A comparison of three brain atlases for MCI prediction

Kenichi Ota [a], Naoya Oishi [a*], Kengo Ito [b], Hidenao Fukuyama [a], the SEAD-J Study Group [1]

[a] Human Brain Research Center, Kyoto University Graduate School of Medicine, Kyoto, Japan

[b] Clinical and Experimental Neuroimaging, National Center for Geriatrics and Gerontology, Obu, Japan

[1] Data used in the preparation of this article were obtained from the Research group of the Studies on Diagnosis of Early Alzheimer's Disease-Japan (SEAD-J), which comprised investigators from nine different facilities. As such, the investigators within the SEAD-J study group contributed to the design and implementation of SEAD-J and/or provided data but did not participate in the analysis or writing of this report.

Keywords: Alzheimer's disease (AD); mild cognitive impairment (MCI); magnetic resonance imaging (MRI); voxel-based morphometry (VBM); support vector machine (SVM); atlas-based parcellation.

Abstract

-Background

Although previous voxel-based studies using features extracted by atlas-based parcellation produced relatively poor performances on the prediction of Alzheimer's disease (AD) in subjects with mild cognitive impairment (MCI), classification performance usually depends on features extracted from the original images by atlas-based parcellation. To establish whether classification performance differs depending on the choice of atlases, support vector machine (SVM)-based classification using different brain atlases was performed.

-New Method

Seventy-seven three-dimensional T1-weighted MRI data sets of subjects with amnestic MCI, including 39 subjects who developed AD (MCI-C) within three years and 38 who did not (MCI-NC), were used for voxel-based morphometry (VBM) analyses and analyzed using SVM-based pattern recognition methods combined with a feature selection method based on the SVM recursive feature elimination (RFE) method. Three brain atlases were used for the feature selections: the Automated Anatomical Labeling (AAL) Atlas, Brodmann's Areas (BA), and the LONI Probabilistic Brain Atlas (LPBA40).

-Results

The VBM analysis showed a significant cluster of gray matter density reduction, located at the left hippocampal region, in MCI-C compared to MCI-NC. The SVM analyses with the SVM-RFE algorithm revealed that the best classification performance was achieved by LPBA40 with 37 selected features, giving an accuracy of 77.9%. The overall performance in LPBA40 was better than that of AAL and BA regardless of the number of selected features.

-Conclusions

These results suggest that feature selection is crucial to improve the classification performance in atlas-based analysis and that the choice of atlases is also important.

Introduction

Although there is no real treatment for Alzheimer's disease (AD), medications currently available to alleviate cognitive and behavioral symptoms of AD may delay clinical progression to AD (Petersen et al., 2005). AD pathophysiological processes precede the clinical manifestation of symptoms of AD. The earlier the intervention against symptomatic progression begins, the more effective the intervention may be. Earlier diagnosis or prediction of the conversion of mild cognitive impairment (MCI) to AD is therefore required for earlier intervention. It is necessary to explore effective biomarkers for early AD detection in the earliest stages of the disease.

The recently revised diagnostic criteria for AD incorporated biomarkers of neuronal injury such as brain atrophy (Albert et al., 2011). Currently, even clinical magnetic resonance imaging (MRI) scanners allow the acquisition of images with high spatial resolution, to provide AD biomarkers that reflect brain atrophy (Frisoni et al., 2010). Structural MRI-based features that are widely used in such pattern recognition methods include: voxel-based whole brain volume data, surface-based measures such as cortical thickness, and features based on specific regions of interest (ROIs) such as hippocampus and entorhinal cortex. In this context, MRI studies across the whole brain, such as voxel-based morphometry (VBM) techniques (Ashburner and Friston, 2000), could be more useful in investigating biomarkers for early AD detection, rather than those based on specific ROIs. For example, Karas et al. (2004) conducted a VBM study to analyze patterns of gray matter (GM) loss and revealed that subjects with MCI had significant local GM reductions in the medial temporal lobe, the insula, and the thalamus compared to normal elderly controls. GM loss on structural MRI can be shown in a specific topographic pattern involving temporal lobes and parietal cortices (Jack et al., 2011). VBM studies have also demonstrated that patterns of GM loss correlate

with neurofibrillary tangle pathology (Vemuri et al., 2008; Whitwell et al., 2008). AD neuropathology is mainly characterized by the extracellular deposition of fibrillary β-amyloid protein and intracellular formation of neurofibrillary tangles composed of abnormal tau protein (Nelson et al., 2009). Braak et al. (2011) reported that β-amyloid protein and neurofibrillary tangles were significantly correlated and that pathological aggregation of tau protein might begin earlier than previously thought; these authors also suggest that possibly these events occur in subcortical nuclei rather than in the transentorhinal region of the perirhinal cortex.

Pattern recognition methods based on machine-learning techniques such as the support vector machine (SVM) are promising tools for computer-aided diagnosis of AD (Klöppel et al., 2008, 2012). It is, however, currently unrealistic to use a hundred thousand voxels of MRI data for each scan directly for machine-learning-based pattern recognition, because of the possibility of poor generalization from overfitting, which could arise in a case in which the number of features is much larger than the number of subjects. Therefore, dimensionality reduction in the feature space, such as feature extraction and feature selection, is usually necessary to achieve good generalization, which is required for possible clinical utility.

Among various approaches to classifying subjects with MCI using structural MRI, voxel-based methods can be roughly classified into two categories: data-driven adaptive feature extraction methods (Fan et al., 2007; Misra et al., 2009; Davatzikos et al., 2011) and atlas-based parcellation methods, using a predefined brain atlas (Cuingnet et al., 2011; Cho et al., 2012). Data-driven feature extraction methods adaptively define ROIs depending on the input data without a predefined brain atlas, which are difficult to interpret anatomically. On the other hand, atlas-based parcellation methods allow feature vector extraction with good anatomical interpretability. An atlas-based parcellation approach can be used as an

anatomical filter for dimensionality reduction to construct gray matter volumes for dozens of predetermined ROIs from hundreds of thousands of voxels in each subject's whole-brain structural MRI. The extracted regional feature variables for each subject can be employed to create multivariate models to classify each subject into different groups (Whitwell et al., 2011). Examples of such an anatomical brain atlas include Brodmann's areas (BA; Brodmann, 1909) and the Automated Anatomical Labeling Atlas (AAL; Tzourio-Mazoyer et al., 2002). These atlases are publicly available from the Internet, in open source software packages (MRIcro/MRIcron, http://www.mricro.com/). Although previous benchmark studies showed poor classification performance of an atlas-based method (Cuingnet et al., 2011; Cho et al., 2012), classification performance strongly depends on the features that are extracted from the atlas used for parcellation. It remains unclear whether atlas-based parcellation methods applied to different atlases would in fact lead to different classification performances.

The aim of our study was to establish whether the performance when predicting conversion to AD using GM volumes from the structural MRI of subjects with MCI differs depending on the choice of the atlas. To accomplish this goal, SVM-based classification using three different brain atlases was used. We also investigated whether a feature selection method could enhance the classification accuracy.

Materials and methods

Subjects

Studies on the Diagnosis of Early Alzheimer's Disease—Japan (SEAD-J), a prospective multicenter cohort study of subjects with amnestic MCI was started in 2005 by the National Center for Geriatrics and Gerontology, to achieve the early prediction of AD

conversion (Kawashima et al., 2012). In this cohort study, 114 subjects with amnestic MCI were recruited from nine different facilities across Japan (Supplementary Table 1) between January 2006 and March 2007. All of the subjects were living independently in the community at the time of their baseline evaluation. This study was approved by the Ethics Committee at every participating institution. Each subject signed an informed consent form after a full explanation of the procedures had been offered.

Diagnosis of MCI was based on an interview with neurologists that contained evidence of reduced cognitive capacity, normal activities of daily living, and the absence of dementia (Cui et al., 2011). All of the patients were free of significant underlying medical, neurological, or psychiatric illness. The patients were initially accessed using a neuropsychological test battery, including the Mini-Mental State Examination (MMSE; Folstein et al., 1975), Alzheimer's Disease Assessment Scale-Cognitive Subscale, Japanese version (ADAS-J cog; Homma et al., 1992), Clinical Dementia Rating (CDR; Morris, 1993), Geriatric Depression Scale (GDS; Yesavage et al., 1982; Nyunt et al., 2009), Everyday Memory Checklist (EMC; Kazui et al., 2003), and the Wechsler Memory Scale-Revised Logical memory test (WMS-R LM; Sullivan, 1996). In accordance with the inclusion criteria, MCI patients were between 50 and 80 years old, with an MMSE score ≥ 24, a GDS score ≤ 10, a WMS-R LM I score ≤ 13, an LM II part A and part B score (maximum, 50) ≤ 8, and a CDR memory box score restricted to 0.5. Patients who had an educational level (defined as the number of completed years of formal education) of under 6 years were excluded. The patients were observed at 1-year intervals for 3 years and underwent the following standardized procedures. Trained clinicians performed baseline and follow-up 1-year evaluations. The CDR, MMSE, EMC, and WMS-R-LM were completed at each visit during follow-up. ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) and MRI were optional during follow-up. The ADAS-J cog was also administered as an option in

selected centers. Conversion to dementia was established when CDR became ≥ 1. The diagnosis of AD was made in a given center if a patient fulfilled both CDR ≥ 1 and the National Institute of Neurological Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) probable AD criteria (McKhann et al., 1984). The diagnosis of other causes was based on established clinical criteria for each disease, including vascular dementia (VaD) (Devanand et al., 2010), dementia with Lewy bodies (DLB) (McKeith et al., 1996), frontotemporal dementia (FTD) (McKhann et al., 2001), and Creutzfeldt-Jakob disease (CJD) (Knopman et al., 2001).

Of all of the 114 participants, 37 subjects were excluded from our analyses due to the following reasons: 2 had no baseline three-dimensional T1-weighted MRI scans, 3 converted to non-AD dementia (VaD, DLB, and FTD), 23 withdrew from the study within 3 years, and 9 were excluded due to the lack of whole-brain coverage in their baseline T1-weighted MRI scans.

As a result, we identified 77 subjects with amnestic MCI from the SEAD-J comprising 39 who developed AD within 3 years (AD converters, MCI-C; 20 females, 19 males; age ± SD = 71.3 ± 6.7 years ranging from 50 to 79 years; and MMSE = 25.6 ± 1.8 ranging from 24 to 30) and 38 who did not (non-converters, MCI-NC; 22 females, 16 males; age ± SD = 70.6 ± 6.9 years ranging from 55 to 79 years; and MMSE = 27.0 ± 2.0 ranging from 24 to 30). Among these AD converters, 21 converted to AD within 1 year after inclusion, 14 within 2 years, and 4 within 3 years. Baseline demographic data and neuropsychological test results of these subjects are shown in Table 1. The two groups significantly differed in the MMSE, ADAS-J cog, WMS-R-LM, and GDS scores at the baseline. No significant differences were observed in the age, gender, and education.

MRI acquisition and preprocessing

A three-dimensional structural MRI at the baseline was acquired on each subject with T1-weighted gradient echo sequences on 1.5 T or 3.0 T MRI scanners at the nine facilities. Details about MRI acquisition conditions are provided in Supplementary Table 2.

We used the SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8) and VBM8 Toolbox (Kurth et al., 2010; http://dbm.neuro.uni-jena.de/vbm) on MATLAB 7.12 for preprocessing the baseline MRI data. The images were first segmented into GM, white matter and cerebrospinal fluid using Unified Segmentation (Ashburner and Friston, 2005) implemented in SPM8 and a technique based on the maximum *a posteriori* (MAP) estimation (Rajapakse et al., 1997) and the Partial Volume Estimation (PVE) (Tohka et al., 2004) implemented in VBM8 with standard parameters. Then, the segmented images were spatially normalized using the Diffeomorphic Anatomical Registration using the Exponentiated Lie algebra (DARTEL) algorithm (Ashburner, 2007). Jacobian modulation was applied to compensate for the effect of spatial normalization and to restore the original absolute GM density in the segmented GM images. The normalized, segmented, and modulated images were smoothed with an 8-mm full-width at half-maximum isotropic Gaussian kernel.

Voxel-based morphometry (VBM) analysis

The smoothed baseline MR images were analyzed using a conventional VBM method (Ashburner and Friston, 2000) to investigate the differences in the density of the GM between the MCI-C and the MCI-NC group. We conducted a statistical analysis that included an adjustment for age, gender, and scan facilities as covariates using SPM8. The statistical threshold was set at $p<0.001$, was uncorrected for multiple comparisons and was cluster-level corrected for multiple comparisons ($p<0.05$).

Feature extraction using atlas-based parcellation

We performed atlas-based parcellation to extract feature vectors from the segmented, normalized MR images using three different anatomically labeled brain atlases: the AAL atlas, BA, and the LONI Probabilistic Brain Atlas (LPBA40; Shattuck et al., 2008) (Fig. 1).

The AAL atlas is a single-subject atlas based on the Montreal Neurological Institute (MNI) Colin27 T1 atlas. This MNI single-subject brain template was obtained from 27 high-resolution T1-weighted scans of a young male. Each acquisition was spatially normalized to the MNI305 average template using a linear nine-parameter transformation (Holmes et al., 1998). In each hemisphere, 45 ROIs were drawn manually every 2 mm on the axial slices of the MNI single-subject brain. In addition, AAL includes a cerebellar parcellation with 26 ROIs (Schmahmann et al., 1999, 2000) (Fig. 1). Finally, 116 ROIs were defined, including the cerebellum for the AAL atlas.

The BA of the human cortex originally shows 43 cytoarchitectonic areas (Brodmann, 1909), where areas with the numbers 12–16 and 48–51 are not shown (Zilles et al., 2010). We used the BA atlas in MRIcro with 41 areas, where BA 31 (dorsal posterior cingulate area), 33 (pregenual area), and 52 (parainsular area) were not included, and BA 48 (retrosubicular area) was included. For the purpose of the comparison with the other two atlases, we subdivided each of the 41 areas of BA symmetrically with respect to the mid-sagittal plane to obtain 82 ROIs in total (Fig. 1). Both AAL and BA in MRIcro were based on the "ch2" image, which was created using 27 scans from a single individual (Holmes et al., 1998).

The LPBA40 atlas is a population-based probabilistic atlas that is constructed from high-resolution T1-weighted MRI scans of 40 healthy, normal volunteers comprising 20 males and 20 females with the average age of 29.2 ± 6.3 years (mean ± S.D.; min = 19.3, max

= 39.5) (Shattuck et al., 2008). LPBA40 has three variants depending on the spatial normalization strategy. We used the LPBA40/AIR version. Each of the 40 volumes was aligned to the ICBM 452 T1 Warp 5 Atlas (ICBM452W5). The ICBM452W5 was created from 452 brains, and each volume was normalized to MNI305 average brains using a linear 12-parameter transformation and a subsequent non-linear 5th-order polynomial warping. In LPBA40, a total of 56 structures were manually labeled, including 50 cortical structures, 4 subcortical nuclei, the brainstem, and the cerebellum (Fig. 1).

We obtained deformation fields in the same manner as the above spatial normalization. Then we applied the forward deformation field to each atlas ROIs to map to MNI space. The mean GM density within each ROI, as calculated by modulation with the Jacobian, was computed with linear regression to adjust image quality differences among the facilities and used as the feature vectors.

Classification using a support vector machine (SVM)

Multivariate pattern recognition analysis using machine-learning methods has been applied to the classification of MCI subjects. In particular, SVM (Vapnik, 1998) is one of the widely used methods because of its remarkable performance of classification as well as the simplicity of its theory and implementation. For example, Aguilar et al. (2013) demonstrated that SVM was superior to other multivariate classifiers for classification of subjects with AD and cognitive normal and SVM also provided similar predictive values for MCI differentiation although there were no significant differences between classifiers. Accordingly, we chose SVM as a classifier for predicting the conversion from MCI to AD because we anticipated achieving a better classification performance.

We used an SVM classifier with a radial basis function (RBF) in accordance with the practical guide by Hsu et al. (2003). The classification performance was assessed using a leave-one-out cross-validation (LOOCV) strategy. We used MATLAB 7.6 and the LIBSVM library (Chang and Lin, 2011; Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm) to implement an RBF-kernel SVM with LOOCV. The hyperparameters (C, γ) of the RBF kernel were optimized using a two-step grid-search technique with 5-fold cross-validation according to the recommendation described in a practical guide to SVM classification (Hsu et al., 2003). First, the best pair of C_{coarse} and γ_{coarse} was found by a coarse grid-search on $\log_2 C = -5, -3, \ldots, 15$ and $\log_2 \gamma = -15, -13, \ldots, 3$. Then, the best pair of C_{fine} and γ_{fine} was obtained by a fine grid-search on $\log_2 C = C_{coarse}-2, C_{coarse}-1.75, \ldots, C_{coarse}+1.75, C_{coarse}+2$ and $\log_2 \gamma = \gamma_{coarse}-2, \gamma_{coarse}-1.75, \ldots, \gamma_{coarse}+1.75, \gamma_{coarse}+2$. The best ($C_{fine}$, γ_{fine}) was used to generate the final classifier for each training set.

Feature selection using the SVM recursive feature elimination (SVM-RFE)

In general, feature sets that are extracted from an input data set still contain redundant or irrelevant features as well as those that are important for classification. The higher the dimension the feature space is, the worse the performance of the classification. The ultimate goal of pattern recognition is to improve the generalization.

To enhance the performance of SVM-based classification, we performed a feature selection method that was based on the support vector machine recursive feature elimination (SVM-RFE) algorithm (Guyon et al., 2002). The SVM-RFE algorithm uses SVM to produce a feature ranking. Features that do not contribute to separation are eliminated according to the feature ranking. We used a linear kernel in the SVM-RFE procedure in the same manner as the original algorithm proposed by Guyon et al. (2002), and implemented the algorithm using

MATLAB 7.6 according to the literature (Guyon et al., 2002). The hyperparameter C of the linear kernel was optimized using a two-step grid-search technique with 5-fold cross-validation in a similar manner to the above. A total of 77 feature ranked lists were obtained by the following SVM-RFE LOOCV iterative process:

Inputs: feature vectors $\mathbf{X}_0 = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_i, \mathbf{x}_n]^T$ and class labels $\mathbf{y} = [y_1, y_2, ..., y_i, ..., y_n]^T$

for $i = 1$ to n

Split $\mathbf{X}_0$ into a test set (subject i) and a training set (the remaining subjects)

Initialize:

Subset of surviving features: $\mathbf{s} = [1, 2, ..., m]$

Feature ranked list: $\mathbf{r} = []$

repeat

Restrict training set to good feature indices

Optimize hyperparameters of linear and RBF kernel SVM classifier

Train the classifiers

Compute the weight vector

Compute the weight magnitude as ranking criterion

Find the feature with smallest ranking criterion

Update feature ranked list

Eliminate the feature with smallest ranking criterion

Predict the test set with the RBF kernel classifier

until $\mathbf{s} = []$

end for

Output: feature ranked list $\mathbf{r}$

After SVM-RFE, we computed classification measures for feature j; accuracy ACC_j as the percentage of trials that were correctly classified, sensitivity SEN_j as the percentage of trials

that were correctly classified as MCI-C, and specificity SPC_j as the percentage of trials that were correctly classified as MCI-NC. The number of features that provided the best accuracy in the SVM-RFE procedure was used as the number of features selected after SVM-RFE. As a result of SVM-RFE, we obtained a 77 x m matrix L (Fig. 2), where m is the number of features of each data set. Each row of the matrix L corresponds to the feature ranked list r for the ith LOOCV iteration. The feature ranked list r is a 1 x j matrix of the features ordered by relevance. The first element of the feature ranked list r had the index of the most relevant feature. More specifically, L(i, j) refers to the feature number of the rank j of the ith LOOCV iteration. To obtain a final ranking of features, we first converted the matrix L to the RFE rank matrix R according to the equation R(i, L(i, j)) = j. We then converted the resulting RFE rank matrix R to the RFE rank score matrix according to the following equation:

$S(i, j) = 0 \ (R(i, j) > k)$

or

$S(i, j) = (k - R(i, j) + 1) / k \ (R(i, j) \leq k)$

where S(i, j) is the score for feature j of the ith LOOCV iteration ($0 \leq S \leq 1$); k is the number of selected features after SVM-RFE for the data set; and R(i, j) is the rank number of feature j in the matrix R ($1 \leq R \leq m$; m is the number of features of the data set). Then we computed the sum of each column of the RFE rank score matrix S and normalized the resulting matrix between 0 and 1 to obtain a 1 x m feature rank score matrix. We sorted these feature rank scores into numerical order to obtain a final selection ranking of features after SVM-RFE for each atlas data set (Supplementary Table 3).

Statistical analysis

In order to investigate whether the classification accuracies were significantly different from one another, we applied leave-one-out cross validation a hundred times using

different atlases and with features selected through the SVM-RFE procedure. Two-way ANOVA followed by Tukey's multiple comparison test was performed for statistical analysis of atlas and feature selection using the statistical software packages R version 2.15.2 (The R Foundation for Statistical Computing, http://www.r-project.org/). To examine the behavior of the classifiers, we generated receiver operating characteristics (ROC) curves and computed areas under the curve (AUC) and 95% confidence intervals (CI) using the pROC package for R (Robin et al., 2011). CIs for AUCs were computed with DeLong's method. Pairwise comparisons of ROC curves were performed using DeLong's test implemented in the pROC package. LOOCV was also used to test these final classifiers.

Results

VBM analysis

Fig. 3 shows the result of VBM analysis. We found a significant cluster of GM density reduction in MCI-C compared with MCI-NC. The 2,391 mm^3 cluster was located in the parahippocampal gyrus and hippocampus on the left side (peak voxel at MNI coordinate −26, −42, −9). The reverse contrast (MCI-NC<MCI-C) showed no significant regions of GM loss.

Features selected with SVM-RFE

Fig. 4 illustrates the RFE rank score matrices S from three brain atlases. The vertical axis of the matrix represents a subject number, i.e., each step of the LOOCV procedure. The horizontal axis represents the number of features in each atlas. The top-ranked feature (the

feature that was selected last during the SVM-RFE procedure) having a score of 1 was colored in white. The score of the features that were not selected was 0 and colored in black. For example, in the map from the AAL, for the left hippocampus, feature number 37 was the most often selected during the LOOCV procedure. As can be seen from Fig. 4, "hot" regions that were often selected and "cold" regions that were rarely selected almost tidily line up vertically, suggesting that similar regions were selected through each step of the LOOCV procedure.

Fig.5 lists the regions that were selected with the SVM-RFE procedure, which revealed the highest performance in the three brain atlases. In AAL, 20 regions out of 116 were selected, 20 out of 82 in BA, and 37 out of 56 in LPBA40. The left hippocampus was in the highest rank in AAL, which is consistent with the results of the VBM analysis. In BA, the left parahippocampal region was ranked first. In LPBA40, the left parahippocampal gyrus was ranked second following the left inferior occipital gyrus.

Fig. 6 illustrates selected region maps with the SVM-RFE procedure overlaid to representative structural MR images. The regions with the highest rank are colored in white and the lowest rank are colored in black. The results of AAL and BA were similar, e.g., the left hippocampal region was the most often selected. In contrast, more regions that included the left hippocampal region were often selected in LPBA40.

SVM classification combined with SVM-RFE

Fig. 7 shows the plots of the classification accuracy versus the number of selected features in the dataset extracted with each brain atlas and selected based on the SVM-RFE. First, the classification accuracy differs depending on the choice of atlases. LPBA40 allowed the highest accuracy, AAL was the second, and BA was the lowest. Second, all the plots had

a peak between the minimum and maximum numbers of features. The number of features at the peak, i.e., the number of features selected in the SVM-RFE procedure, was the largest in LPBA40.

Table 2 lists the sets of accuracy, sensitivity, and specificity for the classification without feature selection and with feature selection based on the SVM-RFE algorithm. As a result of the SVM-RFE-based feature selection, 20 features were selected in AAL and BA, and 37 features were selected in LPBA40. The feature selection method improved the classification accuracy in all of the atlases. The best classification performance was obtained by using LPBA40 with 37 features, giving a correct classification rate of 77.9%, a sensitivity of 76.9%, and a specificity of 78.9%. AAL with 20 features distinguished MCI-C from MCI-NC with 71.4% accuracy, 69.2% sensitivity, and 73.7% specificity. BA with 20 features reached 67.5% accuracy, 64.1% sensitivity, and 71.1% specificity.

Fig. 8 demonstrates classification accuracies obtained with features extracted using different atlases (left) and features further selected through the SVM-RFE procedure (right). Note that the accuracies shown in Fig. 8 were obtained with the fixed set of 37 features that we chose based on the ranking after SVM-RFE, while the accuracies shown in Fig. 7 were calculated with the 80 LOOCV results of different classifiers that were trained with different combinations of 37 features. These combinations of 37 features were similar but not identical to one another and to the fixed set based on the feature ranking. There were significant main effects and interactions of atlas and feature selection ($p<0.0001$) except between AAL and BA without feature selection ($p = 0.16$).

Fig. 9 shows ROC curves with AUC and 95% CI obtained with different atlases using the original features and the features further selected through the SVM-RFE procedure. All *p*-values for feature selection differences were smaller than 0.0001 in all the atlases. Without feature selection, *p*-values for atlas differences were 0.84 for AAL vs. BA, 0.0055 for AAL

vs. LPBA40, and 0.014 for BA vs. LPBA40. In contrast, using feature selection, there were no significant differences (p>0.05) between the pairs of atlases.

Discussion

This study focused on feature extraction using atlas-based parcellation and feature selection based on the SVM-RFE algorithm in SVM-based classification using GM volumes from baseline structural MRI of subjects with amnestic MCI. To date, we are not aware of any study that has demonstrated a comparison of brain atlases for feature extraction.

The SEAD-J study showed a higher conversion rate for year 1 compared with the ADNI study (Kawashima et al., 2012). The inclusion criteria of SEAD-J were different from that of ADNI, for example, in WMS-R LM II score. The cohort of SEAD-J included amnestic MCI patients with severer verbal memory deficits compared with ADNI. Tabert et al. (2006) reported that deficits in verbal memory strongly predicted conversion to AD. Thus, this higher conversion rate might be due to the severity of memory deficit of the SEAD-J cohort, which is likely to be attributed to the inclusion criteria of SEAD-J.

We classified 77 subjects in this study into late MCI (LMCI) and early MCI (EMCI) on the basis of their objective memory loss measured by education-adjusted scores on WMS-R LM II according to the definition of LMCI and EMCI in the inclusion criteria of the ADNI 2 study (page 27 of the ADNI 2 Procedures Manual, http://adni.loni.ucla.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf). As a result, 60 subjects (77.9% of total) were classified into LMCI and 17 subjects (22.1%) EMCI. More specifically, 39 converters (MCI-C) in our study consisted of 37 LMCI (94.9%) and 2 EMCI (5.1%), whereas 38 non-converters (MCI-NC) included 23 LMCI (60.5%) and 15 EMCI (39.5%). The

proportions of LMCI and EMCI between the MCI-C and MCI-NC groups were significantly different (Fisher's exact test, p-value = 0.00028).

The result of the VBM analysis was consistent with that of a previous meta-analysis of VBM studies (Ferreira et al., 2011). The correspondence could demonstrate the validity of the MRI data and the methodology of VBM that we used in this study. The results have their own limitations, which are derived from the MRI data being acquired on multiple scanners at different research institutions. After adjusting for interscanner variability in the quality of the MRI acquisitions in both VBM and SVM analyses, we obtained classification accuracies in the range of 55%-78% (mean±SD = 65.8%±9.1%) that were comparable with those of previous studies (56%-82%, mean±SD = 66.8%±7.0%) (Eskildsen et al., 2013). These results indicate that the effect of scanner differences on the results of this study might not be significant, as shown in a previous VBM study using MRI data from different scanners (Stonnington et al., 2008). In addition, because the goal of this study was to evaluate the relative differences in the classification performances of the different brain atlases, the MRI data from multiple scanners probably made no remarkable difference in the results of the comparison among the atlases.

We applied an LOOCV technique where a test set and a training set were initially separated before the SVM-RFE procedure. The training set at each step of the SVM-RFE did not include the test set. The cross-validation procedure may prevent the overfitting problem (Hsu et al., 2003). Thus, it may be unlikely that the classification accuracies we obtained are inflated accuracies due to overfitting.

Different atlases for parcellation may cause differences in feature vectors constructed from the original whole-brain volumetric image on the basis of the following major factors: (1) a method to parcel the whole brain and (2) number or size of regions. Different atlases having different number of ROIs provide different feature vectors having different inter-

regional correlations (Wang et al., 2009; Faria et al., 2012). Multivariate analysis utilizes the spatial covariance structure in the data (Habeck et al., 2008). Differences in topological patterns of feature vectors in feature space thus may affect the decision boundary of a multivariate classifier. Accordingly, different inter-regional correlations due to different parcellation atlases can influence the classification accuracy in the multivariate pattern analysis.

The overall classification performance in this study was better than or comparable with the results of previous studies on the early prediction of AD using MRI-based biomarkers (Cuingnet et al., 2011; Davatzikos et al., 2011; Wolz et al., 2011; Cho et al., 2012; Eskildsen et al., 2013). The study demonstrated the classification performance differed across atlases when no feature selection was applied, using the same dataset and the same methods except for different atlases to define ROIs in voxel-based analysis. Although there were no significant differences in AUC, classification accuracies revealed significant differences across atlases when SVM-RFE was applied. To find the "optimal" atlas for AD prediction, however, replication in another cohort would be required to demonstrate that the found prediction accuracy was not merely by chance on the particular cohort studied.

Then, what are the reasons for providing such a considerable disparity in the performance across the three atlases? Although underlying causality remains unknown, clues for solving this question, if any, could be found in the differences between the atlases. The AAL and BA atlases with the “ch2” image gave similar classification accuracies, whereas LPBA40 with the ICBM452W5 template differed in their classification performance from the other two. Therefore, we mainly contrast AAL with LPBA40 for simplicity.

Brain atlases are classified into two categories: single-subject topological atlases and population-based probabilistic atlases (Cabezas et al., 2011). AAL is a single-subject atlas that is based on the brain of a young male (Tzourio-Mazoyer et al., 2002), whereas LPBA40

is a probabilistic atlas created from 40 MRI volumes (Shattuck et al., 2008). We speculate that this difference might be a major important difference between AAL and LPBA40. No single brain is representative of a population because of the neuroanatomical variability across individuals (Devlin and Poldrack, 2007). There is, therefore, no "correct" single-subject atlas. For example, the MNI single-subject brain has some problems because of anatomical variation and methodological limitations in spatial normalization.

Regarding anatomical variation, Tzourio-Mazoyer et al. (2002) mentioned that the MNI single-subject brain of AAL showed an atypical rightward asymmetry of the planum temporale (PT). The PT is a triangular structure that is located on the superior temporal gyrus (STG) and that has extensive connections to (and from) other regions of the brain. The PT could be engaged in mediating sensorimotor control processing such as speech motor processing (Zheng, 2009). PT asymmetry might be influenced by gender, and this rightward anatomical variation in the MNI brain of a young man was found in only approximately 10% of the subjects in a previous study (Shapleske et al., 1999). Chance et al. (2011) reported that microanatomical changes in cortical minicolumn organization of the association cortex in the PT (BA22) were detected in the early stages of MCI as well as AD. Such minicolumn measures in the temporal lobe reportedly reflect selective regional vulnerability to AD tangle pathology and differential involvement in the cognitive deficit of AD (Chance et al., 2006). Involvement of the superior temporal cortex in early atrophic changes in AD was also found in a VBM study on patterns of GM loss in MCI and AD (Karas et al., 2004). Furthermore, the SVM analyses in this study demonstrated that BA22 and STG in the right hemisphere were selected via the SVM-RFE procedure in BA and LPBA40, respectively, whereas in AAL, the STG in each hemisphere was eliminated. STG was also chosen by a feature selection method that was different from the SVM-RFE procedure for the classification of MCI using a linear discriminant analysis (Eskildsen et al., 2013). These findings suggest that the atypical PT

asymmetry in the MNI single-subject brain might pertain to the relatively poor performance of a whole-GM SVM classification of MCI using AAL-based parcellation.

Tzourio-Mazoyer et al. (2002) also reported that several sulcal patterns in the MNI single-subject brain, such as the Rolandic sulcus and the precentral sulcus in the left hemisphere, had a low probability with reference to Ono's atlas of sulci (Ono et al., 1990). In AAL, the ROIs in each hemisphere were defined using sulcal landmarks as the limits of the ROIs on the outer surface of the brain. The internal limit of the regions was extended beyond the gray matter layer, because AAL was originally intended to provide a standard reference frame of anatomical localization for functional neuroimaging studies with generally lower spatial resolution compared to anatomical MRI (Tzourio-Mazoyer et al., 2002). However, sulcal and gyral patterns are extremely variable, and macroanatomical landmarks do not match cytoarchitectonic borders in almost all of the cases (Amunts et al., 2007). In fact, AAL provides peak labeling, not precise anatomical localization, for structural imaging studies (Tzourio-Mazoyer et al., 2007). Thus, a single-subject atlas such as AAL does not represent the individual diversity of human anatomy (Toga et al., 2007; Cabezas et al., 2011).

These issues in AAL suggest that care must be taken to apply AAL to structural MRI analyses of subjects with MCI, as many experts suggested (Devlin and Poldrack, 2007; Toga et al., 2006, 2007; Tzourio-Mazoyer et al., 2002, 2007; Evans et al., 2012).

Although LPBA40 better represents the MCI cohort in this study compared with AAL, LPBA40 might not be the best choice. Population-based templates also lack inter-subject correspondence in cortical folding (Mangin et al., 2010). Furthermore, LPBA40 also differed from the MCI cohort in this study in terms of age, race, and disease, which is similar to in AAL. Cortical thickness analysis with the surface-based atlases in FreeSurfer (Desikan et al., 2006; Destrieux et al., 2010; http://surfer.nmr.mgh.harvard.edu/) using the same dataset as in this study might provide improved classification performance. Future studies that use a

disease-specific population-based atlas for MCI would also better serve the early detection of AD (Toga et al., 2006, 2007).

Another difference between the atlases is the number of ROIs in each atlas. Although a left hippocampal region was detected as a robust discriminating region by a univariate analysis (Fig. 3), our multivariate analysis demonstrated that classification using multiple features rather than a sole well-discriminating predictor could lead to better performance. When using the AAL atlas, the left hippocampus (Region 37) was most often selected as the last remaining feature after SVM-RFE. Using only this feature can provide a relatively good accuracy around 70% for the dataset we used in this study. In some circumstances, when adding more features, the resulting set of features may be more discriminating as a whole than the only feature in multivariate analysis. In other cases, the resulting set of features may be less discriminating than the original set and the resulting accuracy may be worse than when not adding the features. A set of features generated from 116 regions of the AAL atlas may generally contain a lot of less discriminating features. Therefore feature selection can be crucial when using atlases. A previous multivariate analysis of MR images of subjects with MCI also achieved improved accuracy in classification using linear discriminant analysis (LDA) with multiple features (Wolz et al., 2011). Our study suggests that an optimal number of regions could result in good performance in multivariate analysis and that too many regions also could lead to poor performance due to overfitting. However, it is difficult to determine the optimal number of regions in advance of feature extraction. During the SVM-RFE procedure, regions that do not contribute well to the separation are removed from the original feature set according to a feature-ranking algorithm. Whether an individual region separates the classes well or not is determined by how to parcellate a brain template. Multimodal probabilistic atlases generated by integrating the cytoarchitectonic, receptor

architectonic and functional imaging data (Toga et al., 2006) will play an important role in MRI data analyses.

BA and AAL gave similar results in both the classification performance and the regions selected through the SVM-RFE procedure. A parcellation that is too coarse would not reflect the underlying cytoarchitecture in each coarse region, as Amunts et al. (2007) concluded from their classification results using unsupervised cluster analysis on seven occipital areas of ten human brains. One of the possible explanations for why BA has poorer performance compared with AAL was the coarse parcellation in BA, although the ROI generation of BA by subdivision at the mid-sagittal plane might be imprecise. However, the poor classification performance in BA could primarily be attributed to the same problems in the MNI single-subject brain as in AAL, because LPBA40, which has the smallest number of regions, provided the best performance.

As seen from Figs. 5 and 6, the left hippocampal region was not consistently selected among the atlases. Fig. 10 showed the left parahippocampal region and left hippocampal region in each atlas corresponding to the regions detected by the VBM analysis shown in Fig. 3. The left hippocampal region differed across the atlases as shown in Fig. 10. Moreover, because the number of ROIs was also different among the atlases, correlations between the left hippocampal region and the other regions, i.e., correlation or covariance patterns also differed across the atlases. Multivariate analysis utilizes the spatial covariance structure in the data (Habeck et al., 2008). Different covariance patterns due to different atlas-based parcellations might cause the inconsistency in the selection of the left hippocampal region among the atlases as well as the differences in the performance of multivariate classifiers.

Similarly to this study, Chu et al. (2012) also employed similar methods for MCI prediction. Namely, they also used the LPBA40 atlas for atlas-based parcellation, the SVM-RFE method for feature selection, and SVMs for classification among AD, MCI, and normal

controls. However, differently from our study, they basically adopted a voxel-wise data-driven feature selection approach using high-dimensional whole brain voxel data (299,477 voxels) as the original input features. Moreover, they used seven atlas-based ROIs and two combinations thereof as prior knowledge for feature selection. The regions were chosen arbitrarily based on findings from previous mass-univariate VBM analyses. The classification accuracies using different ROIs for classifying MCI-C and MCI-NC were up to 65%, and the region combining hippocampus and parahippocampal gyrus (11,031 voxels) were superior to other regions. From the results, they also suggested that covariance between information encoded in the ROIs may help classification. The regions selected using SVM-RFE were widely distributed across the brain, which is similar to our results on the LPBA40 atlas. This also suggests that inter-regional covariance or correlation may play an important role as a biomarker for early detection of AD.

As Faria et al. (2012) noted, a typical whole brain MR image has approximately more than hundreds of thousands of voxels, and correlations between these enormous number of voxels exceed 5 billion. Moreover, the signal from each voxel is so noisy that it is practically challenging for us to produce a good feature representation for a classification task from the high-dimensional original data. Thus, future work towards challenges of learning automatically hidden topological structures or deep architectures from the original data such as unsupervised feature learning (Coates et al., 2011) and deep learning (Bengio, 2009) could allow us to identify a good representation of features for classification in various neuroimaging data.

Conclusions

In conclusion, this study showed that the performance of SVM-based classification of MCI using GM volumes from structural MRI at the baseline differs depending on the choice of atlases that defines ROIs. LPBA40, a population-based probabilistic atlas, was superior to AAL, a single-subject atlas, in classification performance using the SVM-RFE procedure. The result suggests that feature selection is crucial to improve classification performance and that the feature selection method based on the SVM-RFE algorithm effectively enhanced the classification accuracy regardless of the choice of atlas. The choice of atlases for feature extraction is also important when using no feature selection. The appropriate selection of ROIs combined with a feature selection technique in a voxel-based approach has the potential of further improving the classification performance. Moreover, atlas-based parcellation methods can be applied to analyses using other modalities such as resting state functional connectivity MRI studies (Wang et al., 2009; Faria et al., 2012) and multi-modal studies combining structural MRI with other modalities. This study will provide implications for future atlas-based analyses using multivariate pattern analysis methods on a wide range of issues and modalities.

Acknowledgments

This work was supported by the Health Labour Sciences Research Grand from the Ministry of Health, Labour, and Welfare of Japan (H17-Tyojyu-023) and the Research Funding for Longevity Sciences from National Center for Geriatrics and Gerontology, Japan (20-1). The authors appreciate the efforts and contributions of those who engaged in the subjects' care and the collection of MRI images and clinical reports.

References

Aguilar C, Westman E, Muehlboeck J-S, Mecocci P, Vellas B, Tsolaki M, et al. Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. Psychiatry Res 2013;212:89–98.

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270–279.

Amunts K, Schleicher A, Zilles K. Cytoarchitecture of the cerebral cortex—more than localization. NeuroImage 2007;37:1061–5.

Ashburner J, Friston KJ. Voxel-based morphometry—the methods. NeuroImage 2000;11:805–21.

Ashburner J, Friston KJ. Unified segmentation. NeuroImage 2005;26:839–51.

Ashburner J. A fast diffeomorphic image registration algorithm. NeuroImage 2007;38:95–113.

Bengio Y. Learning deep architectures for AI. Foundat and Trends Mach Learn 2009;2:1–127.

Braak H, Thal DR, Ghebremedhin E, Del Tredici K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. J Neuropathol Exp Neurol 2011;70:960–9.

Brodmann K. Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Leipzig: Barth; 1909.

Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB. A review of atlas-based segmentation for magnetic resonance brain images. Comput Methods Programs Biomed 2011;104:e158–77.

Chance SA, Casanova MF, Switala AE, Crow TJ, Esiri MM. Minicolumn thinning in temporal lobe association cortex but not primary auditory cortex in normal human ageing. Acta Neuropathol. 2006;111:459–64.

Chance SA, Clover L, Cousijn H, Currah L, Pettingill R, Esiri MM. Microanatomical correlates of cognitive ability and decline: normal ageing, MCI, and Alzheimer's disease. Cereb Cortex 2011;21:1870–8.

Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol 2011;2:27:1–27.

Cho Y, Seong J-K, Jeong Y, Shin SY, Alzheimer's Disease Neuroimaging Initiative. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. NeuroImage 2012;59:2217–30.

Chu C, Hsu A-L, Chou K-H, Bandettini P, Lin C, Alzheimer's Disease Neuroimaging Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. NeuroImage 2012;60:59–70.

Coates A, Lee H, Ng AY. An analysis of single-layer networks in unsupervised feature learning. In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), JMLR W&CP 15, 2011.

Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, et al. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLoS ONE 2011;6:e21896.

Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert M-O, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. NeuroImage 2011;56:766–81.

Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging 2011;32:2322.e19–27.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 2006;31:968–80.

Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 2010;53:1–15.

Devanand DP, Van Heertum RL, Kegeles LS, Liu X, Jin ZH, Pradhaban G, et al. (99m)Tc hexamethyl-propylene-aminoxime single-photon emission computed tomography prediction of conversion from mild cognitive impairment to Alzheimer disease. Am J Geriatr Psychiatry 2010;18:959–72.

Devlin JT, Poldrack RA. In praise of tedious anatomy. NeuroImage 2007;37:1033–41.

Eskildsen SF, Coupé P, García-Lorenzo D, Fonov V, Pruessner JC, Collins DL, et al. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage 2013;65:511–21.

Evans AC, Janke AL, Collins DL, Baillet S. Brain templates and atlases. NeuroImage 2012;62:911–22.

Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements. IEEE Trans Med Imaging 2007;26:93–105.

Faria AV, Joel SE, Zhang Y, Oishi K, van Zjil PCM, Miller MI, et al. Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. NeuroImage 2012;61:613–621.

Ferreira LK, Diniz BS, Forlenza OV, Busatto GF, Zanetti MV. Neurostructural predictors of Alzheimer's disease: A meta-analysis of VBM studies. Neurobiol Aging 2011;32:1733–41.

Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 2010;6:67–77.

Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389–422.

Habeck C, Foster NL, Perneczky R, Kurz A, Alexopoulos P, Koeppe RA, et al. Multivariate and univariate neuroimaging biomarkers of Alzheimer's disease. NeuroImage 2008;40:1503–1515.

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. J Comput Assist Tomogr 1998;22:324–33.

Homma A, Fukuzawa K, Tsukada Y, Ishii T, Hasegawa K, Mohs RC. Development of a Japanese version of Alzheimer's Disease Assessment Scale (ADAS). Jpn J Geriatr Psychiary 1992;3:647–55.

Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification. Tech. rep., Department of Computer Science, National Taiwan University. 2003.

Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging-

Alzheimer‘s Association workgroups on diagnostic guidelines for Alzheimer’s disease. Alzheimers Dement 2011;7:257–262.

Karas GB, Scheltens P, Rombouts SARB, Visser PJ, van Schijndel RA, Fox NC, et al. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. NeuroImage 2004;23:708–16.

Kawashima S, Ito K, Kato T, the SEAD-J Study Group. Inclusion criteria provide heterogeneity in baseline profiles of patients with mild cognitive impairment: comparison of two prospective cohort studies. BMJ Open 2012;2:e000773.

Kazui H, Watamori TS, Honda R, Mori E. [The validation of a Japanese version of the Everyday Memory Checklist]. No To Shinkei 2003;55:317–25.

Klöppel S, Abdulkadir A, Jack CR, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. NeuroImage 2012;61:457–63.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131:681–9.

Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1143–53.

Kurth F, Luders E, Gaser C. VBM8-Toolbox Manual. 2010 (available at http://dbm.neuro.uni-jena.de/vbm8/VBM8-Manual.pdf).

Mangin J-F, Jouvent E, Cachia A. In-vivo measurement of cortical morphology: means and meanings. Curr Opin Neurol 2010;23:359–67.

McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson DW, Hansen LA, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies

(DLB): report of the consortium on DLB international workshop. Neurology 1996;47:1113–24.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer‘s disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease. Neurology 1984;34:939–44.

McKhann GM, Albert MS, Grossman M, Miller B, Dickson D, Trojanowski JQ, et al. Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease. Arch Neurol 2001;58:1803–9.

Misra C, Fan Y, Davatzikos C. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. NeuroImage 2009;44:1415–22.

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–4.

Nelson PT, Braak H, Markesbery WR. Neuropathology and cognitive impairment in Alzheimer disease: a complex but coherent relationship. J Neuropathol Exp Neurol 2009;68:1–14.

Nyunt MSZ, Fones C, Niti M, Ng T-P. Criterion-based validity and reliability of the Geriatric Depression Screening Scale (GDS-15) in a large validation sample of community-living Asian older adults. Aging Ment Health 2009;13:376–82.

Ono M, Kubick S, Abernathey C. Atlas of the cerebral sulci. New York: Thieme; 1990.

Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 2005;352:2379–2388.

Rajapakse JC, Giedd JN, Rapoport JL. Statistical approach to segmentation of single-channel cerebral MR images. IEEE Trans Med Imaging 1997;16:176–86.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.

Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, et al. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. NeuroImage 1999;10:233–60.

Schmahmann JD, Doyon J, Toga AW, Evans AC, Petrides M. MRI atlas of the human cerebellum. San Diego: Academic Press; 2000.

Shapleske J, Rossell SL, Woodruff PW, David AS. The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. Brain Res Rev 1999;29:26–49.

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, et al. Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 2008;39:1064–80.

Stonnington CM, Tan G, Klöppel S, Chu C, Draganski B, Jack CR, et al. Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease. NeuroImage 2008;39:1180–5.

Sullivan K. Estimates of interrater reliability for the Logical Memory subtest of the Wechsler Memory Scale-Revised. J Clin Exp Neuropsychol 1996;18:707–12.

Tabert MH, Manly JJ, Liu X, Pelton GH, Rosenblum S, Jacobs M, et al. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. Arch Gen Psychiatry 2006;63:916–924.

Toga AW, Thompson PM, Mori S, Amunts K, Zilles K. Towards multimodal atlases of the human brain. Nat Rev Neurosci 2006;7:952–66.

Toga AW, Thompson PM. What is where and why it is important. NeuroImage 2007;37:1045–9.

Tohka J, Zijdenbos A, Evans A. Fast and robust parameter estimation for statistical partial volume models in brain MRI. NeuroImage 2004;23:84–97.

Tzourio-Mazoyer N, Hervé PY, Mazoyer B. Neuroanatomy: Tool for functional localization, key to brain organization. NeuroImage 2007;37:1059–60.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. NeuroImage 2002;15:273–89.

Vapnik VN. Statistical learning theory. New York: Wiley; 1998.

Vemuri P, Whitwell JL, Kantarci K, Josephs KA, Parisi JE, Shiung MS, et al. Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. NeuroImage 2008;42:559–567.

Wang J, Wang L, Zang Y, Yang H, Tang H, Gong Q, et al. Parcellation-dependent small-world brain functional networks: a resting-state fMRI study. Hum Brain Mapp 2009;30:1511–1523.

Whitwell JL, Josephs KA, Murray ME, Kantarci K, Przybelski SA, Weigand SD, et al. MRI correlates of neurofibrillary tangle pathology at autopsy: A voxel-based morphometry study. Neurology 2008;71:743–749.

Whitwell JL, Jack CR, Przybelski SA, Parisi JE, Senjem ML, Boeve BF, et al. Temporoparietal atrophy: a marker of AD pathology independent of clinical diagnosis. Neurobiol Aging 2011;32:1531–1541.

Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS ONE 2011;6:e25446.

Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17:37–49.

Zheng ZZ. The functional specialization of the planum temporale. J Neurophysiol 2009;102:3079–81.

Zilles K, Amunts K. Centenary of Brodmann's map—conception and fate. Nat Rev Neurosci 2010;11:139–45.

Figure captions

Fig. 1.

Three brain atlases used for feature extraction, overlaid to representative structural MR images. AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas.

Fig. 2.

Schematic representation of how to obtain a final ranking of features after SVM-RFE. In the figure, m refers to the number of features of a data set.

Fig. 3.

Result of VBM analysis: In the left hippocampus and parahippocampal gyrus, only a significant cluster of gray matter density reduction in subjects with amnestic MCI who converted to AD within three years (MCI-C) compared to subjects who did not (MCI-NC) (p<0.001, uncorrected for multiple comparisons and p<0.05, cluster-level corrected for multiple comparisons).

Fig. 4.

RFE rank score matrices from three brain atlases. The vertical axis of the map represents the subject number, i.e., each step of the leave-one-out cross-validation (LOOCV) procedure. The horizontal axis represents the number of features in each atlas. The top-ranked features having a score of 1, i.e., the feature last selected during the SVM-RFE procedure, are colored in white, while the features that were not selected (score 0) are colored in black. AAL,

Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas.

Fig. 5.

Final rankings of features of three brain atlases and bar plots of feature rank score as a result of the SVM-RFE procedure. Each bar plot for each region is colored differently according to its anatomical location. AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas.

Fig. 6.

Selected region maps from the SVM-RFE procedure, which revealed the highest performance, overlaid to representative structural MR images. The regions with the highest rank are colored in white and those with the lowest rank are colored in black. AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas.

Fig. 7.

Plots of the classification accuracy versus the number of features in the dataset extracted with each brain atlas. AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas.

Fig. 8.

Classification accuracies obtained with features extracted using different atlases (left) and features further selected through the SVM-RFE procedure (right). Values are mean and error bars represent standard errors. * $p<0.0001$, two-way ANOVA followed by Tukey's multiple comparison test. AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40,

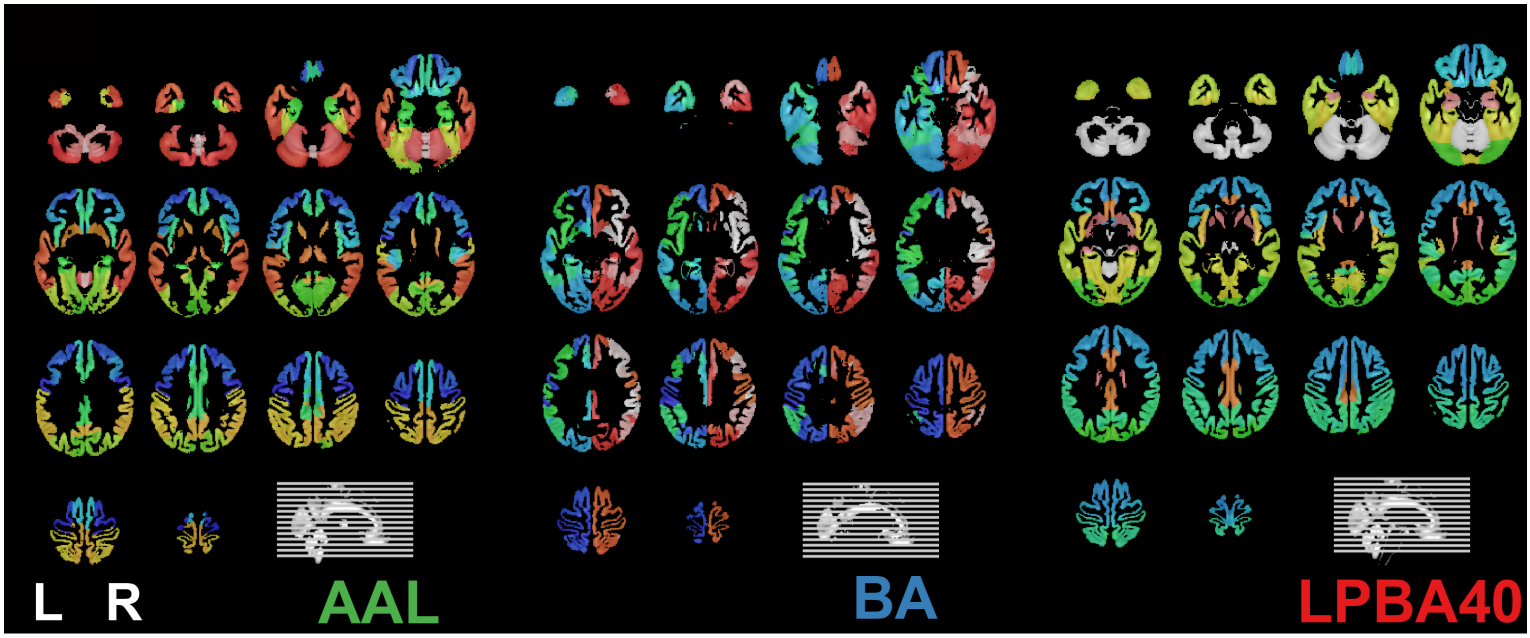
LONI Probabilistic Brain Atlas; SVM-RFE, Support vector machine-recursive feature elimination.

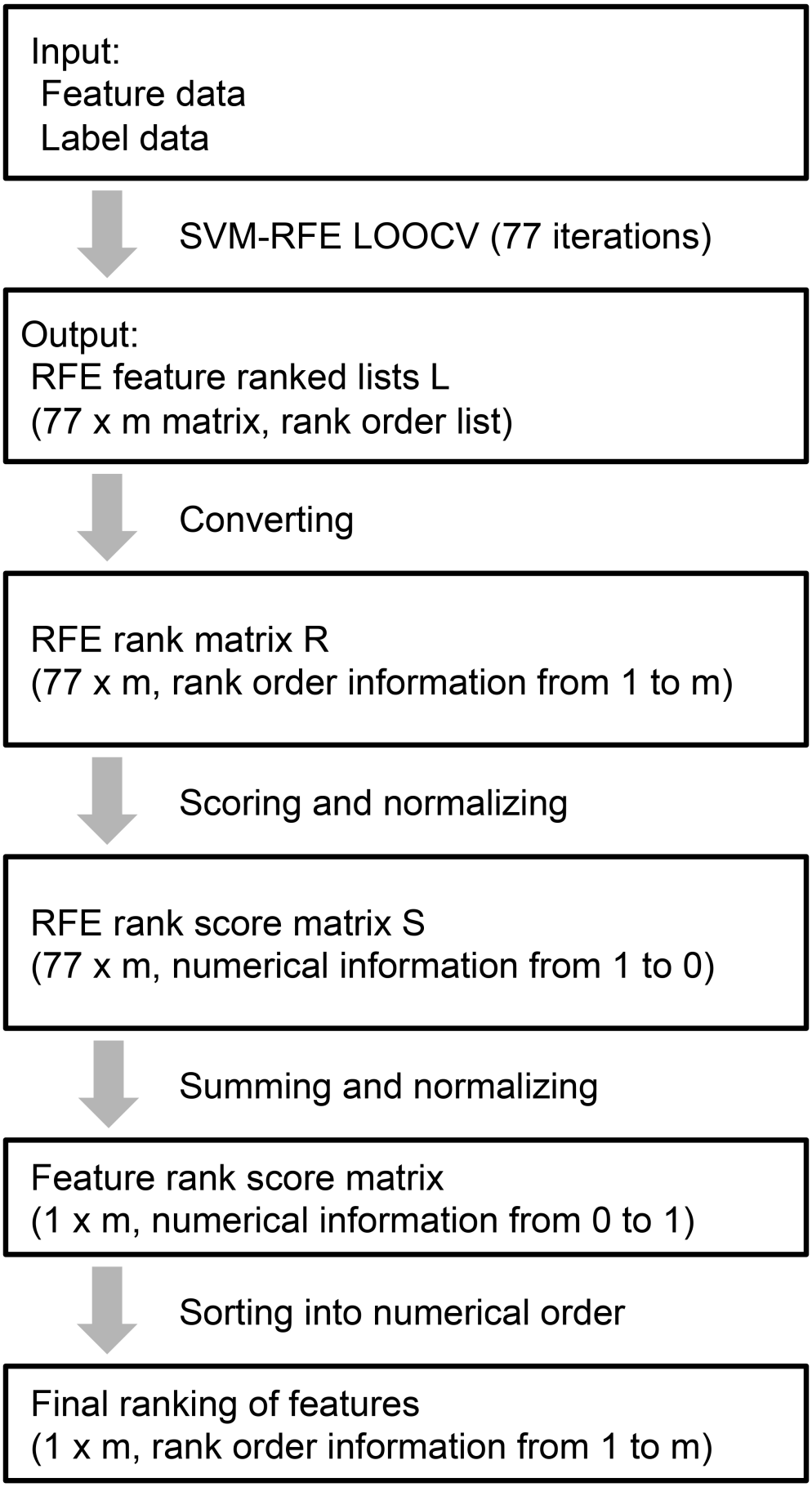
Fig. 9.

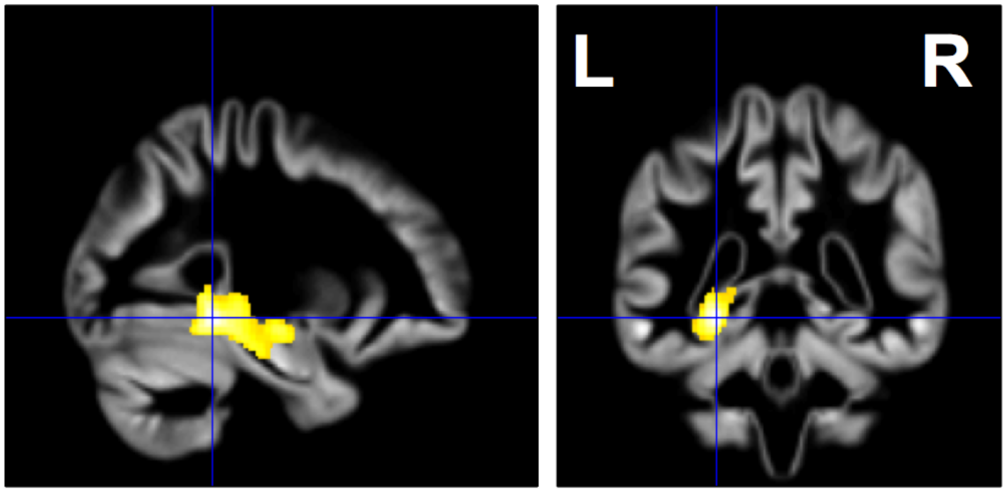
Receiver operating characteristics (ROC) curves with the areas under the curve (AUC) and 95% confidence intervals (CI) obtained with different atlases using the original features and the features further selected through the SVM-RFE procedure. CIs for AUCs were computed with DeLong's method. SVM-RFE, Support vector machine-recursive feature elimination; AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas.

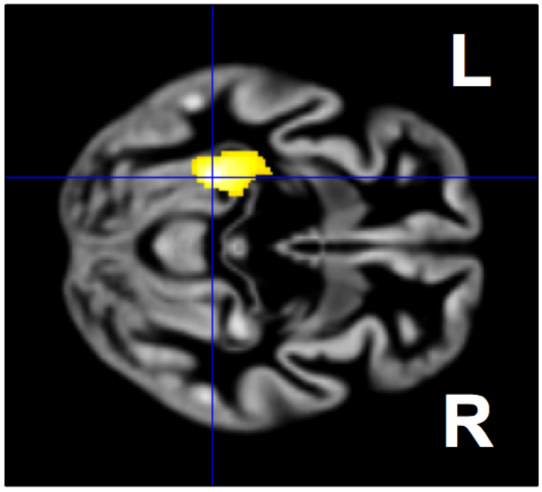
Fig. 10.

Left parahippocampal regions and left hippocampal regions in three brain atlases corresponding to the regions detected by the VBM analysis. The left parahippocampal and the hippocampal regions are colored in white and red, respectively. AAL, Automated Anatomical Labeling; BA, Brodmann's areas; LPBA40, LONI Probabilistic Brain Atlas; L PHG, Left parahippocampal gyrus; L HC, Left hippocampus.


L R
AAL
BA
LPBA40

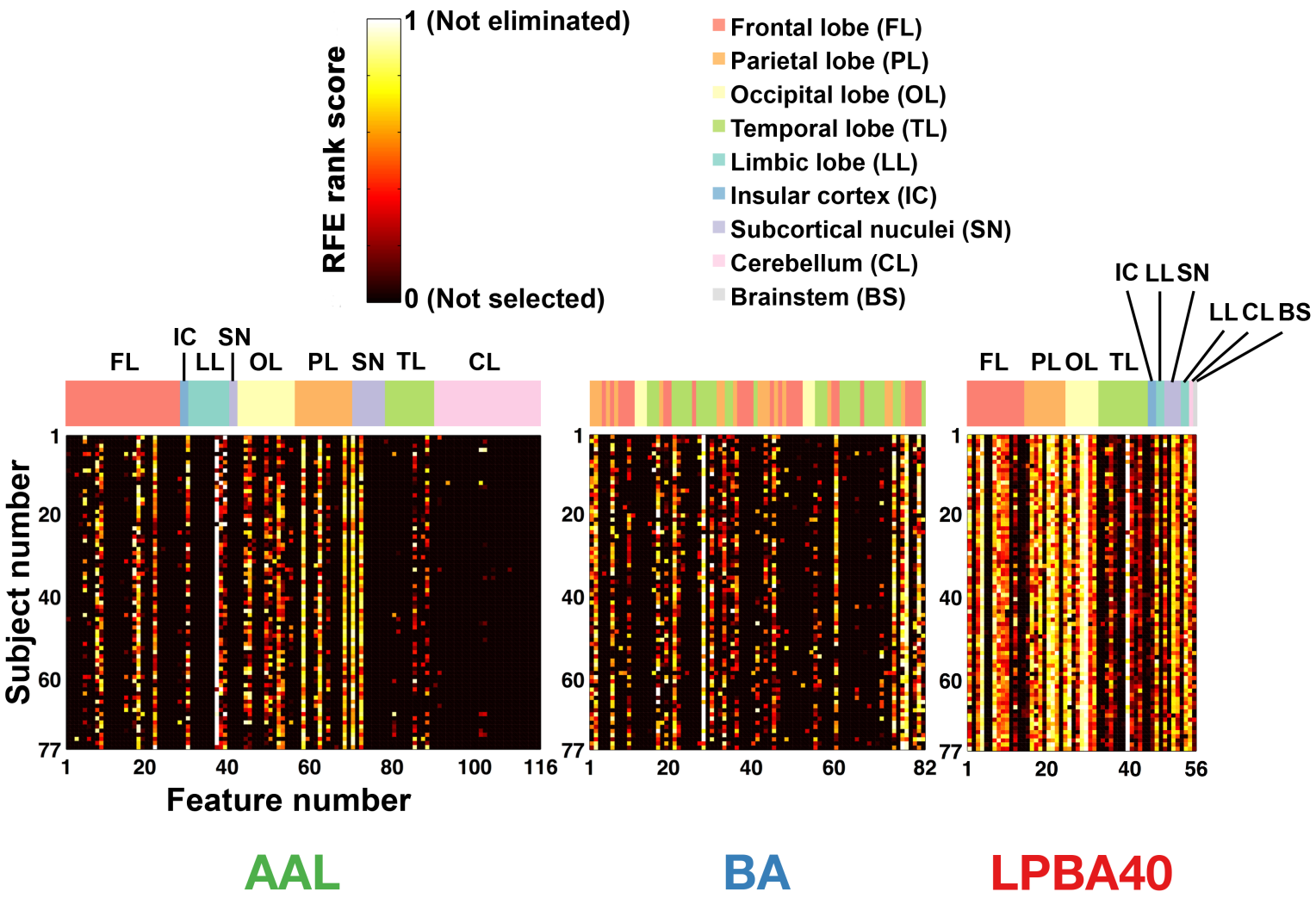

Input:
Feature data
Label data
SVM-RFE LOOCV (77 iterations)
Output:
RFE feature ranked lists L
(77 x m matrix, rank order list)
Converting
RFE rank matrix R
(77 x m, rank order information from 1 to m)
Scoring and normalizing
RFE rank score matrix S
(77 x m, numerical information from 1 to 0)
Summing and normalizing
Feature rank score matrix
(1 x m, numerical information from 0 to 1)
Sorting into numerical order
Final ranking of features
(1 x m, rank order information from 1 to m)

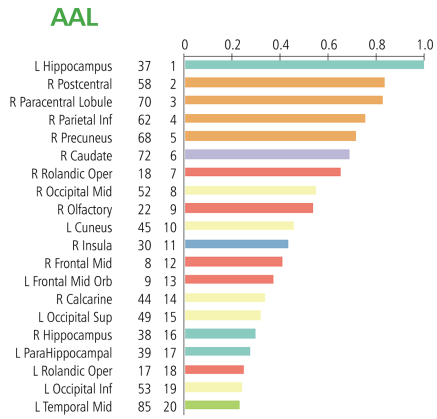


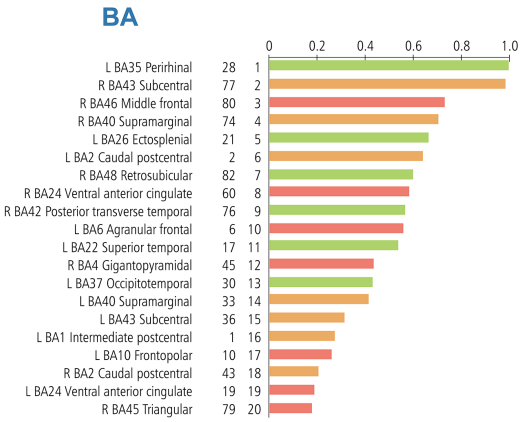


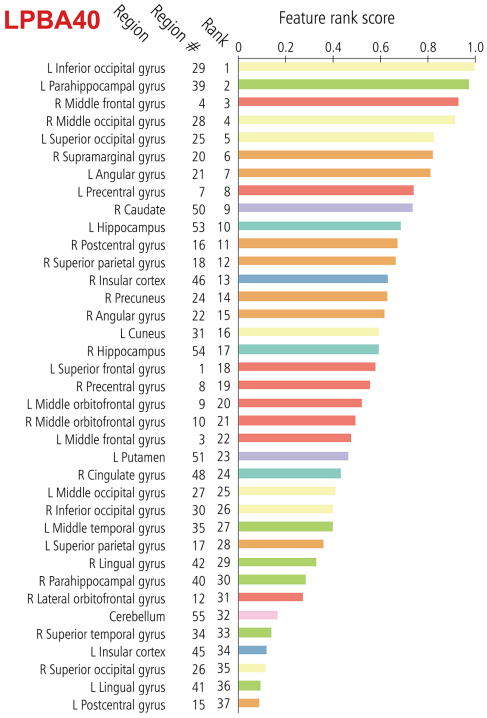
Region	Coordinate MNI space x	y	z	Z score	Cluster size
Left parahippocampal gyrus	−26	−42	−9	4.93	2391*
Left hippocampus	−28	−15	−14	4.22	

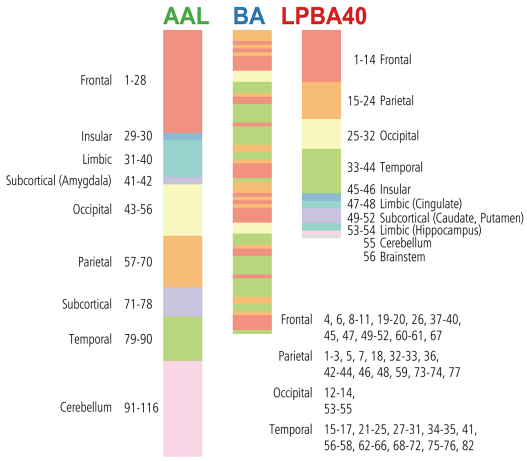
* Uncorrected $p<0.001$,
cluster level $p<0.05$ (corrected for multiple comparisons).


1 (Not eliminated)
RFE rank score
0 (Not selected)
Frontal lobe (FL)
Parietal lobe (PL)
Occipital lobe (OL)
Temporal lobe (TL)
Limbic lobe (LL)
Insular cortex (IC)
Subcortical nuculei (SN)
Cerebellum (CL)
Brainstem (BS)
FL IC LL SN OL PL SN TL CL
FL PL OL TL IC LL SN LL CL BS
Subject number
Feature number
1 20 40 60 77
1 20 40 60 80 100 116
1 20 40 60 82
1 20 40 56
AAL
BA
LPBA40

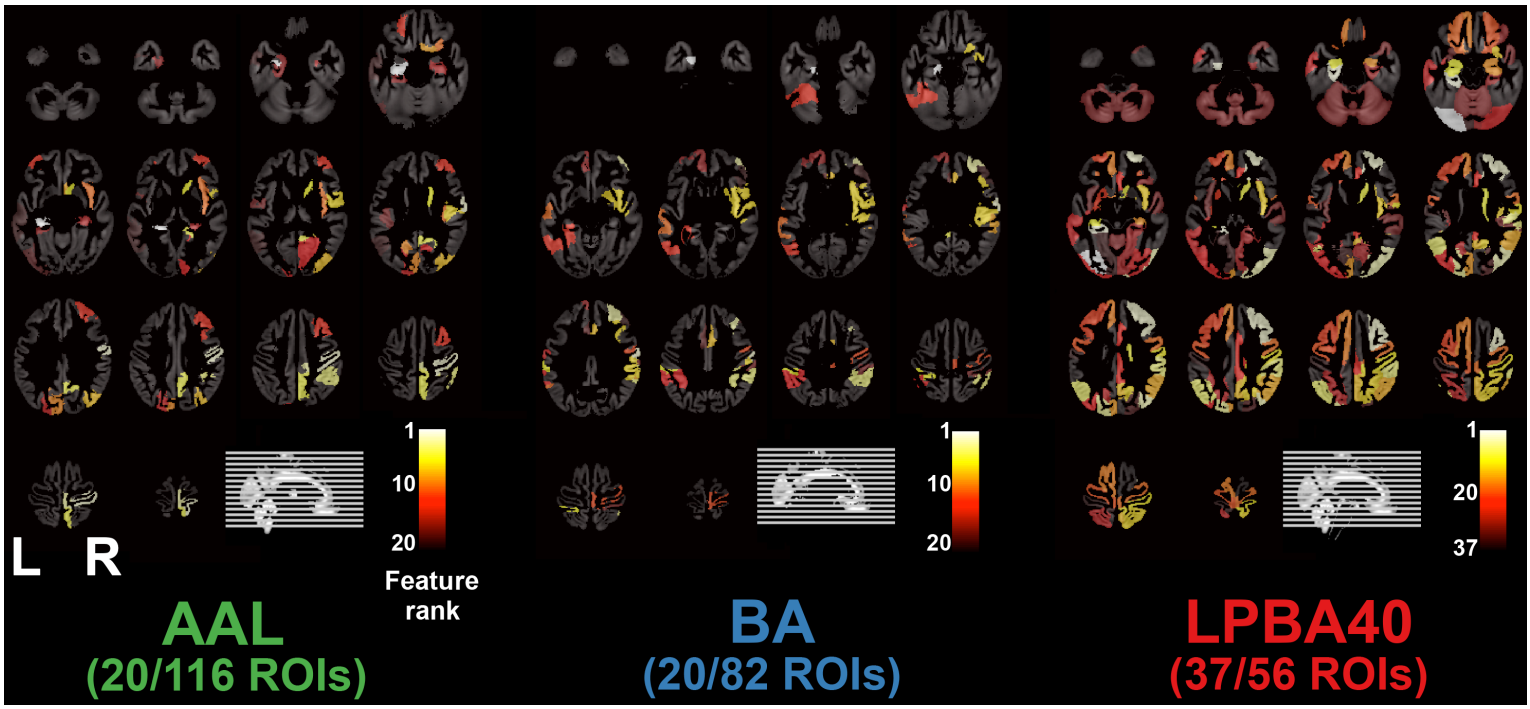


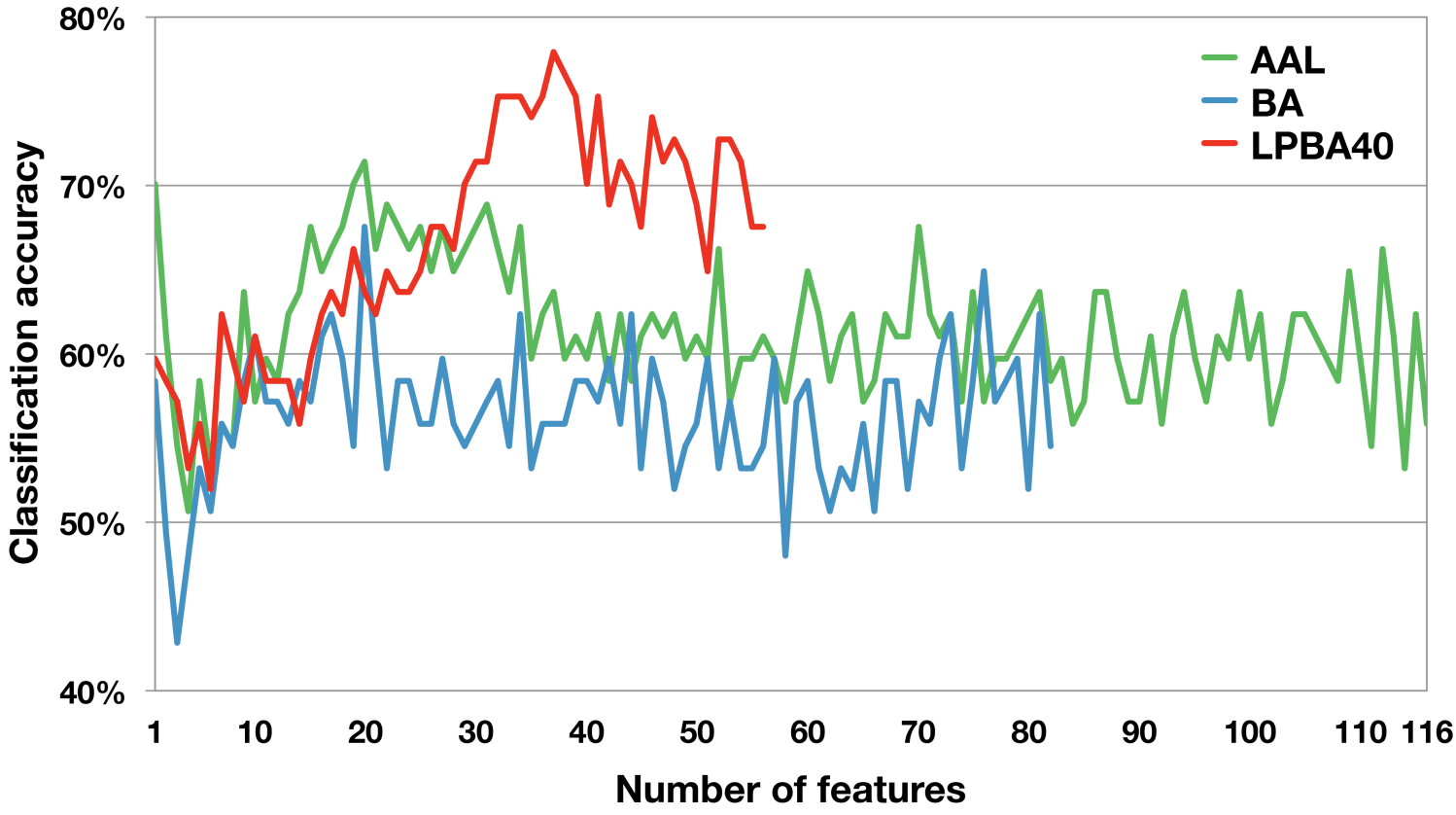


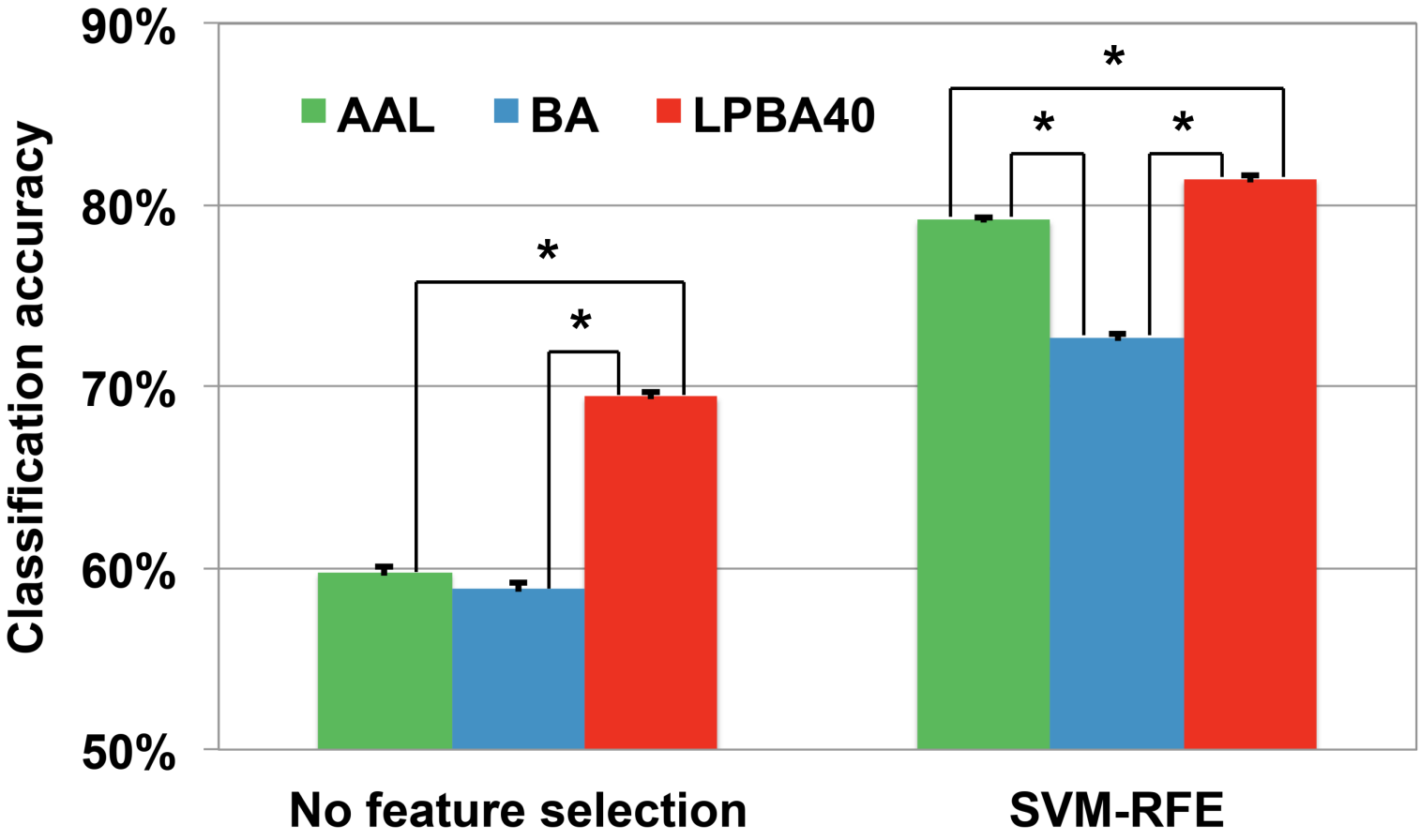


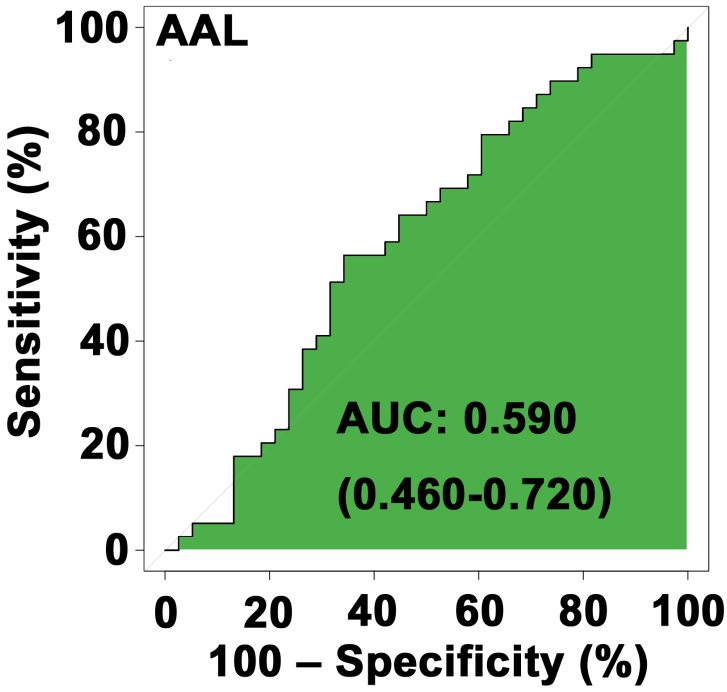


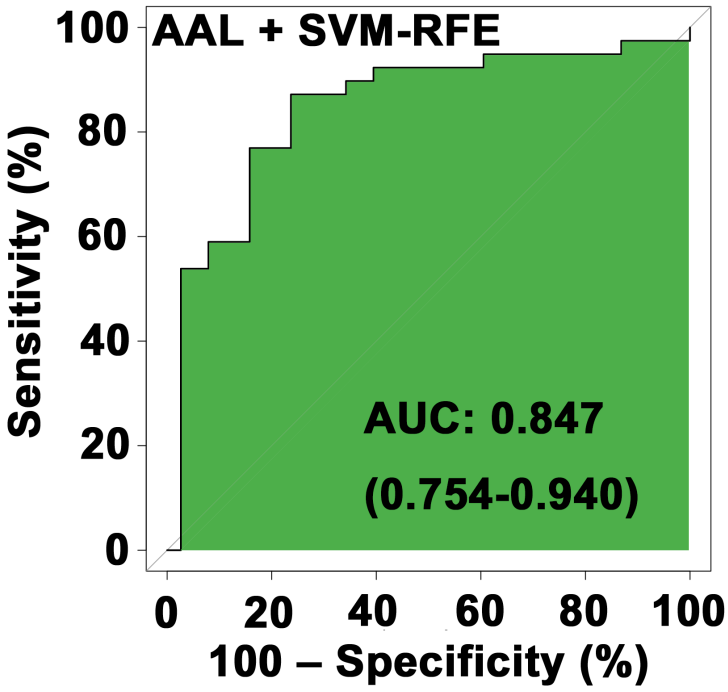
	Number of regions		
	AAL	BA	LPBA40
Frontal lobe	28	26	14
Parietal lobe	14	18	10
Occipital lobe	14	6	12
Temporal lobe	12	32	8
Limbic lobe	10	0	4
Insular cortex	2	0	2
Subcortical nuculei	10	0	4
Cerebellum	26	0	1
Brainstem	0	0	1
Total	116	82	56

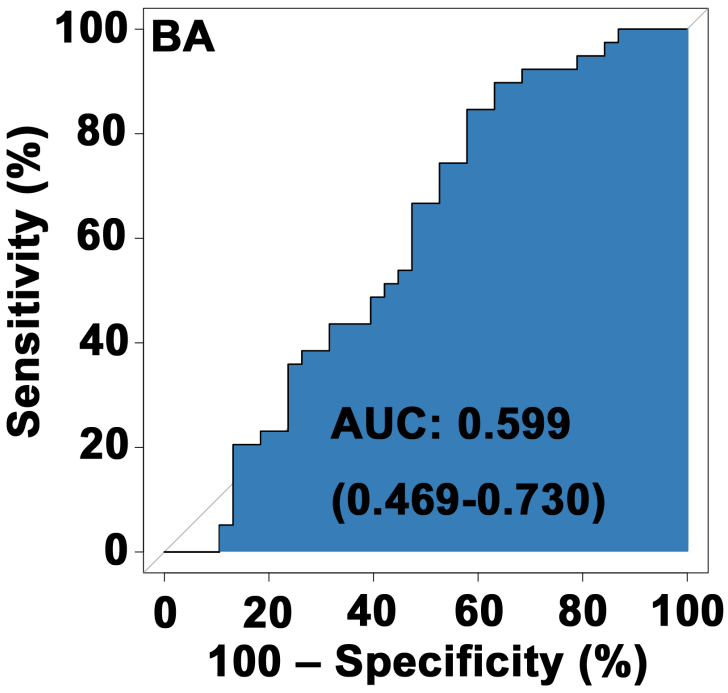

L R
1
10
20
Feature rank
AAL
(20/116 ROIs)
1
10
20
BA
(20/82 ROIs)
1
20
37
LPBA40
(37/56 ROIs)

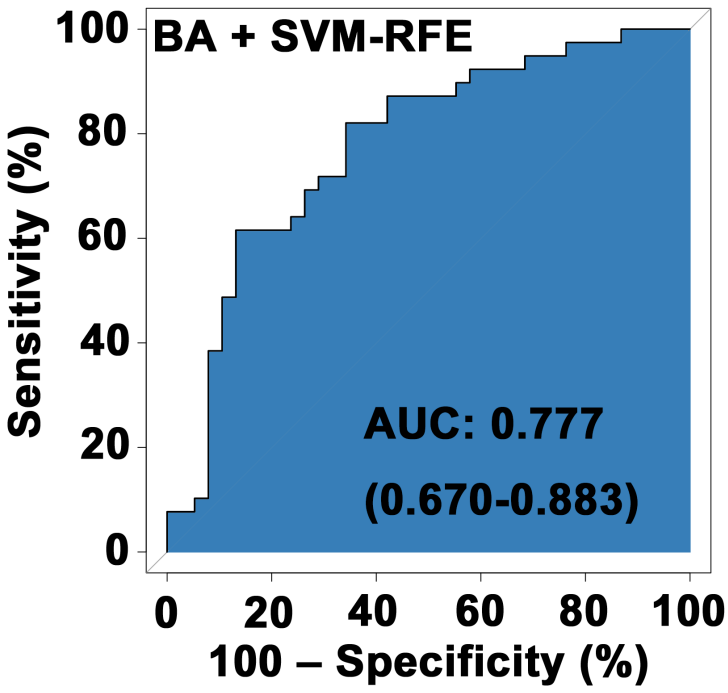

80%
70%
60%
50%
40%
Classification accuracy
1
10
20
30
40
50
60
70
80
90
100
110
116
Number of features
AAL
BA
LPBA40

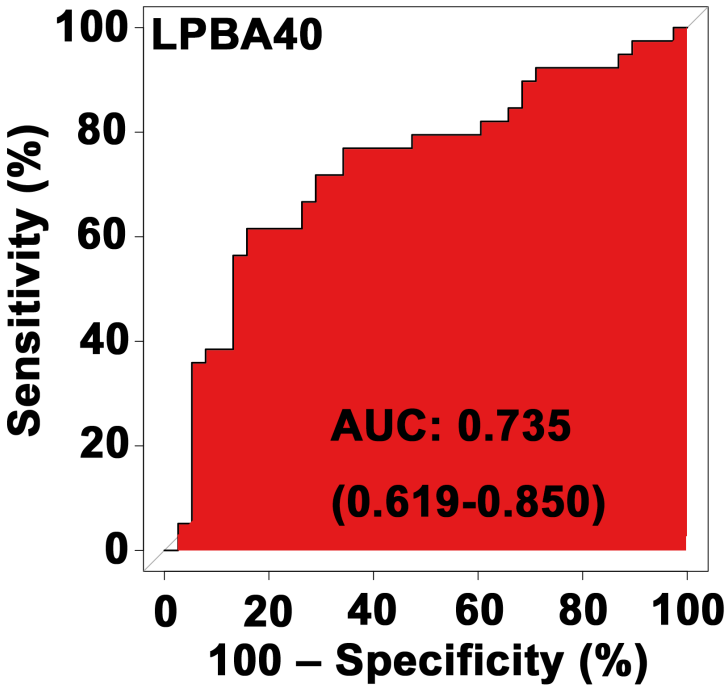

90%
80%
70%
60%
50%
Classification accuracy
AAL
BA
LPBA40
*
*
*
*
*
No feature selection
SVM-RFE

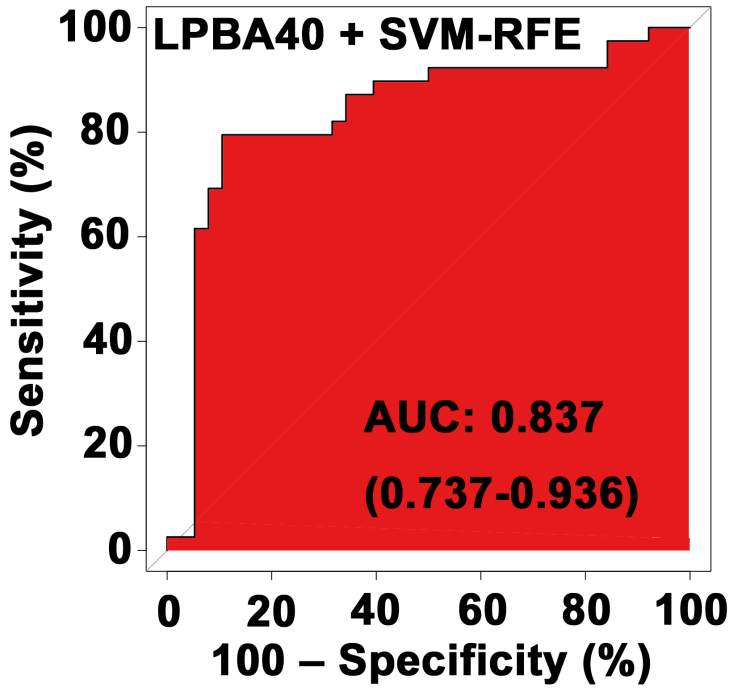

AAL
AUC: 0.590
(0.460-0.720)
Sensitivity (%)
100 – Specificity (%)
0
20
40
60
80
100


AAL + SVM-RFE
AUC: 0.847
(0.754-0.940)
Sensitivity (%)
100 – Specificity (%)
0
20
40
60
80
100


BA
AUC: 0.599
(0.469-0.730)
Sensitivity (%)
100 – Specificity (%)
0
20
40
60
80
100


BA + SVM-RFE
AUC: 0.777
(0.670-0.883)
Sensitivity (%)
100 – Specificity (%)
0
20
40
60
80
100


LPBA40
AUC: 0.735
(0.619-0.850)
Sensitivity (%)
100 – Specificity (%)
0
20
40
60
80
100


LPBA40 + SVM-RFE
AUC: 0.837
(0.737-0.936)
Sensitivity (%)
100 – Specificity (%)
0
20
40
60
80
100

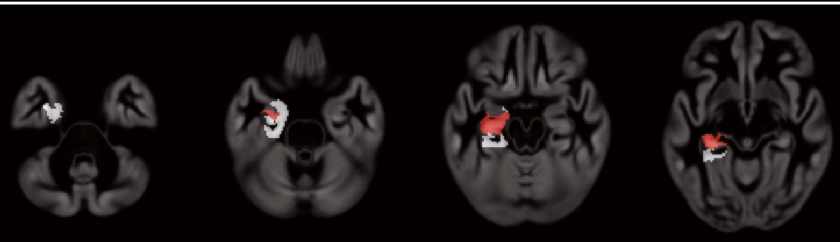
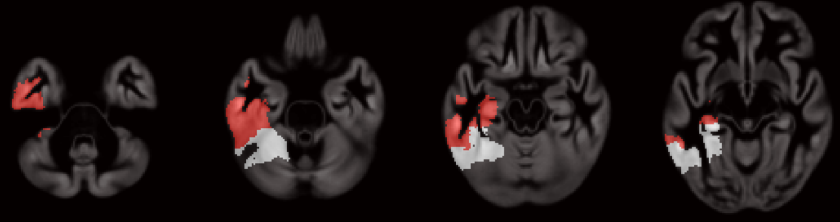
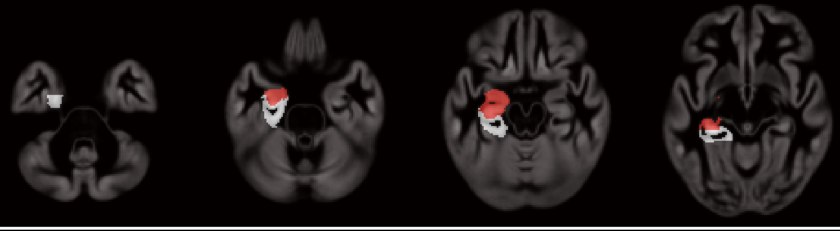
Atlas	Region	Feature #	Feature Rank	Region map
AAL	L PHG	39	17	
	L HC	37	1	
BA	L BA37	30	13	
	L BA20	15	46	
LPBA40	L PHG	39	2	
	L HC	53	10	

Table 1
Demographic and neuropsychological data of MCI-C and MCI-NC subjects at baseline.

	MCI-C (n = 39)	MCI-NC (n = 38)	*p* value
Age [years]	71.3 ± 6.7	70.6 ± 6.9	0.65
Female / Male	20 / 19	22 / 16	0.32
Education [years]	12.2 ± 3.2	11.8 ± 3.1	0.62
WMS-R LM immediate recall	6.5 ± 3.3	9.4 ± 3.1	$<10^{-3}$ *
WMS-R LM delayed recall	1.6 ± 2.2	4.3 ± 2.9	$<10^{-4}$ *
MMSE	25.6 ± 1.8	27.0 ± 2.0	0.003 *
ADAS-J cog	9.9 ± 4.7	7.8 ± 4.6	0.046 *
GDS	4.9 ± 2.3	3.4 ± 1.8	0.004 *

MCI-C, MCI converters; MCI-NC, MCI non-converters; WMS-R LM, Wechsler Memory Scale-Revised Logical memory; MMSE, Mini-Mental State Examination; ADAS-J cog, Alzheimer's Disease Assessment Scale-Cognitive Subscale, Japanese version; GDS, Geriatric Depression Scale.
Age, education, and neuropsychological test scores are shown with mean ± S.D.
* t-test, $p<0.05$.

Table 2
Results of SVM classification using the feature sets extracted with three brain atlases without feature selection and those with a feature selection method based on SVM-RFE. The feature selection technique enhanced the performance of the classification for all of the atlases.

Atlas	Without feature selection				With SVM-RFE			
	Number of features	ACC	SEN	SPC	Number of features	ACC	SEN	SPC
AAL	116	55.8%	56.4%	55.3%	20	71.4%	69.2%	73.7%
BA	82	54.5%	53.8%	55.3%	20	67.5%	64.1%	71.1%
LPBA40	56	67.5%	71.8%	63.2%	37	77.9%	76.9%	78.9%

SVM-RFE, support vector machine-based recursive feature elimination; ACC, accuracy; SEN, sensitivity; SPC, specificity; AAL, Automated Anatomical Labeling; BA, Brodmann's Areas; LPBA40, LONI Probabilistic Brain Atlas.