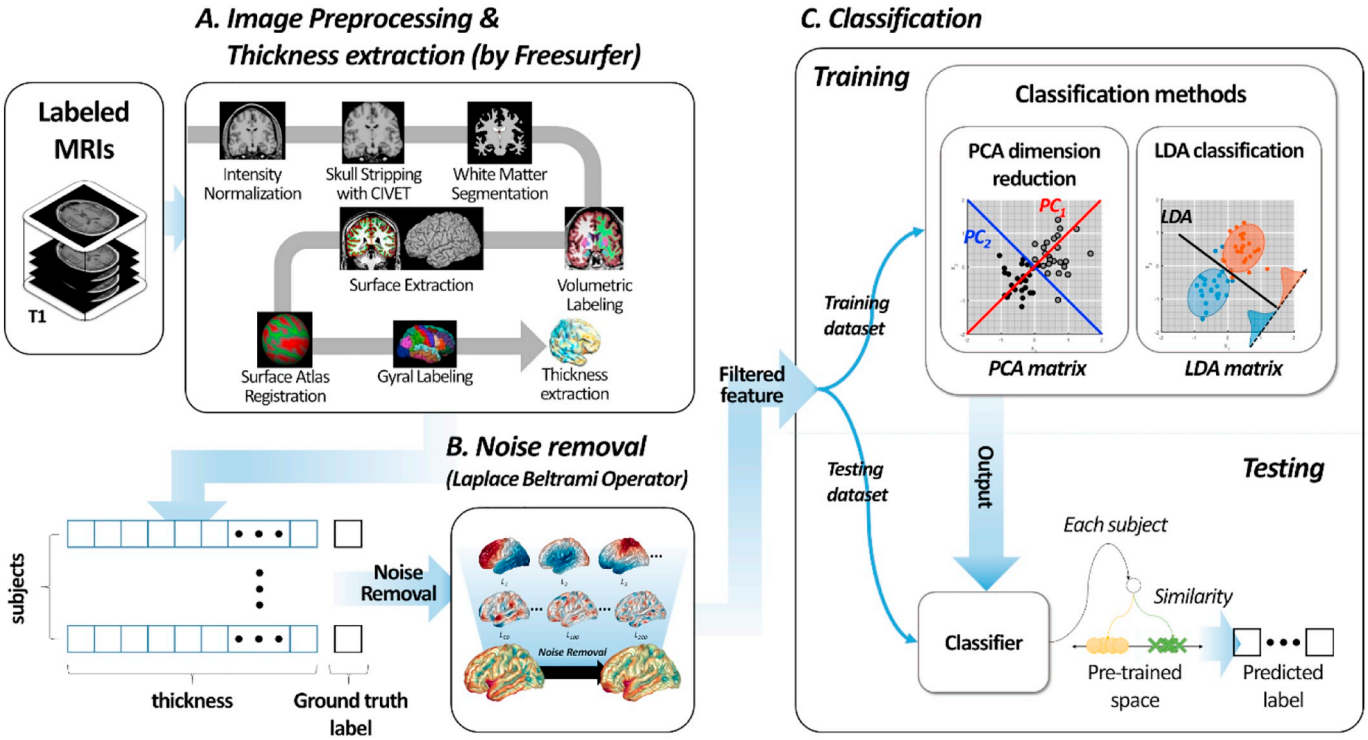

Fig. 1. Overview of the proposed cortical atrophy pattern-based classification method. (A) Image preprocessing and cortical thickness extraction. (B) Noise removal based on the Laplace Beltrami operator. (C) Cortical atrophy pattern-based classification including a training step and a testing step.

includes motion noise correction, space transformation, normalization and skull stripping. Afterwards, we employed the CIVET pipeline for additional correction of the skull-stripped image. Subsequently, the cortical surfaces were constructed for both white and gray matter boundaries. The gray and white matter surfaces were then used for calculating cortical thickness. To achieve correspondence between subjects, the mesh vertices were resampled to have the same number (40,962) of vertices for each hemisphere. Finally, the cortical thicknesses were defined at every vertex as the distance between two corresponding vertices of the gray and white matter surfaces. Throughout the whole image preprocessing pipeline, a neuroanatomist visually checked images, and corrected image processing errors manually. In particular, the image segmentation was carefully examined and corrected manually in subjects with atrophy: for example, svPPA patients often have anterior temporal lobe atrophy to such a degree that FreeSurfer processing may initially fail to detect the gray matter.

Of the initial 339 subjects, eight subjects were excluded due to errors in preprocessing: FreeSurfer failed to produce results in seven subjects, and one subject had an overestimation error which could not be corrected manually. Therefore, MRI scans of 331 subjects (48 AD, 48 bvFTD, 50 svPPA, 39 nfvPPA and 146 CN subjects) were analyzed in this study.

2.6. Hierarchical classification based on cortical atrophy

For hierarchical classification, we used four different pairs of groups for each classifier step. Fig. 2A shows a schematic view of the hierarchical classification. First, we trained a classifier using the CN and Dementia (FTD + AD) groups (Step 1). Subsequently, another classifier was trained using the FTD and AD groups (Step 2). Next, the FTD classifier was trained using the bvFTD and PPA groups (Step 3). Finally, the PPA classifier was trained to distinguish between svPPA and nfvPPA (Step 4). The hierarchical classification was performed with these four classifiers. Specifically, a single subject was classified using the Step 1 classifier. If the subject was classified as a patient, the subject was then tested using the Step 2 classifier. Again, if the subject was classified as an FTD patient, then Step 3 classifier was applied. This hierarchical process was performed consecutively through the entire tree until the subject was finally classified into one of the final clinical labels. Using the cortical thickness data of each subject, we applied the Laplace Beltrami operator for noise removal (Fig. 1B). This scheme transforms the cortical thickness data from the geometrical domain into frequency space, and represents the original data using oscillations of alternating thin and thick cortices across the cortical surface.(Qiu et al., 2006; Vallet and Lévy, 2008) In the frequency domain, high frequency components were considered as noise, and thus the lower frequency components were used as features in classification. For each subject, the original cortical thickness data was sampled at 81,924 vertices, which was then transformed to about 250-dimensional frequency domain. The detailed process of noise removal was described in our previous work. (Cho et al., 2012) The classifier was then constructed by applying both PCA and LDA to the cortical thickness data(Belhumeur et al., 1997; San Lee et al., 2018) (Fig. 1C). PCA was applied for the purpose of dimension reduction, which transforms the 250-dimensional feature data to much lower dimensional space. The detailed information on the feature transformation was described in Supplementary Table S1. Finally, the individual cortical thickness data that was not included in the training data set was tested to obtain a prediction label (Fig. 1C).

For comparison purposes, two additional experiments were performed. First, classification using a single, five-label LDA classifier employing the same learning procedure (Fig. 2B) was performed to demonstrate how the hierarchical scheme improved the classification performance. Additionally, pairwise classifications without the use of the Laplace-Beltrami operator were conducted to evaluate how much this noise-removal step contributed to our classification performances.

In order to evaluate the classification performance, we used a k-fold cross-validation (CV) approach (Fig. 1C). For each classification step in the hierarchical tree, the set was randomly split into $k = 10$ independent subsets. Nine subsets were used for training, and the remaining subset was used for testing. We performed the cross-validation 1000 times for reliability and calculated the mean accuracy, sensitivity and specificity.

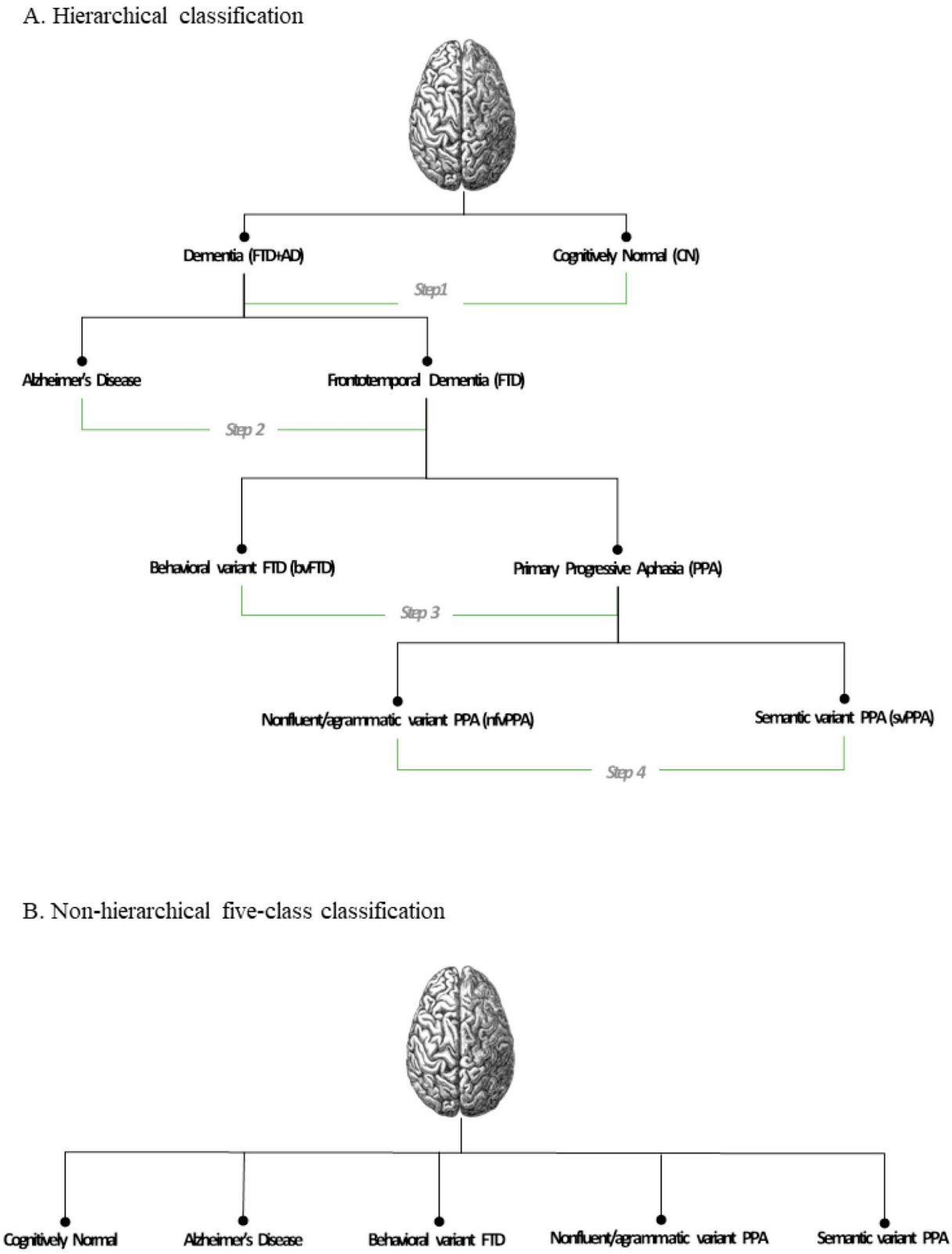

Fig. 2. Schematic view of (A)hierarchical and (B)non-hierarchical classification.

2.7. Discriminative region analysis

We extracted the discriminative regions for our classifiers, which provided topographic patterns representing the contribution of each brain region to the discriminability between the two groups being compared. We obtained the discriminative regions for each classification by visualizing the weight vector of the classifier similarly to the method of Haufe et al.(Haufe et al., 2014) Generally, the discriminative regions are the brain regions with relative importance in classification and are obtained as:

$$D = \sum_x \times w$$

(w: weight vector, $\Sigma_x = XX^T$ is the covariance of the feature vector)

Since, in our approach, we used additional steps for noise removal and PCA dimension reduction, the weight vector was defined by multiplication between the PCA matrix and the LDA matrix ($w = M_{PCA} \times M_{LDA}$). Additionally, X is the filtered feature in the frequency domain. We then obtained the discriminate pattern (D_{freq}) on frequency space as:

$$D_{freq} = \sum_{x_{filtered}} \times (M_{PCA} \times M_{LDA}).$$

This discriminative pattern on frequency domain was projected back into the surface domain by shifting back D_{freq} using the center (PCA_{mean}) and MHT (M_{MHT}). Finally, the discriminated region ($D_{surface}$) on the surface was obtained as:

$$D_{surface} = (D_{freq} + PCA_{mean}) \times (M_{MHT})^T$$

For visualization, we colored $D_{surface}$ on the template surface, and the warm/cool color represents the importance of the feature for each group, with darker colors indicating greater importance.

2.8. Statistics

We compared the demographic and clinical data among groups using one-way analysis of variance (ANOVA) tests. Continuous variables were expressed as mean ± standard deviation (SD). Statistical analyses were performed using R version 3.5.0.

3. Results

3.1. Clinical characteristics

As shown in Table 1, of the 331 subjects, 153 (46.2%) were men. The mean age of subjects in the FTD group was 65.5 ± 11.8 years (bvFTD: 62.4 ± 9.4, svPPA: 65.6 ± 7.9, and nfvPPA: 68.9 ± 8.6). The interval between the onset of symptoms and MRI acquisition was 3.4 ± 2.4 years. There were no differences in age, years of education, and time from symptom onset to MRI among groups. There was no difference in mini-mental status examination (MMSE) score among FTD subtypes and the AD group.

3.2. Classification performance

The results from classification using the entire hierarchical tree showed that each subject was classified into one of the five clinical labels with 75.8% accuracy. (Tables 2A-1) Supplementary Fig. S1A shows a confusion matrix of the hierarchical classification approach. Within the confusion matrix, the numbers in the boxes located diagonally from the top-left corner to the bottom-right corner indicate the cumulative accuracies per diagnostic subgroup (CN 89.3%, AD 73.1%, bvFTD 57.2%, nfvPPA 51.9%, and svPPA 75.6%).

Table 2
Classification performances.

	Accuracy	Sensitivity	Specificity	AUC
A-1. The entire hierarchical tree approach				
	75.8%	69.4%	93.2%	
A-2. Performances of the pairwise classifiers from four steps				
Step 1 (CN vs Dementia)	86.1% (85.9–86.3%)	87.0%	85.4%	0.917
Step 2 (AD vs FTD)	90.8% (90.5–91.1%)	87.5%	92.0%	0.955
Step 3 (bvFTD vs PPA)	86.9% (86.2–87.5%)	92.1%	77.1%	0.865
Step 4 (nfvPPA vs svPPA)	92.1% (91.6–92.7%)	97.4%	88.0%	0.955
B. A single, multi-label classification performance				
	73.0%	67.1%	92.6%	

Accuracies for pairwise classifiers were shown with 95% confidence intervals in brackets.
Abbreviations: AUC = Area under receiver operating characteristic curve, FTD = frontotemporal dementia, bvFTD = behavioral variant frontotemporal dementia, PPA = primary progressive aphasia, svPPA = semantic variant primary progressive aphasia, nfvPPA = non-fluent/agrammatic variant primary progressive aphasia, AD = Alzheimer's disease, CN = cognitively normal.

Table 2A-2 shows the performance of each classification step. In step 1, the accuracy in discriminating between CN subjects and Dementia (FTD + AD) patients was 86.1%. In step 2, AD and FTD patients were classified with 90.8% accuracy. Within the FTD group, the classifier discriminated between bvFTD and PPA patients with 86.9% accuracy (step 3). The accuracy in discriminating between PPA clinical syndromes was 92.1% in step 4. The receiver operating characteristic (ROC) curves for steps 1 to 4 are shown in Supplementary Fig. S2. The areas under the ROC curves for steps 1 to 4 were 0.917, 0.955, 0.865, and 0,955, respectively.

For comparison with the hierarchical classifier, the classification performance of a single, multi-label classifier which does not use the hierarchical scheme is shown in Table 2B. This classifier demonstrated an accuracy of 73.0%. The confusion matrix of this multi-label classifier is shown in Supplementary Fig. S1B.

Supplementary Table S2 further shows the classification performance without the application of the Laplace Beltrami operator to cortical thickness data. In the overall hierarchical steps, we obtained 3–4% improvements in accuracies by using the operator to reduce noise components in the cortical thickness data. Fig. 3 depicts the discriminative regions on the atlas surface for our classifiers.

For classification between Dementia and CN groups, the left fronto-parieto-temporal areas, right anterior temporal area, and right superior frontal gyrus distinguished demented subjects from cognitively normal subjects. In step 2, the bilateral precuneus and lateral parietal, right posterior temporal and lateral occipital, and left frontal regions

Table 1
Clinical characteristics of participants.

	Total	FTD			AD	CN
		bvFTD	svPPA	nfvPPA		
Number	331	48	50	39	48	146
Age, years	65.4 ± 11.8	62.4 ± 9.4	65.6 ± 7.9	68.9 ± 8.6	65.7 ± 7.6	65.5 ± 15.0
Gender (M/F)	153/178	26/22	29/21	17/22	25/23	56/90
Education, years	10.8 ± 5.3	12.5 ± 5.2	11.0 ± 4.8	11.5 ± 5.1	10.9 ± 2.7	10.0 ± 6.0
K-MMSE score	23.2 ± 7.4	19.6 ± 6.7	18.8 ± 8.8	19.5 ± 7.9	17.7 ± 5.8	28.6 ± 1.7[†]
Years from first symptom	3.4 ± 2.4	3.2 ± 2.4	3.4 ± 2.5	3.0 ± 2.0	4.1 ± 2.5	

Abbreviations: N = number, FTD = frontotemporal dementia, bvFTD = behavioral variant frontotemporal dementia, svPPA = semantic variant primary progressive aphasia, nfvPPA = non-fluent/agrammatic variant primary progressive aphasia, AD = Alzheimer's disease, CN = cognitively normal, K-MMSE = Korean mini-mental state examination.
[†] $p < 0.05$.

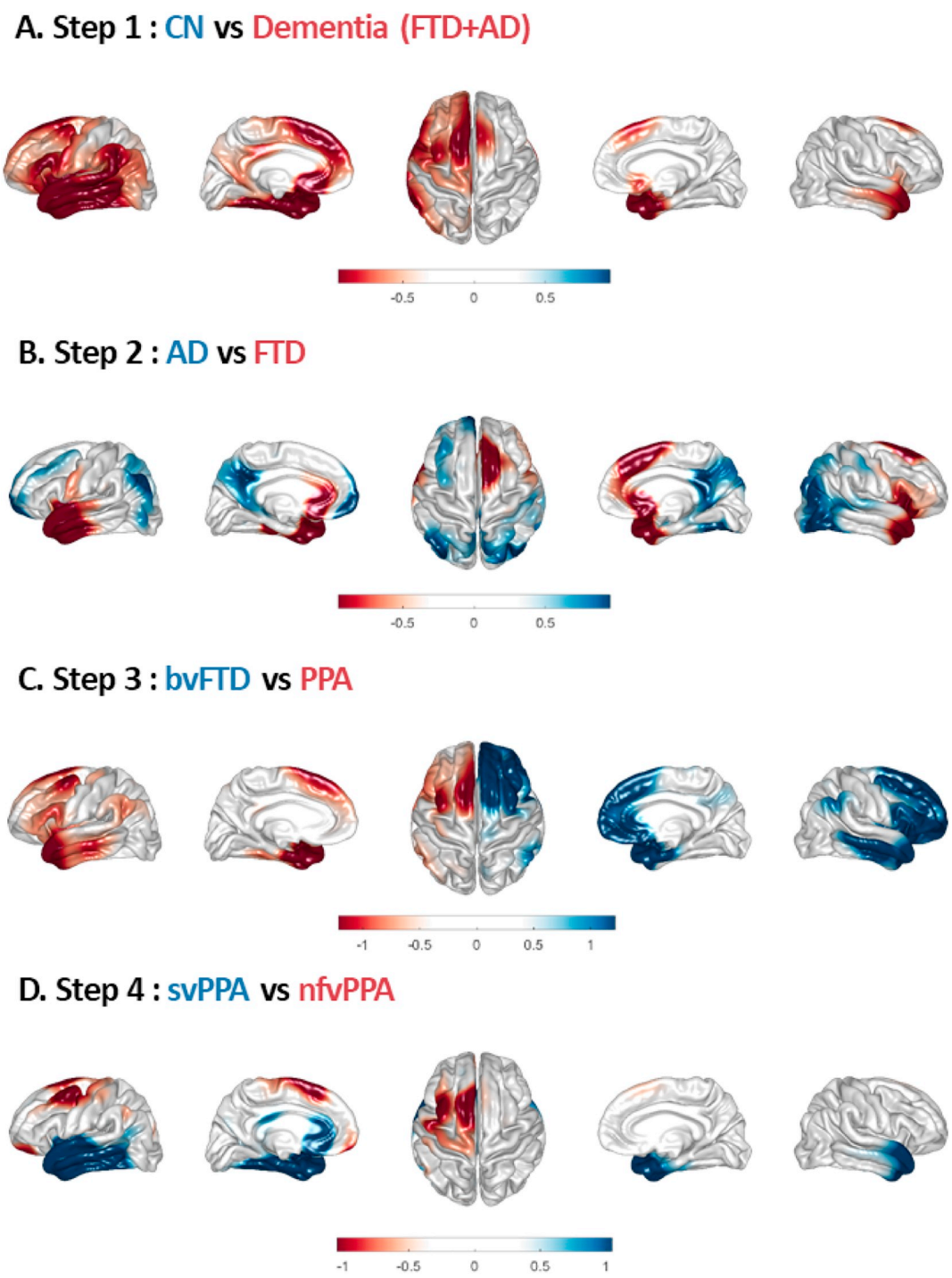

Fig. 3. Discriminative regions for each step. Each discriminative area corresponds to the group written in the same color.

distinguished AD from FTD, while the bilateral anterior temporal, anterior cingulate and right frontal regions distinguished FTD from AD. The left frontotemporal region and left inferior parietal lobule distinguished PPA from bvFTD, while the right frontotemporal regions significant influenced the in discrimination of bvFTD from nfvPPA (Step 3). For distinguishing svPPA from nfvPPA, the bilateral anterior temporal regions and left anterior cingulate cortex were of significantly influential in identifying svPPA, whereas the left frontal lobar regions were significant for identifying nfvPPA (Step 4).Abbreviations: FTD = frontotemporal dementia, bvFTD = behavioral variant frontotemporal dementia, PPA = primary progressive aphasia, svPPA = semantic variant primary progressive aphasia, nfvPPA = non-fluent/agrammatic variant primary progressive aphasia, AD = Alzheimer's disease, CN = cognitively normal

3.3. Post-hoc assessment for misclassified subjects

We further performed a post-hoc assessment for misclassified subjects. A total of 89 out of 742 pairwise classifications did not match the clinical diagnosis. In a visual review of these scans, 22 of them only had subtle atrophy which were not suggestive of a single clinical diagnosis (Fig. 4A). Nine scans showed significant atrophy but the spatial pattern of cortical atrophy was shared by more than one clinical diagnosis

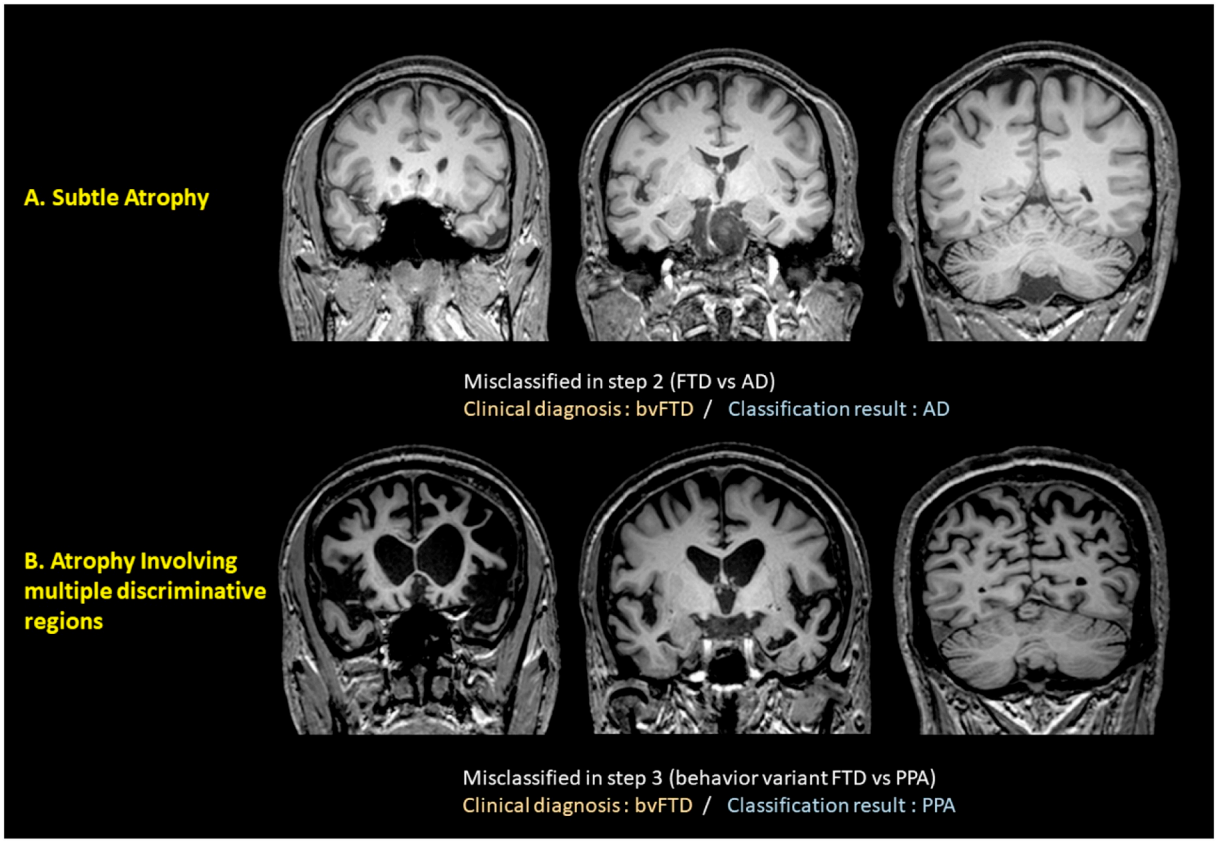

Fig. 4. Examples of misclassified subjects. (A) The MRI scan shows no definite atrophy. (B) The MRI scan shows significant atrophy in the bilateral frontotemporal areas. The atrophy is slightly worse on the left side, which might have led to the misclassification.

(Fig. 4B). For the remainder of the misclassified subjects, it was not obvious from visual review why misclassification occurred. There were subjects who were misclassified in multiple steps. Fourteen subjects were misclassified in two steps, and one subject was misclassified in three steps (Supplementary Table S3).

4. Discussion

We developed a machine learning-based automated classifier for differential diagnosis of FTD clinical syndromes. We included carefully phenotyped FTD patients, for whom precise clinical diagnoses were made through a consensus decision, for the development of our classifier. This classifier was successful in discriminating among CN, AD, and FTD subtypes using MRI-based cortical thickness data. Since AD is the most prevalent cause of dementia and the most important etiology to be considered for patients with cognitive decline, we also included AD patients in our classification model. Methodologically, the proposed method based on the Laplace operator removed high-frequency components of cortical thickness data as noise, which made the classification more sensitive. This was possible because we were able to overcome the spatial variance due to noise while maintaining significant differences of shape especially in FTD and AD groups. Thus, we believe that one advance of our study was the application of a Laplace Beltrami operator to the cortical thickness data, which allowed us to reduce the contribution of noise to the classification. Moreover, visualization of discriminative regions was possible by transforming the discriminative patterns in the frequency domain to that in the surface domain. Therefore, our study clearly shows that the classifier models discriminate each patient group with relative importance weights distributed across multiple brain regions. Furthermore, automated classifiers are expected to help in the clinical diagnosis of patients with subtle cortical atrophy. While a clinician might be biased to look for only a few well-known structural changes in structural MRI, our automated classifier can identify minute changes in co-varying regions. For example, in svPPA, clinicians may only pay attention to the anterior temporal lobe, while our study demonstrates that the anterior cingulate cortex also has significance in discriminating svPPA from nfvPPA (Fig. 3).

Compared with previous classifier studies, we used more diverse diagnostic categories of neurodegenerative dementia. Previous studies introduced classifiers discriminating AD patients from cognitively normal subjects, (Wee et al., 2013; Westman et al., 2013) as well as those distinguishing between AD, FTD, and CN groups(Davatzikos et al., 2008; Klöppel et al., 2008; Raamana et al., 2014; Moller et al., 2016; Bron et al., 2017; Bouts et al., 2018; Kloppel et al., 2008; Kloppel et al., 2015). Other studies demonstrated classifiers for individual level classification of PPA subtypes. (Wilson et al., 2009; Bisenius et al., 2017) In contrast, we built a more comprehensive classifier, which not only discriminates FTD, AD, and CN from each other, but also further classifies three clinical syndromes of FTD. We achieved good to excellent accuracies for classification between groups, especially between dementia groups. In discriminating FTD from AD, our classifier had an accuracy of 90.8%, demonstrating similar performance compared to literature reporting 72% to 89.2% accuracies. (Davatzikos et al., 2008; Klöppel et al., 2008; Raamana et al., 2014; Moller et al., 2016; Bron et al., 2017; Bouts et al., 2018; Kloppel et al., 2008; Kloppel et al., 2015) Our model classified subjects with nfvPPA and svPPA with 92.1% accuracy, which was similar to or higher than the 78% to 89% accuracies reported in previous studies. (Wilson et al., 2009; Bisenius et al., 2017; Agosta et al., 2015) We believe that both the surface-based feature extraction and Laplace-Beltrami operator-based noise removal have contributed to the improvement in performance.

The patterns of discriminative regions for our classifiers are similar to previously known cortical atrophic patterns in each clinical

syndrome. Although the discriminative regions of a certain group can vary depending on the combinations of groups compared, they generally reflect the structural changes occurring in that group. Compared with PPAs, the bvFTD group's discriminative regions showed right frontal predominance. This may have been influenced by the cortical atrophy pattern of bvFTD, which is known from past studies. (Rosen et al., 2002; Seeley et al., 2008) The discriminative regions of svPPA were most prominent in the left anterior temporal area, which is also consistent with results from previous studies. (Galton et al., 2001) The left frontal region was crucial in discriminating nfvPPA from other FTD clinical syndromes, also consistent with the previously known left frontal dominance in atrophy pattern in this condition. (Gorno-Tempini et al., 2006) Thus, our classifiers reflect the known cortical atrophy pattern for each clinical subtype.

The four steps of our hierarchical classifier emulate the clinical decision process for diagnosing FTD patients. First, it is important to determine whether the patient's complaints are due to normal aging. (step 1) When the patient seems to have behavioral or language symptoms that cannot be considered as changes of normal aging, the physician still has to rule out AD (step 2) since the prevalence of AD is much higher than FTD, and AD patients can show similar symptoms. If the patient's symptoms and signs suggest FTD, the clinician tries to determine which clinical subtype is the most likely (steps 3–4). The classification accuracy of step 4 was highest among the four steps, probably due to the distinctiveness of the cortical atrophy pattern in svPPA patients. (Davies et al., 2009; Hodges and Patterson, 2007; Brambati et al., 2015) Although our main goal was to develop a hierarchical classifier to emulate the clinical diagnostic process, the classifiers for each step can also be utilized in clinical practice. For example, it is sometimes difficult to determine whether a patient's behavioral symptoms suggest bvFTD or AD with prominent frontal dysfunction. Additionally, it is often hard to tell which type of PPA a patient with mild language dysfunction has. In these cases, individual classifiers can be used selectively.

We found that about 12% of classifications did not match the clinical diagnosis. This might be related to subtle atrophy or diffuse severe atrophy at the individual level. In our experiments, many of the misclassified patients (about 23%) had subtle cortical atrophy, which made it difficult for the classifier to capture the characteristic atrophy pattern. Indeed, such subtle changes in the cerebral cortex were barely detectable even in visual assessment performed by an expert neurologist. Another main reason for misclassification stems from the shared spatial pattern of cortical atrophy across multiple clinical diagnoses. As the supervised learning proposed in this study tries to detect different cortical atrophy patterns in two clinical diagnoses, similar atrophy patterns between them could lead to a misclassification. This issue is a well-known overfitting problem in supervised learning, and has been a major obstacle to the application of computer-aided diagnosis to medical images due to limited numbers of patients. Once we have more data, we believe this overfitting problem could be resolved, resulting in an improved performance for classification.

This study has several limitations. First, we did not have pathological diagnoses for most FTD patients, although we included carefully phenotyped patients. It is important for a subject's clinical syndrome and underlying disease to be distinguished, as not only could patients with clinical AD dementia potentially have underlying frontotemporal lobar degeneration, but patients with FTD clinical syndromes could also have AD pathology. The purpose of this study was to predict clinical syndromes rather than underlying diseases. However, to enhance the homogeneity within the groups, we used the results of amyloid PET scans for inclusion of AD and CN subjects. Second, four FTD patients were amyloid (+) on PET, which leaves open the possibility that frontal variant AD might be included in the FTD group. However, since we diagnosed FTD patients through a consensus decision committee, amyloid (+) in these patients might be incidental findings. In fact, the prevalence of amyloid (+) in the FTD group seemed to be similar to that of cognitively normal individuals.(Engler et al., 2008) Third, the discriminative regions depicted in our figure provide information on the statistically significant regions used in discriminating one group from another, but do not clearly demonstrate whether the cortical regions are thicker or thinner. Fourth, because we developed the classifier using carefully phenotyped subjects, performance may not be as high when applied to patients in the early stages of these diseases whose clinical phenotypes are not conclusive yet. In future studies, it would be meaningful to conduct similar analysis using MRI scans acquired in the early stage of the clinical course from subjects who were later carefully phenotyped. Finally, there may be a concern for overfitting when training an LDA classifier with a relatively small number of data samples. We indeed compared the classification performance by incorporating additional regularization terms for the LDA classifier, which unfortunately could not improve the 10-fold CV performances. Our future studies will focus on developing computationally more generalizable methods with comparable or better classification accuracy using an external dataset for validation if possible.

5. Conclusion

With our fully automated classifier, cortical thickness data alone could classify FTD clinical subtypes and AD with good to excellent accuracy. Our classifier may help clinicians to diagnose FTD subtypes with subtle cortical atrophy and facilitate the selection of appropriate interventions.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2019.101811.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) [No. NRF-2017R1A2B2005081]; Research of Korea Centers for Disease Control and Prevention [2018-ER6203-01]; and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) [No. 2016R1A2B4014398].

Acknowledgements

None.

References

Agosta, F., Ferraro, P.M., Canu, E., Copetti, M., Galantucci, S., Magnani, G., Marcone, A., Valsasina, P., Sodero, A., Comi, G., et al., 2015. Differentiation between subtypes of primary progressive aphasia by using cortical thickness and diffusion-tensor MR imaging measures. Radiology 276 (1), 219–227.

Bang, J., Spina, S., Miller, B.L., 2015. Frontotemporal dementia. Lancet 386 (10004), 1672–1682.

Barthel, H., Gertz, H.J., Dresel, S., Peters, O., Bartenstein, P., Buerger, K., Hiemeyer, F., Wittemer-Rump, S.M., Seibyl, J., Reininger, C., et al., 2011. Cerebral amyloid-beta PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. Lancet Neurol. 10 (5), 424–435.

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. 19 (7), 711–720.

Bisenius, S., Mueller, K., Diehl-Schmid, J., Fassbender, K., Grimmer, T., Jessen, F., Kassubek, J., Kornhuber, J., Landwehrmeyer, B., Ludolph, A., et al., 2017. Predicting primary progressive aphasias with support vector machine approaches in structural MRI data. NeuroImage Clin. 14, 334–343.

Bouts, M., Moller, C., Hafkemeijer, A., van Swieten, J.C., Dopper, E., van der Flier, W.M., Vrenken, H., Wink, A.M., Pijnenburg, Y.A.L., Scheltens, P., et al., 2018. single subject classification of Alzheimer's disease and Behavioral variant Frontotemporal dementia using anatomical, diffusion tensor, and resting-state functional magnetic resonance imaging. J. Alzheimers Dis. 62 (4), 1827–1839.

Brambati, S.M., Amici, S., Racine, C.A., Neuhaus, J., Miller, Z., Ogar, J., Dronkers, N., Miller, B.L., Rosen, H., Gorno-Tempini, M.L., 2015. Longitudinal gray matter contraction in three variants of primary progressive aphasia: a tenser-based morphometry study. NeuroImage 8, 345–355 Supplement C.

Bron, E.E., Smits, M., Papma, J.M., Steketee, R.M.E., Meijboom, R., de Groot, M., van Swieten, J.C., Niessen, W.J., Klein, S., 2017. Multiparametric computer-aided

differential diagnosis of Alzheimer's disease and frontotemporal dementia using structural and advanced MRI. Eur. Radiol. 27 (8), 3372–3382.
Cho, Y., Seong, J.-K., Jeong, Y., Shin, S.Y., 2012. Initiative AsDN: individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage 59 (3), 2217–2230.
Davatzikos, C., Resnick, S.M., Wu, X., Parmpi, P., Clark, C.M., 2008. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. NeuroImage 41 (4), 1220–1227.
Davies, R.R., Scahill, V.L., Graham, A., Williams, G.B., Graham, K.S., Hodges, J.R., 2009. Development of an MRI rating scale for multiple brain regions: comparison with volumetrics and with voxel-based morphometry. Neuroradiology 51 (8), 491–503.
Engler, H., Santillo, A.F., Wang, S.X., Lindau, M., Savitcheva, I., Nordberg, A., Lannfelt, L., Långström, B., Kilander, L., 2008. In vivo amyloid imaging with PET in frontotemporal dementia. Eur. J. Nucl. Med. Mol. Imaging 35 (1), 100–106.
Farrar, G., 2017. Regional visual read inspection of [18F] flutemetamol brain images from end-of-life and amnestic MCI subjects. J. Nucl. Med. 58 (supplement 1), 1250.
Galton, C.J., Patterson, K., Graham, K., Lambon-Ralph, M.A., Williams, G., Antoun, N., Sahakian, B.J., Hodges, J.R., 2001. Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. Neurology 57 (2), 216–225.
Gorno-Tempini, M.L., Ogar, J.M., Brambati, S.M., Wang, P., Jeong, J.H., Rankin, K.P., Dronkers, N.F., Miller, B.L., 2006. Anatomical correlates of early mutism in progressive nonfluent aphasia. Neurology 67 (10), 1849.
Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., Ogar, J.M., Rohrer, J.D., Black, S., Boeve, B.F., et al., 2011. Classification of primary progressive aphasia and its variants. Neurology 76 (11), 1006–1014.
Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.
Hodges, J.R., Patterson, K., 2007. Semantic dementia: a unique clinicopathological syndrome. Lancet Neurol. 6 (11), 1004–1014.
Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, J.C.R., Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681–689.
Kloppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., et al., 2008. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain 131 (Pt 11), 2969–2974.
Kloppel, S., Peter, J., Ludl, A., Pilatus, A., Maier, S., Mader, I., Heimbach, B., Frings, L., Egger, K., Dukart, J., et al., 2015. Applying automated MR-based diagnostic methods to the memory clinic: a prospective study. J. Alzheimers Dis. 47 (4), 939–954.
McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7 (3), 263–269.
Moller, C., Pijnenburg, Y.A., van der Flier, W.M., Versteeg, A., Tijms, B., de Munck, J.C., Hafkemeijer, A., Rombouts, S.A., van der Grond, J., van Swieten, J., et al., 2016. Alzheimer disease and Behavioral variant Frontotemporal dementia: automatic classification based on cortical atrophy for single-subject diagnosis. Radiology 279 (3), 838–848.
Pereira, J.M., Williams, G.B., Acosta-Cabronero, J., Pengas, G., Spillantini, M.G., Xuereb, J.H., Hodges, J.R., Nestor, P.J., 2009. Atrophy patterns in histologic vs clinical groupings of frontotemporal lobar degeneration. Neurology 72 (19), 1653–1660.
Qiu, A., Bitouk, D., Miller, M.I., 2006. Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace-Beltrami operator. IEEE Trans. Med. Imaging 25 (10), 1296–1306.
Raamana, P.R., Rosen, H., Miller, B., Weiner, M.W., Wang, L., Beg, M.F., 2014. Three-class differential diagnosis among Alzheimer disease, Frontotemporal dementia, and controls. Front. Neurol. 5, 71.
Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G.P., Onyike, C.U., et al., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134 (9), 2456–2477.
Rosen, H.J., Gorno–Tempini, M.L., Goldman, W.P., Perry, R.J., Schuff, N., Weiner, M., Feiwell, R., Kramer, J.H., Miller, B.L., 2002. Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology 58 (2), 198.
San Lee, J., Kim, C., Shin, J.-H., Cho, H., D-s, Shin, Kim, N., Kim, H.J., Kim, Y., Lockhart, S.N., 2018. Na DL: machine learning-based individual assessment of cortical atrophy pattern in Alzheimer's disease Spectrum: development of the classifier and longitudinal evaluation. Sci. Rep. 8 (1), 4161.
Seelaar, H., Rohrer, J.D., Pijnenburg, Y.A.L., Fox, N.C., van Swieten, J.C., 2011. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. J. Neurol. Neurosurg. Psychiatry 82 (5), 476.
Seeley, W.W., Crawford, R., Rascovsky, K., Kramer, J.H., Weiner, M., Miller, B.L., Gorno-Tempini, M.L., 2008. Frontal Paralimbic network atrophy in very mild Behavioral variant Frontotemporal dementia. Arch. Neurol. 65 (2), 249–255.
Vallet, B., Lévy, B., 2008. Spectral geometry processing with manifold harmonics. In: Computer Graphics Forum: 2008: Wiley Online Library, pp. 251–260.
Vieira, R.T., Caixeta, L., Machado, S., Silva, A.C., Nardi, A.E., Arias-Carrión, O., Carta, M.G., 2013. Epidemiology of early-onset dementia: a review of the literature. Clin. Pract. Epidemiol. Ment. Health 9, 88–95.
Wee, C.Y., Yap, P.T., Shen, D., Initiative, AsDN, 2013. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum. Brain Mapp. 34 (12), 3411–3425.
Westman, E., Aguilar, C., Muehlboeck, J.-S., Simmons, A., 2013. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topogr. 26 (1), 9–23.
Wilson, S.M., Ogar, J.M., Laluz, V., Growdon, M., Jang, J., Glenn, S., Miller, B.L., Weiner, M.W., Gorno-Tempini, M.L., 2009. Automated MRI-based classification of primary progressive aphasia variants. NeuroImage 47 (4), 1558–1567.