



## Clinical Neuroscience

## Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease



Kenichi Ota<sup>a,b</sup>, Naoya Oishi<sup>a,c,\*</sup>, Kengo Ito<sup>d</sup>,  
Hidenao Fukuyama<sup>a,b</sup>, the SEAD-J Study Group<sup>1</sup> for the Alzheimer's Disease Neuroimaging Initiative<sup>2</sup>,

<sup>a</sup> Human Brain Research Center, Kyoto University Graduate School of Medicine, 54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan

<sup>b</sup> Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, 54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan

<sup>c</sup> Department of Psychiatry, Kyoto University Graduate School of Medicine, 54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan

<sup>d</sup> Department of Clinical and Experimental Neuroimaging, National Center for Geriatrics and Gerontology, 7-430 Morioka-cho, Obu-shi, Aichi 474-8511, Japan

## HIGHLIGHTS

- We investigate imaging-based prediction of conversion from MCI to AD.
- We compare unimodal and multimodal features from MRI, FDG-PET and their combinations.
- Multimodal features are found to be better than unimodal features to predict AD.
- FDG-PET is found to be better than MRI, particularly when using AAL-based features.
- SVM-RFE can improve the predictive accuracy when using atlas-based imaging features.

## ARTICLE INFO

## Article history:

Received 26 January 2015

Received in revised form 27 July 2015

Accepted 18 August 2015

Available online 28 August 2015

## Keywords:

Alzheimer's disease

Mild cognitive impairment

<sup>18</sup>F-fluorodeoxyglucose positron emission tomography

Magnetic resonance imaging

Support vector machine

Feature selection

## ABSTRACT

**Background:** The choice of biomarkers for early detection of Alzheimer's disease (AD) is important for improving the accuracy of imaging-based prediction of conversion from mild cognitive impairment (MCI) to AD. The primary goal of this study was to assess the effects of imaging modalities and brain atlases on prediction. We also investigated the influence of support vector machine recursive feature elimination (SVM-RFE) on predictive performance.

**Methods:** Eighty individuals with amnesic MCI [40 developed AD within 3 years] underwent structural magnetic resonance imaging (MRI) and <sup>18</sup>F-fluorodeoxyglucose positron emission tomography (FDG-PET) scans at baseline. Using Automated Anatomical Labeling (AAL) and LONI Probabilistic Brain Atlas (LPBA40), we extracted features representing gray matter density and relative cerebral metabolic rate for glucose in each region of interest from the baseline MRI and FDG-PET data, respectively. We used linear SVM ensemble with bagging and computed the area under the receiver operating characteristic curve (AUC) as a measure of classification performance. We performed multiple SVM-RFE to compute feature ranking. We performed analysis of variance on the mean AUCs for eight feature sets.

**Abbreviations:** AAL, Automated Anatomical Labeling; A $\beta$ , amyloid- $\beta$ ; AD, Alzheimer's disease; ADAS-J cog, Alzheimer's Disease Assessment Scale-Cognitive Subscale Japanese version; ADNI, Alzheimer's Disease Neuroimaging Initiative; aMCI, amnesic mild cognitive impairment; ANCOVA, analysis of covariance; ANOVA, analysis of variance; AUC, area under the curve; CMRglc, cerebral metabolic rate for glucose; FDG-PET, <sup>18</sup>F-fluorodeoxyglucose positron emission tomography; GDS, Geriatric Depression Scale; GM, gray matter; LOOCV, leave-one-out cross-validation; LPBA40, LONI Probabilistic Brain Atlas; MCI, mild cognitive impairment; MCI-C, MCI converter; MCI-NC, MCI nonconverter; MMSE, Mini-Mental State Examination; MR, magnetic resonance; MRI, magnetic resonance imaging; MSVM-RFE, multiple support vector machine recursive feature elimination; NFT, neurofibrillary tangle; PVE, partial volume effects; RFE, recursive feature elimination; ROC, receiver operating characteristic; ROI, region of interest; SEAD-J, Studies on Diagnosis of Early Alzheimer's Disease-Japan; SVM, support vector machine; VBM, voxel-based morphometry; WMH, white matter hyperintensity; WMS-R-LM, Wechsler Memory Scale-Revised Logical memory test.

\* Corresponding author at: Human Brain Research Center, Kyoto University Graduate School of Medicine, 54 Shogoin-Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan. Tel.: +81 75 751 3695; fax: +81 75 751 3202.

E-mail address: [noishi@kuhp.kyoto-u.ac.jp](mailto:noishi@kuhp.kyoto-u.ac.jp) (N. Oishi).

<sup>1</sup> 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.

<sup>2</sup> Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

**Results:** The interactions between atlas and modality choices were significant. The main effect of SVM-RFE was significant, but the interactions with the other factors were not significant.

**Comparison with existing method:** Multimodal features were found to be better than unimodal features to predict AD. FDG-PET was found to be better than MRI.

**Conclusions:** Imaging modalities and brain atlases interact with each other and affect prediction. SVM-RFE can improve the predictive accuracy when using atlas-based features.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Alzheimer's disease (AD), which is the main cause of dementia, is a slowly progressive neurodegenerative disorder that leads to declines in memory and other cognitive abilities (Alzheimer's Association, 2013). The revised diagnostic criteria and guidelines for AD (Jack et al., 2011) proposed three stages of AD, including preclinical AD (Sperling et al., 2011), mild cognitive dementia (MCI) due to AD (Albert et al., 2011), and dementia due to AD (McKhann et al., 2011). MCI is a heterogeneous clinical entity that pertains to characteristics between those associated with normal aging and AD, and some individuals with MCI develop AD later (Petersen et al., 1999, 2001).

Among the neuropathological hallmarks of AD, neurofibrillary tangles (NFTs), and senile plaques are considered essential for neuropathological diagnosis of AD (Hyman et al., 2012). NFTs are, at least initially, intraneuronal fibrils primarily composed of hyperphosphorylated tau protein, whereas senile plaques are extracellular deposits of amyloid- $\beta$  (A $\beta$ ) peptides. Progression of these AD neuropathological changes probably begins decades before the onset of cognitive decline (Mufson et al., 2012). Early detection of AD is, therefore, important as a basis for early intervention with disease-modifying drugs (Giacobini and Gold, 2013).

Three imaging biomarkers, as biomarkers to identify brain changes that precede the earliest symptoms, are included in the research criteria for diagnosis of MCI due to AD (Albert et al., 2011). Positron emission tomography (PET) amyloid imaging can measure and visualize A $\beta$  deposition. Hippocampal volume or medial temporal atrophy on magnetic resonance imaging (MRI) and brain glucose hypometabolism on  $^{18}$ F-fluorodeoxyglucose (FDG)-PET imaging are measures reflecting neuronal injury, namely, general damage to neurons and synapses (Jack and Holtzman, 2013). Besides structural MRI and FDG-PET, imaging techniques reflecting neuronal injury include single photon emission tomography (SPECT) perfusion imaging (Ito et al., 2013), diffusion tensor imaging (DTI) (Oishi et al., 2011), functional MRI (fMRI) (Li et al., 2015), and MRI perfusion (Chao et al., 2010). Among these, available data for MRI-related biomarkers except structural MRI are limited and less validated. These neuronal injury markers on MRI or FDG-PET are considered to be less direct or nonspecific evidence of AD, not direct evidence of the presence of A $\beta$  or tau (Albert et al., 2011).

However, these markers are considered to be associated with synaptic loss, which is one of the major neuropathological findings in the brains of individuals with early AD (Scheff et al., 2006). Synaptic loss and neuronal loss are the major pathological substrates of cortical atrophy (Serrano-Pozo et al., 2011) and correlates with cognitive decline (Terry et al., 1991). Longitudinal progression of cognitive decline correlates brain glucose metabolic changes (Shokouhi et al., 2013). Synaptic loss occurs in the limbic regions and the neocortex in individuals with amnesic MCI (aMCI) (Scheff and Price, 2006). Different from the distribution of A $\beta$  deposition, temporospatial accumulation of NFTs originates in the entorhinal cortex constituting the anterior portion of the parahippocampal gyrus and extends through the limbic regions to the neocortex (Braak and Braak, 1991).

Structural MRI and FDG-PET are topographical biomarkers that can help to characterize clinical subtypes with distinct regional patterns of cortical hypometabolism in FDG-PET and of cortical atrophy on structural MRI. FDG-PET has good sensitivity in detection of early brain dysfunction in AD among topographical markers (Dubois et al., 2014). In addition, structural MRI and FDG-PET are less invasive than CSF biomarkers and less expensive than Amyloid PET imaging. On the basis of these foundations, MRI and FDG-PET, particularly their multimodal combination (Price, 2012), can provide valuable biomarkers for early detection of AD. FDG-PET abnormalities are known to precede any cognitive symptoms in individuals who later develop AD (Jack et al., 2010). However, the relative diagnostic abilities of MRI and FDG-PET and of combinations of these different modalities for early disease detection remain controversial (Karow et al., 2010; Mosconi et al., 2006). To achieve scientific evidence of the diagnostic utility of FDG-PET and of MRI in early diagnosis of AD, the Studies on Diagnosis of Early Alzheimer's Disease—Japan (SEAD-J) (Kawashima et al., 2012) was launched in 2005 along with other multicenter clinical trials.

Predictive models based on machine learning algorithms have been widely used for MCI classification (Cuingnet et al., 2011; Young et al., 2013; Zhang et al., 2011). The choice of distinguishing features has an important role in pattern classification (Duda et al., 2001). Atlas-based parcellation using a predefined anatomical brain atlas is a simple feature extraction method with good interpretability and general versatility (Cuingnet et al., 2011; Zhang et al., 2011). Because of the brain atlas concordance problem (Bohland et al., 2009), the use of different brain atlases for parcellation provides different features and can affect the ability to predict conversion from MCI to AD. We have recently reported the importance of the choice of brain atlases for feature extraction in the prediction of conversion by using atlas-based MR biomarkers (Ota et al., 2014). Differences in imaging modalities can also affect predictive performance. However, the effects of imaging modalities and brain atlases for feature extraction on AD prediction have not been well documented.

In addition to feature extraction, feature selection is also important in view of dimension reduction for improving generalization ability and identifying distinguishing features. The effects of feature selection on AD predictive performance remain controversial (Chu et al., 2012; Cuingnet et al., 2011; Kerr et al., 2014). Our previous results (Ota et al., 2014) for MR-based features suggest that support vector machine (SVM)-based recursive feature elimination (RFE) can be important in AD prediction. However, the effects of the use of FDG-PET features or multimodal combination of MRI and FDG-PET features on feature selection have not been clarified.

The primary goal of this study was to assess the effects of imaging modalities and of brain atlases and their interactions on AD prediction. We performed atlas-based feature extraction from MRI and FDG-PET data by using different brain atlases. Using these unimodal or multimodal imaging feature sets, we performed SVM-based classification of MCI and added SVM-RFE feature selection to also evaluate the influence of feature selection on the classification performance.

**Table 1**  
Baseline characteristics of 80 participants with mild cognitive impairment (MCI).

Characteristic	Converters (n = 40)	Nonconverters (n = 40)	p-value	Effect size (r)
<b>Demographics</b>				
Mean age, y ± SD	71.4 ± 6.7	70.5 ± 6.7	0.55	0.068
Male, n (%)	20 (50)	18 (45)	0.82 <sup>a</sup>	0.025 <sup>b</sup>
Education, y ± SD	12.3 ± 3.3	11.8 ± 3.1	0.46	0.084
<b>Neuropsychological scores, mean ± SD</b>				
WMS-R-LM I (immediate recall)	6.6 ± 3.3	9.6 ± 3.1	<10 <sup>-4</sup> ***	0.42
WMS-R-LM II (delayed recall)	1.7 ± 2.2	4.5 ± 2.9	<10 <sup>-5</sup> ***	0.48
MMSE	25.6 ± 1.7	27.1 ± 2.0	0.0009***	0.36
ADAS-J cog	10.0 ± 4.7	7.7 ± 4.5	0.025 <sup>a</sup>	0.25
GDS	4.8 ± 2.3	3.5 ± 1.8	0.006 <sup>**</sup>	0.30

Converters developed Alzheimer's disease within 3 years after inclusion. SD, standard deviation; 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. Student's *t*-test unless otherwise indicated

\*  $p < 0.05$ .

\*\*  $p < 0.01$ .

\*\*\*  $p < 0.001$ .

<sup>a</sup>  $\chi^2$  test.

<sup>b</sup>  $w$ .

## 2. Materials and methods

### 2.1. Participants

We identified 80 individuals with aMCI from a total of 114 participants in the SEAD-J (Ito et al., 2015). Diagnosis of MCI was based on an interview with neurologists as described previously (Ota et al., 2014). "Conversion" was defined as a change in diagnosis from aMCI to AD during a 3-year follow-up period, and 40 participants (50%) converted to AD within 3 years. We excluded 34 participants from the analysis for the following reasons: two participants did not undergo baseline three-dimensional T1-weighted MRI scans, three participants converted to non-AD dementia (vascular dementia, dementia with Lewy bodies, and frontotemporal dementia), 23 participants withdrew from the study within 3 years, and six participants were excluded because of the lack of whole-brain coverage in their baseline T1-weighted MRI scans. Table 1 shows more details on the participants' characteristics at baseline. The converter group (MCI-C) and nonconverter group (MCI-NC) significantly differed in baseline neuropsychological scores, including the Mini-Mental State Examination (Folstein et al., 1975) ( $p < 0.0009$ ,  $r = 0.36$ ), Alzheimer's Disease Assessment Scale-Cognitive Subscale, Japanese version (Homma et al., 1992) ( $p = 0.025$ ,  $r = 0.25$ ), the Wechsler Memory Scale-Revised Logical memory test (WMS-R-LM) (Sullivan, 1996) (immediate recall:  $p < 10^{-4}$ ,  $r = 0.42$ ; delayed recall:  $p < 10^{-5}$ ,  $r = 0.48$ ), and Geriatric Depression Scale (Nyunt et al., 2009; Yesavage et al., 1982) ( $p = 0.006$ ,  $r = 0.30$ ). No significant

differences were observed in age ( $p = 0.55$ ,  $r = 0.068$ ), gender ( $\chi^2$  test,  $p = 0.82$ ,  $w = 0.025$ ), and education ( $p = 0.46$ ,  $r = 0.084$ ). Student's *t*-test was used to compare baseline differences between the two groups unless otherwise stated. We computed *r* for the *t*-test and *w* for the  $\chi^2$  test as measures of effect size (Cohen, 1992).

### 2.2. Magnetic resonance image acquisition and preprocessing

Three-dimensional structural MRI scans at the baseline were collected by using T1-weighted gradient echo sequences on a variety of 1.5-T MRI scanners at eight sites and a 3.0-T MRI scanner at one site. Of 80 participants, 77 (96%) participants were scanned on the 1.5-T scanners and 3 (4%) on the 3-T scanner. Details about MRI acquisition parameters are provided in Supplementary Table 2. We performed the following voxel-based morphometry (VBM) preprocessing procedures using the SPM8 software package (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8>) and VBM8 Toolbox (Kurth et al., 2010) (<http://dbm.neuro.uni-jena.de/vbm>) in MATLAB 7.12. We performed VBM procedures including segmentation, spatial normalization, and smoothing as described previously (Ota et al., 2014).

### 2.3. Fluorodeoxyglucose positron emission tomography image acquisition and preprocessing

FDG-PET scans were collected in a resting state in a dark room 40–60 min after venous injection of <sup>18</sup>F-FDG (Kawashima et al.,

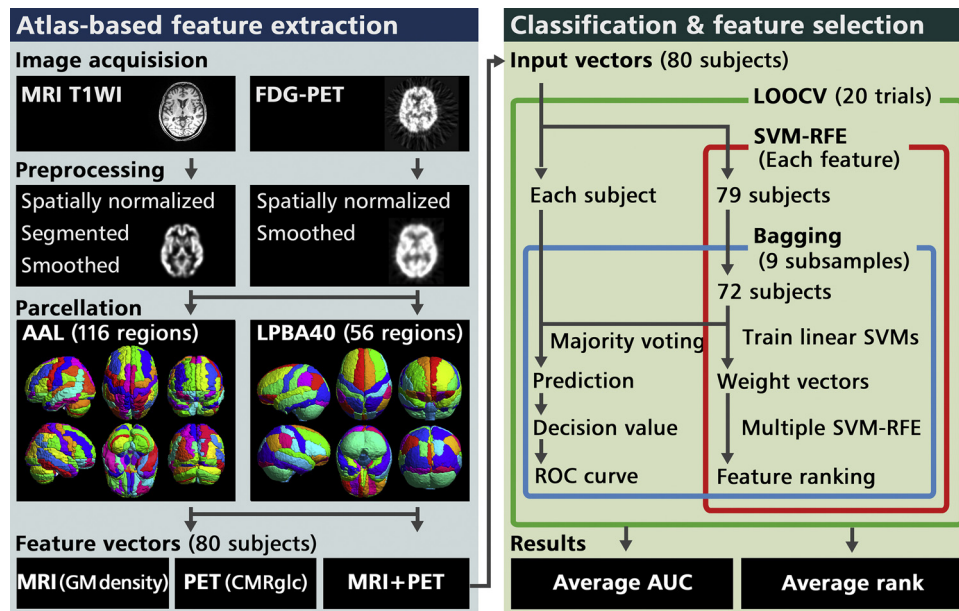
**Table 2**  
Regions of significant brain gray matter loss on magnetic resonance imaging and significant brain glucose hypometabolism on fluorodeoxyglucose positron emission tomography in converters relative to those in nonconverters.

Region	MNI coordinate			Peak-level T	Cluster size
	x	y	z		
<b>MRI</b>					
Left parahippocampal gyrus	-27	-44	-3	4.65	1146 <sup>*</sup>
Left hippocampus	-24	-34	-6	4.04	
<b>FDG-PET</b>					
Left precuneus	-6	-63	27	4.87	4310 <sup>**</sup>
Right precuneus	8	-63	33	4.37	
Left inferior temporal gyrus	-56	-31	-23	4.62	1310 <sup>*</sup>
Left middle temporal gyrus	-57	-40	-11	4.08	
Left angular gyrus	-46	-57	30	4.40	1163 <sup>*</sup>

MNI, Montreal Neurological Institute; MRI, magnetic resonance imaging; FDG-PET, fluorodeoxyglucose positron emission tomography.

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

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



**Fig. 1.** Flow diagram representing steps for feature extraction, classification, and feature selection processes. MRI T1WI, magnetic resonance imaging T1-weighted image; FDG-PET, fluorodeoxyglucose positron emission tomography; AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas; GM, Gray matter; CMRglc, cerebral metabolic rate of glucose; LOOCV, leave-one-out cross-validation; SVM, Support vector machine; SVM-RFE, Support vector machine recursive feature elimination; ROC, Receiver operating characteristic; MRI, magnetic resonance imaging; PET, positron emission tomography; AUC, Area under the ROC curve.

2012). The whole brain was used as the reference region for intensity normalization to compare the effects of two different brain atlases on the ability of prediction because of its higher signal-to-noise ratio compared with intensity normalization to the cerebellum (Dukart et al., 2010). The average whole brain uptake of the converters was  $10,143 \pm 10,289$ , that of nonconverters,  $8451 \pm 10,912$ , ( $t$ -test,  $p=0.48$ , Cohen's  $d=0.11$ ). The images were normalized by using an in-house FDG-PET template created from FDG-PET and MRI brain scans of 23 normal elderly individuals. Partial volume effects (PVE) were not corrected according to a previous finding that PVE correction did not affect the detection of hippocampal hypometabolism in aMCI when using the global mean for scaling (Mevel et al., 2007). The resulting images were smoothed by using an 8-mm isotropic Gaussian filter to blur individual variations in gyral anatomy and to increase the signal-to-noise ratio (Mevel et al., 2007).

#### 2.4. Voxel-based comparison

To determine the differences in MRI and FDG-PET data between converters and nonconverters, we performed voxel-based comparison of gray matter (GM) maps and of FDG-PET images between converters and nonconverters by using the SPM8 software package. We used analysis of covariance (ANCOVA) models with age, gender, and scan sites as covariates according to a previous report (Ota et al., 2014). The thresholds for the statistical parametric maps were set to  $p=0.001$ , uncorrected for multiple comparisons at a voxel level, and  $p=0.05$ , family-wise error corrected at a cluster level. Clusters were visualized by using the BrainNet Viewer 1.43 (<http://www.nitrc.org/projects/bnv/>) (Xia et al., 2013).

#### 2.5. Feature extraction

To assess the effects of brain atlases for template-based region-of-interest (ROI) analysis on the performance of prediction, we compared two brain atlases for simplicity: the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) and the LONI Probabilistic Brain Atlas (LPBA40) (Shattuck et al., 2008) (Fig. 1).

The AAL atlas has 116 ROIs, including 90 cerebral regions and 26 cerebellar regions, whereas the LPBA40 has 56 ROIs, including 54 cerebral regions, the brainstem, and the cerebellum. Using these two brain atlases, we extracted feature vectors from the preprocessed baseline MRI, which was modulated with the Jacobian, and from the FDG-PET data, as described previously (Ota et al., 2014). Accordingly, we obtained eight different feature sets, including two MR-based features (MRI-AAL and MRI-LPBA40) representing GM density in each ROI, two PET-based features (PET-AAL and PET-LPBA40) representing relative cerebral metabolic rate for glucose in each ROI, and four multimodal features (MRI-AAL + PET-AAL, MRI-AAL + PET-LPBA40, MRI-LPBA40 + PET-AAL, and MRI-LPBA40 + PET-LPBA40).

#### 2.6. Classification

Fig. 1 is a schematic overview of our classification pipeline. SVMs (Vapnik, 1998) are one of the most popular supervised learning models and have been applied to the prediction of conversion from MCI to AD (Chu et al., 2012; Ota et al., 2014). To focus on comparison across imaging modalities and brain atlases for feature extraction, we used linear SVMs with the regularization parameter  $C$  of 1 for computational simplicity in the present study as described in a prior study (Zhang et al., 2011).

To enable statistical analysis of the results, we used a bootstrap aggregating (bagging) method (Breiman, 1996) within each round of LOOCV (Dosenbach et al., 2010). In each LOOCV loop, the original data set having 80 participants was divided into a test set of one participant and a training set comprising 79 participants. Then a bootstrap sample of size 72 (90% of the total number of participants) was obtained by bootstrap resampling of the original training set. In general, a smaller bootstrap sample size provides a better prediction performance and a greater variance. A sample size of 72 was determined arbitrarily with the aim of achieving a smaller variance. This procedure was repeated nine times to obtain a bagged ensemble of nine linear SVMs. The number of subsamples was determined empirically in view of a trade-off between computation time and classification performance. These multiple



linear SVMs were trained on different subsamples of training data to compute the weight vectors and the feature ranking scores, as mentioned below. The final prediction of the bagged ensemble for a given test set was determined by majority voting (Valentini et al., 2004). The final decision value of the bagged ensemble for a test vector was computed by averaging the decision values for the SVMs that voted for the majority class. After each round of LOOCV, we obtained decision values for all participants with regard to a given number of features.

On the basis of these decision values, we generated the receiver operating characteristic (ROC) curve (Metz, 1978) and computed the area under the ROC curve (AUC) as a measure of the performance of a classifier (Fawcett, 2004; Huang and Ling, 2005) by using the pROC package for R (Robin et al., 2011). Other measures, such as accuracy, sensitivity, and specificity with regard to a given number of features, were also computed. We implemented the above classification algorithm by using the e1071 package for R (Dimitriadou et al., 2005) based on the LIBSVM library (Chang et al., 2011).

For the eight different feature sets, we repeated the above LOOCV procedure 20 times and computed mean AUC values for AD prediction across 20 LOOCV tests. To compare the ability of prediction across different feature sets, we formed averaged ROC curves with 95% confidence intervals for each feature set on the basis of the results of 20 LOOCV tests by using Fawcett's vertical averaging algorithm (Fawcett, 2004).

### 2.7. Feature selection

In our previous study (Ota et al., 2014), we used the SVM-RFE feature selection method (Guyon et al., 2002) to improve the ability of prediction and to identify the regions with high discriminating power. In the present study, we newly applied the multiple SVM-RFE (MSVM-RFE) feature selection method (Duan et al., 2005) to compute the feature ranking score for a given feature from weight vectors of multiple linear SVMs trained on bootstrap subsamples of the original training data.

As described above, we had nine linear SVMs trained on different subsamples of the original training data. Let  $w_j$  denote the weight vector of the  $j$ th linear SVM and  $w_{ji}$  be the corresponding weight value associated with the  $i$ th feature. The ranking criterion for the  $i$ th feature  $v_{ji}$  was given as  $v_{ji} = (w_{ji})^2$ . As a result of the majority voting, we selected the ranking criteria  $v_k$  for a feature set  $S_k$ , where  $k$  represents the indices of the linear SVMs that voted for the majority class. The feature ranking scores  $c_k$  for a feature set  $S_k$  was computed as:

$$c_k = \frac{\bar{v}_k}{\sigma_k}$$

where,  $\bar{v}_k$  is the mean of  $v_k$  and  $\sigma_k$  is the standard deviation of  $v_k$ .

On the basis of these feature-ranking scores, the feature with the smallest ranking score was removed, and the feature set was updated at the end of each SVM-RFE loop. After performing 20 LOOCV tests, the final average rank was determined as a mean of 20 average ranks for the respective LOOCV tests. Our implementation of the MSVM-RFE algorithm in R was adapted from [http://www.uccor.edu.ar/paginas/seminarios/Software/SVM\\_RFE\\_R\\_implementation.pdf](http://www.uccor.edu.ar/paginas/seminarios/Software/SVM_RFE_R_implementation.pdf). Selected regions were visualized by using the BrainNet Viewer Version 1.43 (<http://www.nitrc.org/projects/bnvv/>) (Xia et al., 2013).

### 2.8. Statistical analysis

To test the differences between the means of AUC values for eight feature sets, we used  $AUC_{RFE-}$  and  $AUC_{RFE+}$  for each feature set.  $AUC_{RFE-}$  for a feature set was an AUC value that was obtained with

the original features (with no feature selection), and  $AUC_{RFE+}$  was determined as the highest AUC during the SVM-RFE procedure for the feature set. To assess the main effects of different factors (SVM-RFE feature selection, brain atlases for parcellation, and imaging modalities) and interactions among them on the performance of prediction, we initially performed an overall three-way factorial analysis of variance (ANOVA) that included all the factors. The  $2 \times 4 \times 3$  ANOVA design included factor "RFE" with levels "RFE-" and "RFE+," factor "atlas" with levels, "MRI-AAL + PET-AAL," "MRI-AAL + PET-LPBA40," "MRI-LPBA40 + PET-AAL," and "MRI-LPBA40 + PET-LPBA40," and factor "modality" with levels "MRI," "PET," and "MRI + PET." Then two-way ANOVAs (atlas  $\times$  modality) at each level of factor RFE were performed. To examine the effects of atlas and modality in unimodal features, we performed an additional two-way (atlas  $\times$  modality) ANOVA at each level of RFE for four unimodal feature sets. Furthermore, to assess the effects of atlas combinations in multimodal feature sets, we performed a two-way (MRI atlas  $\times$  PET atlas) ANOVA for four multimodal feature sets. We performed post-hoc analysis of interactions by using tests of simple main effects (one-way ANOVA) or Tukey's test, where appropriate. Post-hoc multiple comparisons were performed by using Tukey's test. We performed Type II analysis by using the car package for R (Fox and Weisberg, 2009) to compute sums of squares for ANOVA (Langsrud, 2003). To estimate effect sizes, we computed  $\eta^2$  as the ratio of sum of squares for an effect to the total sum of squares (Cohen, 1973). Values of  $p < 0.05$  were considered to indicate statistical significance. The R Statistical Computing Environment, version 3.0.2 (R Development Core Team, 2013) was used for classification, feature selection, and all statistical analyses except voxel-based comparisons.

## 3. Results

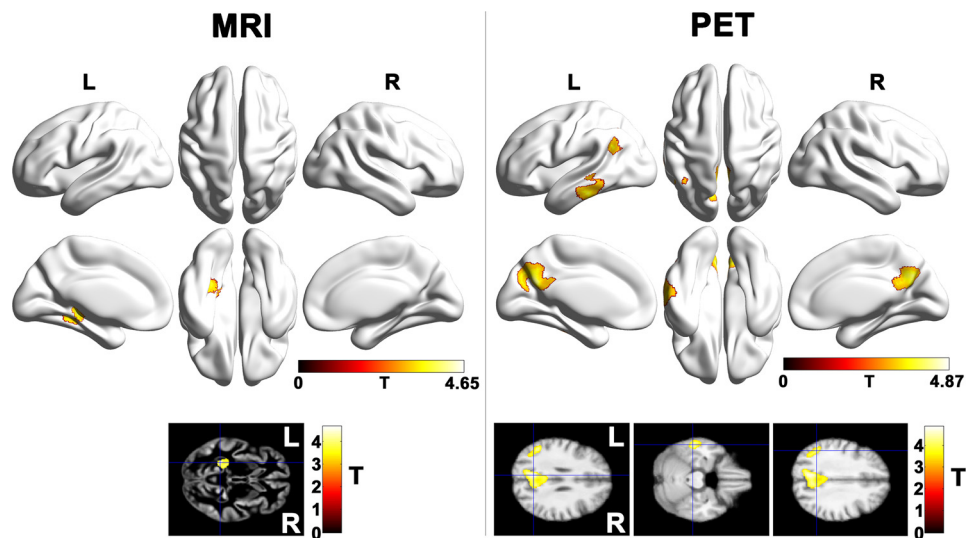
### 3.1. Voxel-based comparison of gray matter volume and cerebral metabolic rate of glucose

As shown in Table 2, we found one significant cluster of GM reduction in the MCI-C group compared with MCI-NC on MRI and three significant clusters of cerebral glucose hypometabolism on FDG-PET. Fig. 2 shows the locations of the clusters mapped onto the surface of the brain. GM loss in the MCI-C group was found in the left parahippocampal gyrus and the left hippocampus. Brain glucose hypometabolism in the MCI-C group was found mostly in the left temporoparietal association cortex, including the precuneus, inferior temporal gyrus, middle temporal gyrus, and angular gyrus. Reverse contrast (MCI-NC < MCI-C) showed no significant regions on MRI or FDG-PET.

### 3.2. Effects of imaging modality, brain atlas, and support vector machine recursive feature elimination on prediction

Fig. 3 shows the plots of the mean AUC values of different feature sets versus the number of features. The arrows on the top of each plot indicate the number of features that provided the highest AUC ( $AUC_{RFE+}$ ) on each curve in the same color. In general, particularly in the multimodal feature sets, the AUC at the right end of each plot (no feature selection) tended to be the highest, whereas the plots of LPBA40-based features, particularly in the unimodal feature sets, showed a local maximum where the number of features was <25. Tests for the statistical significance of the difference between mean AUCs are described below.

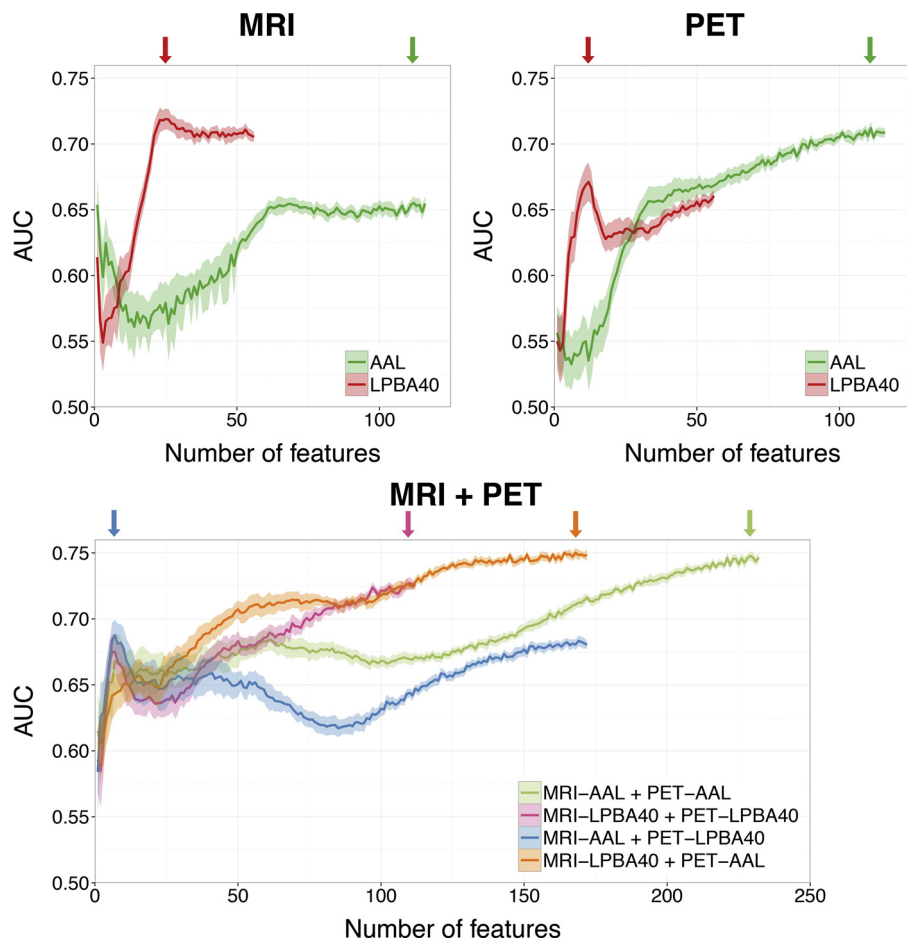
Fig. 4 shows the ROC curves of eight different feature sets for MCI classification. For the MRI-based feature sets, the ROC curve of LPBA40 (shown in red) showed higher sensitivity than that of AAL (shown in green) at almost all specificities. For FDG-PET, in



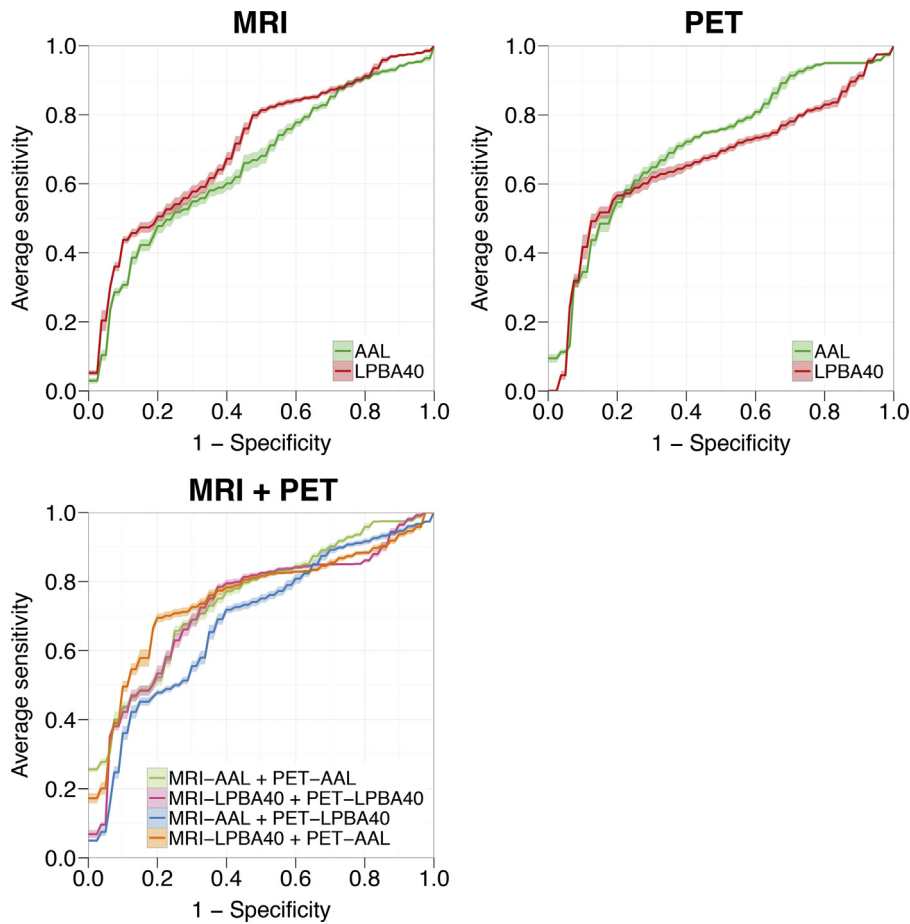
**Fig. 2.** Clusters of significant brain gray matter loss on magnetic resonance imaging (MRI) and significant brain glucose hypometabolism on fluorodeoxyglucose positron emission tomography (FDG-PET) in converters relative to those in nonconverters.

contrast, the ROC curve of AAL (shown in green) demonstrated higher sensitivity than that of LPBA40 (shown in red) at the lower specificity range [false-positive rate (FPR) range of 0.25–0.9]. In the multimodal feature sets, MRI-AAL + PET-AAL (shown in green)

demonstrated higher sensitivity than any of the other feature sets in the FPR range of values  $>0.65$ . MRI-LPBA40 + PET-AAL (shown in orange) showed higher sensitivity than did the other combinations, including unimodal features in the FPR range of values  $<0.2$ .



**Fig. 3.** Classification performance of patients with mild cognitive impairment (MCI) by using support vector machine feature elimination (SVM-RFE) feature selection. Plotted are the mean areas under the receiver operating characteristic (ROC) curves (AUC) resulting from leave-one-out cross-validation tests for MCI classification using magnetic resonance imaging (MRI)-based, positron emission tomography (PET)-based, and multimodal feature sets with respect to the number of features. The shaded region of each plot indicates 95% confidence intervals. The arrows on the top of each plot represent the number of features that provided the highest AUC for each plot in the same color.



**Fig. 4.** Receiver operating characteristic (ROC) curves of eight original feature sets with no feature selection for classification of patients with mild cognitive impairment (MCI). Each curve was formed by vertically averaging 20 ROC curves that resulted from leave-one-out cross-validation tests. The shaded region of each curve indicates 95% confidence intervals.

Table 3 shows the mean AUC and number of features for each feature set for use in our three-way ANOVA. Fig. 5A shows a bar plot of the mean  $AUC_{RFE-}$  and  $AUC_{RFE+}$  values. As shown in Table 4, there were significant main effects for three factors: RFE [ $F(1, 456) = 20.1, p < 10^{-5}, \eta^2 = 0.0065$ ]; atlas [ $F(3, 456) = 333.0, p < 10^{-15},$

$\eta^2 = 0.32$ ]; and modality [ $F(2, 456) = 466.3, p < 10^{-15}, \eta^2 = 0.30$ ]. We found no significant three-way interactions [ $F(6, 456) = 1.5, p = 0.19, \eta^2 = 0.0028$ ]. The two-way interactions between RFE and atlas [ $F(3, 456) = 1.3, p = 0.26, \eta^2 = 0.0013$ ] and between RFE and modality [ $F(2, 456) = 1.1, p = 0.33, \eta^2 = 0.0007$ ] were not significant,

**Table 3**

The mean area under the curve and number of features for each feature set for a three-way ANOVA design.

	R: RFE			
	R1: RFE-		R2: RFE+	
A: Atlas (combination)	Number of features	Mean $AUC_{RFE-}$ (95% CI)	Number of features	Mean $AUC_{RFE+}$ (95% CI)
A1: MRI-AAL + PET-AAL				
M: Modality				
M1: MRI	116	0.655 (0.649–0.660)	112	0.655 (0.650–0.660)
M2: PET	116	0.709 (0.705–0.711)	111	0.712 (0.707–0.717)
M3: MRI + PET	232	0.747 (0.743–0.750)	229	0.748 (0.744–0.751)
A2: MRI-AAL + PET-LPBA40				
MRI	116	0.655 (0.649–0.660)	112	0.655 (0.650–0.660)
PET	56	0.661 (0.656–0.665)	12	0.671 (0.657–0.686)
MRI + PET	172	0.680 (0.678–0.683)	7	0.688 (0.676–0.699)
A3: MRI-LPBA40 + PET-AAL				
MRI	56	0.705 (0.702–0.709)	25	0.719 (0.711–0.727)
PET	116	0.709 (0.705–0.712)	111	0.712 (0.707–0.717)
MRI + PET	172	0.748 (0.744–0.753)	168	0.750 (0.746–0.754)
A4: MRI-LPBA40 + PET-LPBA40				
MRI	56	0.705 (0.702–0.709)	25	0.719 (0.711–0.727)
PET	56	0.661 (0.656–0.665)	12	0.671 (0.657–0.686)
MRI + PET	112	0.725 (0.722–0.728)	110	0.727 (0.722–0.732)

AUC, area under the curve; ANOVA, analysis of variance; RFE, recursive feature elimination; CI, confidence interval; SD, standard deviation; MRI, magnetic resonance imaging; PET, positron emission tomography; AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas.

**Table 4**  
Three-way ANOVA summary table.

Source	(df) <i>F</i>	<i>p</i>	$\eta^2$
RFE	(1, 456) = 20.1	<10 <sup>-5</sup>	0.0065
Atlas	(3, 456) = 333.0	<10 <sup>-15</sup>	0.32
Modality	(2, 456) = 466.3	<10 <sup>-15</sup>	0.30
RFE × Atlas	(3, 456) = 1.3	0.26	0.0013
RFE × modality	(2, 456) = 1.1	0.33	0.0007
Atlas × modality	(6, 456) = 111.7	<10 <sup>-15</sup>	0.22
RFE × atlas × modality	(6, 456) = 1.5	0.19	0.0028
Atlas at modality			
Atlas at MRI	(3, 456) = 227.7	<10 <sup>-15</sup>	0.22
Atlas at PET	(3, 456) = 137.4	<10 <sup>-15</sup>	0.13
Atlas at MRI + PET	(3, 456) = 191.2	<10 <sup>-15</sup>	0.19
Modality at atlas			
Modality at MRI-AAL + PET-AAL	(2, 456) = 449.3	<10 <sup>-15</sup>	0.29
Modality at MRI-AAL + PET-LPBA40	(2, 456) = 44.9	<10 <sup>-15</sup>	0.029
Modality at MRI-LPBA40 + PET-AAL	(2, 456) = 100.5	<10 <sup>-15</sup>	0.065
Modality at MRI-LPBA40 + PET-LPBA40	(2, 456) = 206.6	<10 <sup>-15</sup>	0.13

ANOVA, analysis of variance; RFE, recursive feature elimination; MRI, magnetic resonance imaging; PET, positron emission tomography; AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas; df, degrees of freedom.

\*\*\*  $p < 0.001$ .

whereas the two-way interaction between atlas and modality [ $F(6, 456) = 111.7, p < 10^{-15}, \eta^2 = 0.22$ ] was significant.

Our post-hoc analysis of the atlas × modality interaction demonstrated significant simple main effects of atlas at different levels of modality [ $F(3, 456) = 227.7, p < 10^{-15}, \eta^2 = 0.22$  at M1 (MRI);  $F(3, 456) = 137.4, p < 10^{-15}, \eta^2 = 0.13$  at M2 (PET); and  $F(3, 456) = 191.2, p < 10^{-15}, \eta^2 = 0.19$  at M3 (MRI + PET)] and significant simple main effects of modality at different levels of atlas [ $F(2, 456) = 449.3, p < 10^{-15}, \eta^2 = 0.29$  at A1 (MRI-AAL + PET-AAL);  $F(2, 456) = 44.9, p < 10^{-15}, \eta^2 = 0.029$  at A2 (MRI-AAL + PET-LPBA40);  $F(2, 456) = 100.5, p < 10^{-15}, \eta^2 = 0.065$  at A3 (MRI-LPBA40 + PET-AAL); and  $F(2, 456) = 206.6, p < 10^{-15}, \eta^2 = 0.13$  at A4 (MRI-LPBA40 + PET-LPBA40)].

Regarding atlas combinations, post-hoc multiple comparisons for factor atlas revealed significant differences between any of the four atlas combinations ( $p < 10^{-7}$ ) except between A1 (MRI-AAL + PET-AAL) and A4 (MRI-LPBA40 + PET-LPBA40) ( $p = 0.40$ ). Consequently, marginal means of four groups of atlas combination were ranked in the following order: A3 (MRI-LPBA40 + PET-AAL, 0.724) > A1 (MRI-AAL + PET-AAL, 0.704) and A4 (MRI-LPBA40 + PET-LPBA40, 0.701) > A2 (MRI-AAL + PET-LPBA40, 0.668). Multiple comparisons for atlas at each modality revealed significant differences between any of the different groups ( $p < 10^{-7}$ ) except between A1 (MRI-AAL + PET-AAL) and A3 (MRI-LPBA40 + PET-AAL) ( $p = 0.84$ ). Accordingly, we observed the following rankings for atlas at each modality: at M1 (MRI), LPBA40 > AAL; at M2 (PET), AAL > LPBA40; and at M3 (MRI + PET), A3 (MRI-LPBA40 + PET-AAL) and A1 (MRI-AAL + PET-AAL) > B4 (MRI-LPBA40 + PET-LPBA40) > A2 (MRI-AAL + PET-LPBA40).

Post-hoc multiple comparisons for modality also demonstrated significant differences between MRI and PET, between MRI and MRI + PET, and between PET and MRI + PET, as shown in Fig. 5B. Accordingly, marginal means of three levels of modality were ranked in the following order: MRI + PET (0.727) > PET (0.688) > MRI (0.684). Multimodal feature sets were superior to unimodal feature sets, and PET-based feature sets were better than MRI-based feature sets. Multiple comparisons for modality at each level of factor atlas revealed significant differences between any of the groups ( $p < 0.001$ , except  $p = 0.02$  for between MRI-AAL and PET-LPBA40 at A2) except between MRI-LPBA40 and PET-AAL at A3 ( $p = 0.74$ ). Consequently, the following rankings for modality at each level of atlas were obtained: at A1, M3 (MRI-AAL + PET-AAL) > M2 (PET-AAL) > M1 (MRI-AAL); at

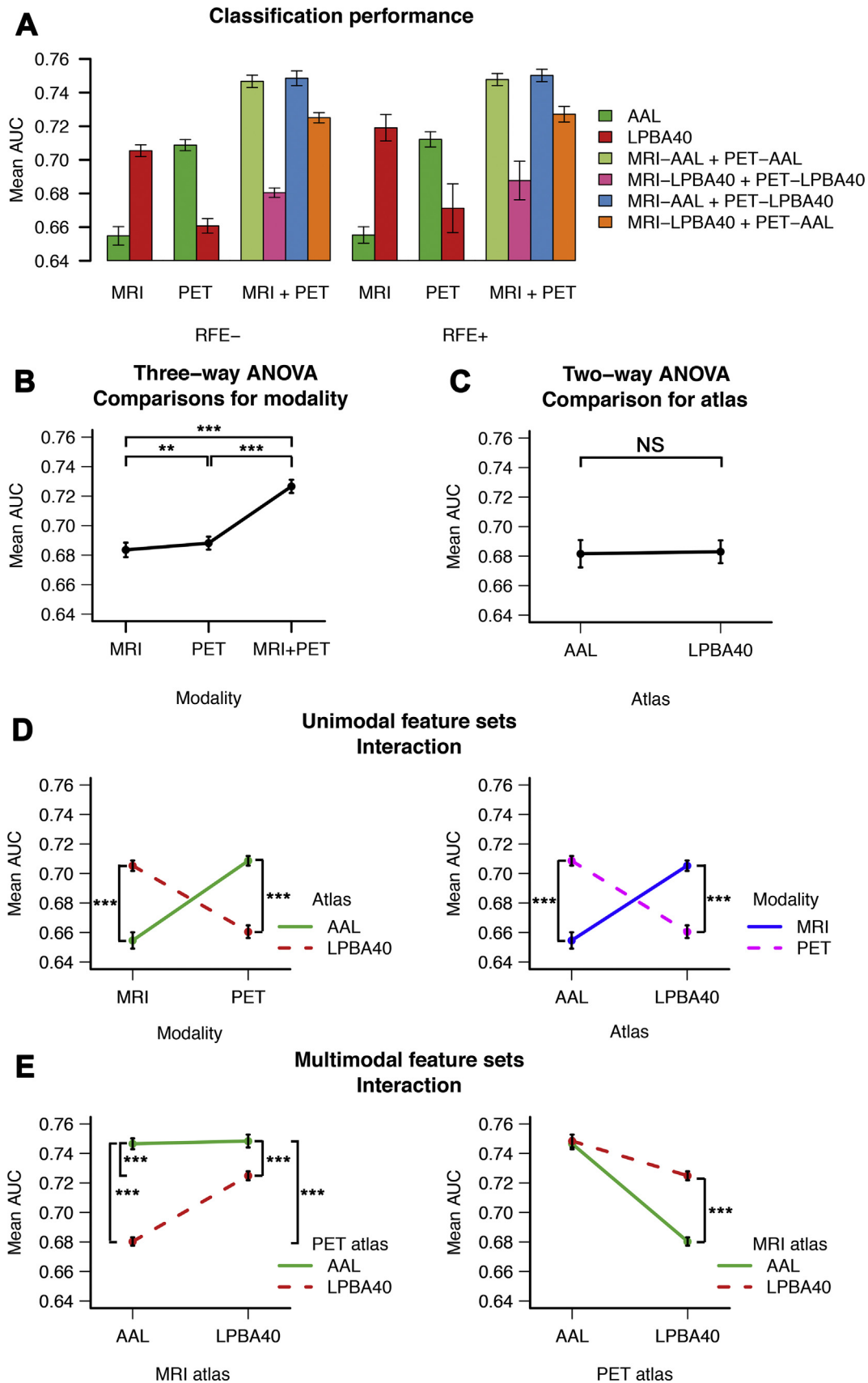
A2, M3 (MRI-AAL + PET-LPBA40) > M2 (PET-LPBA40) > M1 (MRI-AAL); at A3, M3 (MRI-LPBA40 + PET-AAL) > M2 (PET-AAL) and M1 (MRI-LPBA40); and at A4, M3 (MRI-LPBA40 + PET-LPBA40) > M1 (MRI-LPBA40) > M2 (PET-LPBA40). In summary, multimodal combinations were superior to single unimodal feature sets at all levels of atlas. With regard to unimodal features, PET was better than MRI when using AAL, whereas MRI was better than PET when using LPBA40.

Because the lack of significant three-way (RFE × atlas × modality) interactions indicated that the atlas × modality interaction was not significantly different depending on the levels of RFE (RFE- and RFE+), we focused on the atlas × modality interaction at RFE for simplicity. Our additional two-way (atlas × modality) ANOVA for comparing the effects of atlas and modality in unimodal feature sets at RFE- demonstrated a significant main effect of modality [ $F(1, 76) = 5.2, p = 0.02, \eta^2 = 0.008$ ], whereas we found no significant main effect of atlas as shown in Fig. 5C. In contrast, the interaction between atlas and modality was significant (Fig. 5D). Our post-hoc analysis of the atlas × modality interaction revealed the significant simple main effects of modality at AAL and at LPBA40, and of atlas at MRI and at PET.

Another two-way (MRI atlas × PET atlas) ANOVA for comparing the effects of atlases in multimodal feature sets at RFE- showed significant main effects of MRI atlas [ $F(1, 76) = 191.6, p < 10^{-15}, \eta^2 = 0.17$ ] and PET atlas [ $F(1, 76) = 716.0, p < 10^{-15}, \eta^2 = 0.62$ ], and as shown in Fig. 5E, we also found significant interactions between these factors. Our post-hoc analysis of the interaction demonstrated significant simple main effects of MRI atlas at LPBA40 for PET, of PET atlas at AAL for MRI, and of PET atlas at LPBA40 for MRI. We found no significant simple main effect of MRI atlas at AAL for PET.

According to our analysis, multimodal features generally yielded better AUC values than did unimodal features. Comparisons of mean AUC<sub>RFE-</sub> values for multimodal feature sets indicated the following ranking: MRI-LPBA40 + PET-AAL (0.748) and MRI-AAL + PET-AAL (0.747) > MRI-LPBA40 + PET-LPBA40 (0.725) > MRI-AAL + PET-LPBA40 (0.680) (Tukey's test,  $p < 10^{-7}$  except for  $p = 0.997$  for a pair between MRI-LPBA40 + PET-AAL and MRI-AAL + PET-AAL). Regarding MRI-based unimodal features, MRI-LPBA40 (0.705) was significantly better than MRI-AAL (0.655) ( $p < 10^{-7}$ ). In contrast, PET-AAL (0.709) was significantly superior to PET-LPBA40 (0.661) ( $p < 10^{-7}$ ). Comparisons between the unimodal data of MRI and





**Fig. 5.** Statistical analyses. The error bars are 95% confidence intervals. (A) Classification performance. Mean  $AUC_{RFE-}$  and  $AUC_{RFE+}$  values. (B) Comparisons for modality in our three-way ANOVA. Post-hoc multiple comparisons revealed significant differences in the means of RFE- and RFE+ between MRI and PET ( $p = 0.009$ ), between MRI and MRI + PET ( $p < 10^{-7}$ ), and between PET and MRI + PET ( $p < 10^{-7}$ ). (C) Comparison of atlas type in our two-way (atlas  $\times$  modality) ANOVA for unimodal feature sets. The main effect of atlas was not significant [ $F(1, 76) = 0.43, p = 0.51, \eta^2 = 0.0006$ ]. (D) Interaction plots of the two-way ANOVA for unimodal feature sets. Mean  $AUC_{RFE-}$  values for different modalities and for different atlases in unimodal feature sets tested by the additional two-way (atlas  $\times$  modality) ANOVA. The interaction between atlas and modality

**Table 5**

Brain regions that ranked in the top 10% of features for eight feature sets according to the multiple support vector machine feature elimination.

Rank	MRI		PET		MRI + PET							
	MRI	MRI	PET	PET	MRI	PET	MRI	PET	MRI	PET	MRI	PET
	AAL	LPBA40	AAL	LPBA40	AAL	AAL	LPBA40	LPBA40	AAL	LPBA40	LPBA40	AAL
Region	Region	Region	Region	Region	Region	Region	Region	Region	Region	Region	Region	Region
1	L HIP	L PHIP	L T3	L T2	–	R F1	–	R F3	–	L T2	L PHIP	–
2	R F2	L PRE	R HIP	L O1	L HIP	–	L PHIP	–	–	R F3	R IN	–
3	R P2	R F2	R PRE	R F3	–	L Q	–	L O1	L HIP	–	–	R PRE
4	R F3O	R MOFG	L T1P	L PUT	–	R CBLC1	L O3	–	–	R PQ	–	R F1
5	L PRE	L O3	R CBL9	L O2	R IN	–	R O2	–	–	L O1	–	L Q
6	L Q	R O2	VER6	L T1	R HIP	–	–	L T2	–	R POST	–	R CBL9
7	R RO	–	–	–	–	L T3	–	L PUT	R PCL	–	L HIP	–
8	R POST	–	R AMYG	–	L P2	–	R MOFG	–	–	L SMG	–	L T3
9	R PCIN	–	L T2P	–	–	R PRE	R IN	–	L RO	–	L O3	–
10	L O1	–	R P2	–	–	R PAL	L F2	–	–	R LOFG	–	L PUT
11	L SMA	–	L T1	–	–	L AMYG	R CIN	–	L O1	–	–	L O1
12	L F2	–	L CBLC1	–	L RO	–	–	R PQ	L F2	–	–	R CBLC1
13	–	–	–	–	R PCL	–	–	–	R O2	–	–	R PAL
14	–	–	–	–	R O3	–	–	–	R HIP	–	–	R FUS1
15	–	–	–	–	–	R POST	–	–	L Q	–	–	L CBLC1
16	–	–	–	–	–	L CBLC1	–	–	L P2	–	–	R O2
17	–	–	–	–	–	VER7	–	–	R SMA	–	–	R CBL8
18	–	–	–	–	L PHIP	–	–	–	R O1	–	–	L O1
19	–	–	–	–	–	L T2P	–	–	–	–	–	–
20	–	–	–	–	R SMA	–	–	–	–	–	–	–
21	–	–	–	–	–	R CBL9	–	–	–	–	–	–
22	–	–	–	–	–	L F1MO	–	–	–	–	–	–
23	–	–	–	–	–	L T1P	–	–	–	–	–	–
24	–	–	–	–	–	R OC	–	–	–	–	–	–

SVM-RFE, support vector machine-based recursive feature elimination; AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas. The labels for brain regions are defined in Table 6.

FDG-PET revealed no significant differences between MRI-AAL and PET-LPBA40 ( $p=0.32$ ) or between MRI-LPBA40 and PET-AAL ( $p=0.91$ ). No significant differences were observed between any of the other pairs ( $p < 10^{-6}$ ).

### 3.3. Important regions determined by support vector machine recursive feature elimination feature selection

Table 5 lists the top 10% ranked regions determined by the SVM-RFE feature selection method. The label abbreviations are listed in Table 6. Regarding MRI-based features, the left hippocampus ranked first in AAL, and the left parahippocampal gyrus ranked first in LPBA40. For the PET-based features, the left inferior temporal gyrus ranked first in AAL followed by the right hippocampus and right precentral gyrus, whereas the middle temporal gyrus ranked first in LPBA40 followed by the left superior occipital gyrus and the right inferior frontal gyrus. When combining MRI and PET feature vectors, the PET-based regions generally ranked higher than did the MRI-based regions except for the combination of MRI-LPBA40 and PET-AAL.

As shown in Fig. 6, we mapped the regions listed in Table 5 to the surface of the brain. Compared with MRI-based features (upper left) in which the left medial temporal regions in both atlases ranked highest, PET-based features (upper right) selected the lateral temporal regions, not the medial temporal regions. When combining

MRI and PET data (bottom), the right superior frontal gyrus in the AAL ranked high among the PET-based regions.

## 4. Discussion

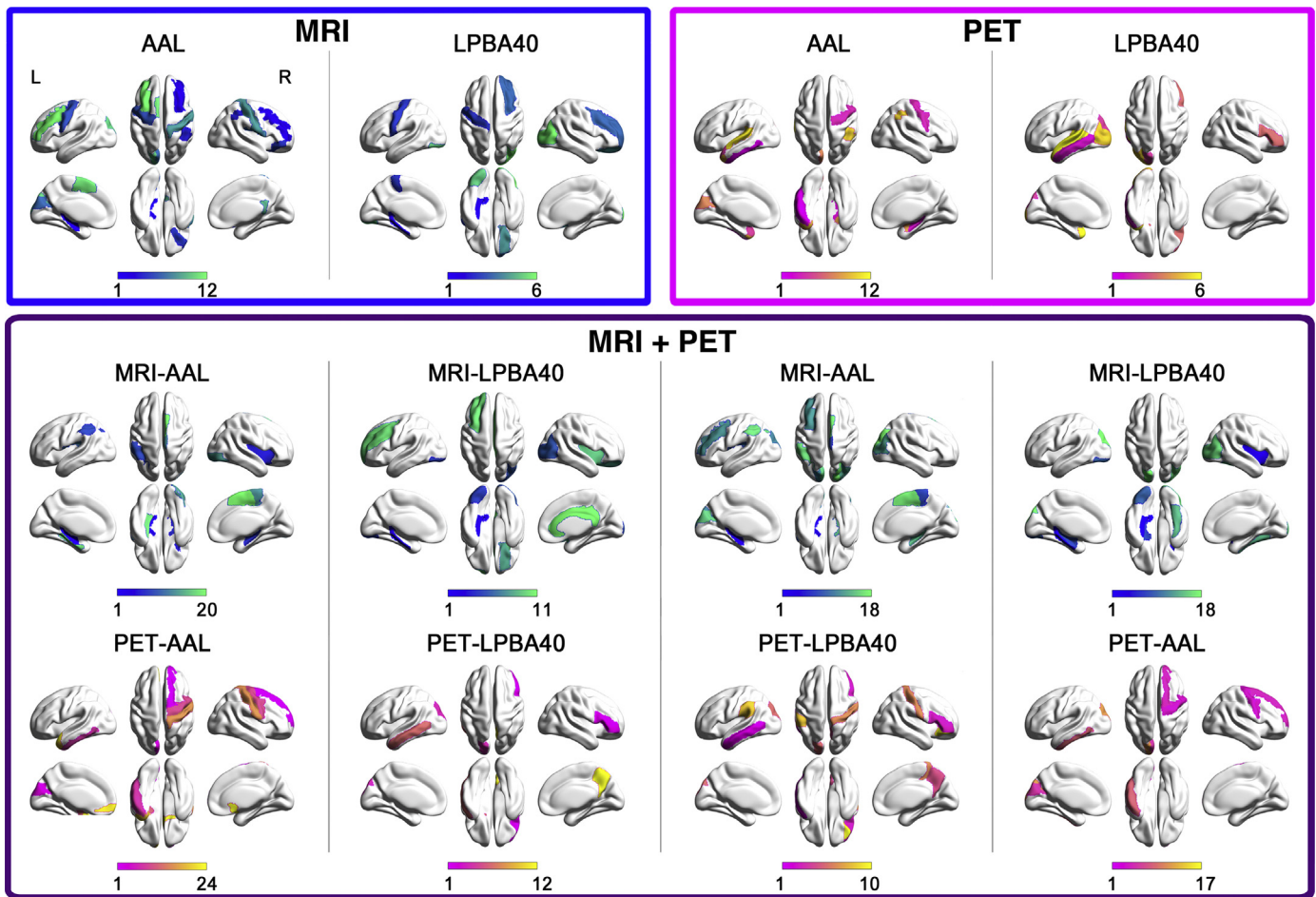
The principal aim of this study was to assess the effects of imaging modalities and brain atlases for feature extraction on prediction of conversion from MCI to AD. The results of this study indicate that preferable brain atlases for feature extraction can differ on the basis of the imaging modality used for unimodal features and for multimodal features and that combining multimodal imaging biomarkers improved the ability to predict conversion from MCI to AD. We believe what is practically important for this prediction is the choice of biomarkers (features) to be used and how to combine these biomarkers (Frisoni et al., 2013).

Our results suggest not only that multimodal features provide better performance for prediction than unimodal features but also that FDG-PET can be superior to MRI for early detection of AD, particularly when using AAL for feature extraction. For example, our two-way ANOVA for comparing the effects of atlases in multimodal feature sets demonstrated that the effect of MRI atlas at AAL for PET was not different. That is, the AAL-based multimodal combination (MRI-AAL + PET-AAL), which contained the weakest feature set (MRI-AAL), was comparable to the better heterogeneous multimodal combination (MRI-LPBA40 + PET-AAL).

was significant [ $F(1, 76) = 593.4, p < 10^{-15}, \eta^2 = 0.88$ ]. All of the simple main effects of modality for AAL [ $F(1, 76) = 355.1, p < 10^{-15}, \eta^2 = 0.53$ ] and for LPBA40 [ $F(1, 76) = 243.5, p < 10^{-15}, \eta^2 = 0.36$ ] and of atlas for MRI [ $F(1, 76) = 312.9, p < 10^{-15}, \eta^2 = 0.46$ ] and for PET [ $F(1, 76) = 280.9, p < 10^{-15}, \eta^2 = 0.42$ ] were significant. (E) Interaction plots of the two-way ANOVA for multimodal feature sets. The mean  $AUC_{RFE}$  values for different atlases for MRI or PET in multimodal feature sets were tested by another two-way (atlas for MRI  $\times$  atlas for PET) ANOVA. The interaction between atlases for MRI and PET was significant [ $F(1, 76) = 162.3, p < 10^{-15}, \eta^2 = 0.14$ ]. The simple main effect of atlas for MRI at AAL for PET was not significant [ $F(1, 76) = 0.60, p = 0.44, \eta^2 = 0.0005$ ], whereas we found the significant simple main effects of atlas for MRI at LPBA40 for PET [ $F(1, 76) = 353.3, p < 10^{-15}, \eta^2 = 0.31$ ], of atlas for PET at AAL for MRI [ $F(1, 76) = 780.1, p < 10^{-15}, \eta^2 = 0.68$ ], and of atlas for PET at LPBA40 for MRI [ $F(1, 76) = 98.2, p < 10^{-14}, \eta^2 = 0.09$ ]. \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . AUC, area under the receiver operating characteristic (ROC) curve; MRI, magnetic resonance imaging; PET, positron emission tomography; ANOVA, analysis of variance; AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas; RFE, recursive feature elimination; NS, not significant.

**Table 6**  
Labels for brain regions defined in the two brain atlases (AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas). The labels for AAL were adapted from Tzourio-Mazoyer et al. (2002) and modified for those for LPBA40.

AAL		LPBA40	
Brain region	Label	Brain region	Label
<b>Central region</b>			
Precentral gyrus	PRE	Precentral gyrus	PRE
Postcentral gyrus	POST	Postcentral gyrus	POST
Rolandic operculum	RO		
<b>Frontal lobe</b>			
Lateral surface			
Superior frontal gyrus, dorsolateral	F1	Superior frontal gyrus	F1
Middle frontal gyrus	F2	Middle frontal gyrus	F2
Inferior frontal gyrus, opercular part	F3OP	Inferior frontal gyrus	F3
Inferior frontal gyrus, triangular part	F3 T		
Medial surface			
Superior frontal gyrus, medial	F1 M		
Supplementary motor area	SMA		
Paracentral lobule	PCL		
Orbital surface			
Superior frontal gyrus, orbital part	F1O	Medial orbitofrontal gyrus	MOFG
Superior frontal gyrus, medial orbital	F1MO	Lateral orbitofrontal gyrus	LOFG
Middle frontal gyrus, orbital part	F2O		
Inferior frontal gyrus, orbital part	F3O		
Gyrus rectus	GR	Gyrus rectus	GR
Olfactory cortex	OC		
<b>Temporal lobe</b>			
Lateral surface			
Superior temporal gyrus	T1	Superior temporal gyrus	T1
Heschl's gyrus	HES		
Middle temporal gyrus	T2	Middle temporal gyrus	T2
Inferior temporal gyrus	T3	Inferior temporal gyrus	T3
<b>Parietal lobe</b>			
Lateral surface			
Superior parietal gyrus	P1	Superior parietal gyrus	P1
Inferior parietal, but supramarginal and angular gyri	P2		
Angular gyrus	AG	Angular gyrus	AG
Supramarginal gyrus	SMG	Supramarginal gyrus	SMG
Medial surface			
Precuneus	PQ	Precuneus	PQ
<b>Occipital lobe</b>			
Lateral surface			
Superior occipital gyrus	O1	Superior occipital gyrus	O1
Middle occipital gyrus	O2	Middle occipital gyrus	O2
Inferior occipital gyrus	O3	Inferior occipital gyrus	O3
Medial and inferior surfaces			
Cuneus	Q	Cuneus	Q
Calcarine fissure and surrounding cortex	V1		
Lingual gyrus	LING	Lingual gyrus	LING
Fusiform gyrus	FUSI	Fusiform gyrus	FUSI
<b>Limbic lobe</b>			
Temporal pole: superior temporal gyrus	T1P		
Temporal pole: middle temporal gyrus	T2P		
Anterior cingulate and paracingulate gyri	ACIN	Cingulate gyrus	CIN
Median cingulate and paracingulate gyri	MCIN		
Posterior cingulate gyrus	PCIN		
Hippocampus	HIP	Hippocampus	HIP
Parahippocampal gyrus	PHIP	Parahippocampal gyrus	PHIP
<b>Insula</b>			
Insula	IN	Insular cortex	IN
<b>Subcortical gray nuclei</b>			
Amygdala	AMYG		
Caudate nucleus	CAU	Caudate	CAU
Lenticular nucleus, putamen	PUT	Putamen	PUT
Lenticular nucleus, pallidum	PAL		
Thalamus	THA		
<b>Cerebellum</b>			
Cerebellum Crus 1	CBL1	Cerebellum	CBL
Cerebellum Crus 2	CBL2		
Cerebellum 3	CBL3		
Cerebellum 4 5	CBL45		
Cerebellum 6	CBL6		
Cerebellum 7b	CBL7B		
Cerebellum 8	CBL8		
Cerebellum 9	CBL9		
Cerebellum 10	CBL10		
Vermis 1 2	VER12		
Vermis 3	VER3		
Vermis 4 5	VER45		
Vermis 6	VER6		
Vermis 7	VER7		
Vermis 8	VER8		
Vermis 9	VER9		
Vermis 10	VER10		
		<b>Brainstem</b>	<b>BS</b>



**Fig. 6.** Top 10 ranked cortical regions for eight feature sets according to support vector machine-based recursive feature elimination feature selection. The cerebellar regions are not shown for simplicity. MRI, magnetic resonance imaging; PET, positron emission tomography; AAL, Automated Anatomical Labeling; LPBA40, LONI Probabilistic Brain Atlas.

The other combinations both comprised PET-LPBA40, which gave an AUC lower than that of PET-AAL but higher than that of MRI-AAL. One possible explanation is that FDG-PET biomarkers, particularly when extracted with AAL, can better detect the difference in temporospatial distribution patterns of synaptic loss in the brains of patients with aMCI. Besides amyloid plaques and NFTs, AD is also characterized by the loss of neurons and their synapses particularly in the cerebral cortex and hippocampus (Scheff and Price, 2006). Although amyloid plaques and NFTs are the major neuropathological hallmarks of AD, amyloid accumulation in the brains of patients with AD is probably not the cause of the neurodegeneration but merely a marker of some upstream alterations that can cause neuronal and synaptic loss (Drachman, 2014). Synaptic loss and neuronal loss are the major pathological substrates of cortical atrophy (Serrano-Pozo et al., 2011). Synaptic loss is the major correlate of cognitive impairment (Terry et al., 1991) and the most reliable index of cognition in both postmortem and biopsied AD brains (DeKosky et al., 1996). *In vivo*  $^{18}\text{F}$ -FDG uptake is strongly correlated with cerebral synaptic density and activity (Rocher et al., 2003). Cognitive decline is strongly associated with glucose hypometabolism in frontal and temporoparietal regions in patients with probable AD (Furst et al., 2012). Temporal and parietal glucose metabolism predict decline in global cognitive function, and medial temporal brain volumes predict memory decline in normal older people (Jagust et al., 2006). These findings and our results, therefore, support that FDG-PET biomarkers and particularly multimodal combinations with other imaging modalities, such as structural MRI, can have an important role in cognitive decline.

From Table 1, it is clear that the converters are more impaired cognitively than non-converters, particularly verbal memory measured by WMS-R-LM. We tested the prediction accuracy using these cognitive scores (Supplementary material). WMS-R-LM II alone demonstrated the highest AUC value among the cognitive feature sets, which outperformed imaging biomarkers alone as we expected (Supplementary Fig. 1). However, some combined feature sets of cognitive scores and imaging biomarkers outperformed cognitive scores alone or imaging biomarkers alone, and the performance varied depending on how to combine these measures (Supplementary Figs. 2 and 3). First, diagnosis of dementia (due to AD) is done clinically. Cognitive scores such as episodic memory measures including WMS-R LM I and II immediate and delayed recall were used for cognitive assessment to differentiate MCI from AD. Thus cognitive scores are somewhat like supervisory signals, and use of cognitive scores for prediction may not be suitable. Second, even if the imaging biomarkers bring little additional discriminative information in terms of predictive accuracy, these injury biomarkers bring additional topographic information that may help to characterize clinical phenotypes (patterns of cortical atrophy or hypometabolism) (Dubois et al., 2014). Therefore, we believe that imaging biomarkers have some advantages to unravel the underlying topographic pathophysiological features compared to cognitive scores.

Regarding the effect of atlas on predictive performance, our two-way ANOVA for comparing the effects of atlas and modality in the unimodal feature sets demonstrated that the performance differed on the basis of the atlases on both MRI and PET. The major difference



between AAL and LPBA40 is the number of ROIs. AAL has 116 ROIs, more than twice the number in LPBA40 (56 ROIs). In addition, AAL includes 26 finely parcellated cerebellar regions, whereas LPBA40 contains the cerebellum as only one large region. These differences in parcellation between these atlases may be associated with the different effects on the classification performance.

Previous studies have shown superior accuracies using other methods than atlas-based approaches. However, in view of clinical applications, for example, electrocardiogram with limited leads, methods for measurement and analysis would be better as simple as possible, as long as the performance is not too poor. Atlas-based analysis is not the best in performance compared to other methods, but it provides an accuracy of around 70%. In addition, atlas-based methods may be good for clinical applications because clinical practitioners usually interpret the results in an anatomical framework. Results therefore should be given regions-of-interest (ROIs) using anatomical notation whatever features are used. Even if data-driven parcellation methods are used, we need to present the results using a reference like the MNI coordinate system. We therefore believe that it is worth using the atlas-based analysis.

Another purpose of this study was to assess the influence of SVM-RFE feature selection on the performance of MCI classification. The importance of feature selection in machine learning studies for AD prediction remains controversial (Chu et al., 2012; Cuingnet et al., 2011; Kerr et al., 2014). In this study, we found a significant main effect of SVM-RFE on the performance of MCI classification. Depending on the atlases and modalities, the influence of SVM-RFE can also differ, although there were no significant interactions between SVM-RFE and the other factors. We think that this may be because of a kind of ceiling effect, which resulted from the difficulty in binary classification of MCI into converters and nonconverters. As compared with discrimination of MCI from or AD from normal age-related deficits (Cuingnet et al., 2011), the accuracy of MCI classification is at most approximately 80% (Cuingnet et al., 2011; Misra et al., 2009; Querbes et al., 2009). This may be partly because of the intrinsic heterogeneity in MCI (Nettiksimmons et al., 2014) or possibly because of variations in the spatial and temporal progression of pathological changes. In addition, “nonconverters” include possible converters in principle because nonconverters may develop AD after 3 years of follow-up.

The optimal prediction accuracy was achieved using multiple features. For MRI, it was around 25 when using LPBA40. This suggests that removing less discriminating set of features by feature selection may improve predictive accuracy. However, for PET and multi-modal approaches, the optimal was without selection. These results should be compared with a recent study, where only 5 features optimize the prediction (Eskildsen et al., 2015). There could be a potential problem of overfitting the model when the number of features is greater than that of subjects. Overfitting leads to poor generalization.

We evaluated this atlas-based approach using the ADNI dataset as an independent cohort (Supplementary Material). Although atlas × modality interaction was also significant in the ADNI dataset, the results of ADNI were different from those of SEAD-J. One possible explanation is sample selection bias. The ADNI dataset and our dataset are different in including criteria (Kawashima et al., 2012), resulting in different baseline characteristics. In addition, individual genetic variations exert lasting influences on brain structures and functions associated with behavior and predisposition to disease (Hibar et al., 2015). This may also be associated with the different results.

Such sample selection bias affects both learning and evaluation (Zadrozny, 2004). Learning methods can be classified into global and local learners. Global learners, such as naive Bayes, soft margin SVM, and decision tree learners, are affected by sample selection

bias, while local learners, such as logistic regression and hard margin SVM, are insensitive to sample selection bias. However, the evaluation step is always affected by sample selection bias.

In ADNI datasets, multimodal datasets did not significantly outperform unimodal datasets. The sample size of the ADNI subjects was 158, almost double the number of subjects SEAD-J, 80. Overfitting in the SEAD-J results for multimodal feature sets cannot be ruled out. If there is no overfitting in the ADNI results, increased-feature sizes including irrelevant features may be associated with no improvement in the performance for the ADNI multimodal feature sets.

Another difference between ADNI and SEAD-J was the effect of feature selection. In the case of ADNI, the effect of SVM-RFE on prediction performance was remarkable compared to SEAD-J, particularly when using unimodal datasets. Difference in sample size between these datasets may also be associated with the difference in the effect of feature selection. Regarding feature selection in small sample size data, the lack of relation between the errors of the best and selected feature sets was observed more evidently in smaller sample sizes than for larger sample sizes (Sima and Dougherty, 2006). In addition, SVM-RFE is generally sensitive to noise and outliers (Nijima and Kuhara, 2006) when it is applied to small sample size data. However, increase in sample size may result in a more heterogeneous sample, and this heterogeneity may make it difficult to predict conversion. Stratification into more homogeneous subgroups may help us predict AD conversion and understand underlying pathophysiological mechanisms.

We believe that SVM-RFE feature selection is useful not only for reducing the dimension of the feature space to avoid the issue of dimensionality and overfitting but also for finding valuable regions in view of the consistency of the selected regions (the left hippocampus on MRI and the inferior temporal gyrus, the precuneus, and the angular gyrus on FDG-PET) observed by using our voxel-based analysis and observed in previous studies (Chételat et al., 2005; Zhang et al., 2011). In this study, the inferior temporal gyrus and the precuneus were identified as important regions for MCI classification by using both SVM-RFE and voxel-based analysis from FDG-PET data. Scheff et al. reported synaptic loss in the inferior temporal gyrus (Scheff et al., 2011) and the precuneus (Scheff et al., 2013) in aMCI patients. According to their studies, the inferior temporal gyrus is affected during the prodromal stage of the disease (Scheff et al., 2011), whereas the precuneus does not show early changes in synaptic decline during the progression of AD (Scheff et al., 2013). On the other hand, significant <sup>18</sup>F-FDG uptake reductions in patients with very mild AD relative to that in normal controls have been found in the precuneus and other neocortical regions, including the posterior cingulate and left temporoparietal and frontal association cortex (Herholz et al., 2002). This spatiotemporal difference in local cerebral glucose metabolism is similar to the characteristic progression patterns of NFTs (Braak and Braak, 1991) and synaptic loss (Scheff and Price, 2006), which originate from the entorhinal cortex and extend through the limbic regions to the neocortex during disease progression. The inferior temporal gyrus has neural interconnections with the structures in the medial temporal cortex, particularly the parahippocampal gyrus (Suzuki and Amaral, 1994) and has an important role in verbal fluency (Scheff et al., 2011). Regional cerebral glucose metabolism in the left inferior temporal region in the brains of patients with mild AD correlates consistently with verbal semantic memory measures (Hirono et al., 2001). A meta-analysis showed that verbal fluency, particularly semantic fluency rather than phonemic fluency, was significantly more impaired than measures of verbal intelligence and psychomotor speed, and episodic memory appeared to be most disrupted by AD (Henry et al., 2004). These findings are consistent with the significant differences in WMS-R-LM in the present study. As mentioned previously (Ota et al., 2014), most of

the participants in this study were subclassified as having late MCI, which is a high risk group for AD (Jessen et al., 2014) and defined by the education-adjusted ranges of the WMS-R-LM II score (Aisen et al., 2010). Decline in verbal functions may be associated with the left-dominant laterality in the significant regions in our voxel-based analysis. These findings raise the possibility that SVM-RFE feature selection could detect temporally different underlying pathological progressions. If so, these findings and our results suggest that glucose hypometabolism in the inferior temporal gyrus might be an early biomarker for discriminating converters from nonconverters. In addition, the left hippocampus (AAL 37) on MRI is a robust biomarker for MCI binary classification as a single feature (Ota et al., 2014). Stratification of MCI into subtypes with combined specific features could provide better diagnostic ability or more detailed information on AD/MCI pathophysiology. To achieve further evidence, evaluation by using other datasets is desired in the future. In addition, it would be useful to apply atlas-based analysis to other imaging biomarkers, such as white matter hyperintensities (Nettiksimmons et al., 2014; Provenzano et al., 2013).

Our validation set up was rather complex. Validation such as 10-fold cross-validation would be easier to be implemented than LOOCV. However, several previous studies used LOOCV (Eskildsen et al., 2013, 2015). A previous study mentioned that 10-fold cross-validation and LOOCV is not significantly different when the number of features is 10 to 100 (Chu et al., 2012, Supplementary data). We therefore consider that use of LOOCV in this study is not a major problem.

A possible limitation of our study is the MRI acquisition on multiple scanners including 1.5-T and 3-T. The numbers of images for 9 sites (A–I) were 3, 15, 15, 5, 7, 16, 7, 2, and 10, respectively ( $\chi^2$  test,  $p=0.001$ ). There could be potential biases. Although the influence of multi-slice imaging on contrast-to-noise ratio between gray and white matter is significantly larger at 3-T than at 1.5-T (Fushimi et al., 2007), the influence may be limited because only 3 participants (4%) were scanned on 3-T in this study.

Atlas-based feature extraction is methodologically simple and has a much lower computational cost than voxel-based adaptive approaches. In addition, atlas-based analysis can provide regional information within each ROI in the brains of subjects that can readily be interpreted by clinicians. In view of possible clinical applications, our results suggest that the use of AAL may be recommended for feature extraction from PET or multimodal biomarkers.

## 5. Conclusions

Our study demonstrated that imaging modalities and brain atlases for feature extraction affect prediction and interact with each other. In addition, SVM-RFE feature selection can itself improve the performance of classification. Regions selected by SVM-RFE were generally consistent with those selected in previous studies and with our results from group comparisons. Besides atlas-based analysis, other techniques for feature extraction include data-driven parcellation methods (Fan et al., 2007) and dimension reduction by factor analysis (Desikan et al., 2010). Data-driven approaches can well reflect the underlying structure of the data, whereas they are usually computationally expensive and the resulting output tends to be complicated and have interpretational difficulties. Dimension reduction by factor analysis can also learn intrinsic data structure without the need for supervisory signals resulting in good generalization, whereas it is not always possible to improve the predictive accuracy. Regarding features for classification, the Ugly Duckling theorem (Watanabe, 1969) indicates that an MCI nonconverter is as similar to an MCI converter as to another nonconverter. Selection of features for classification is arbitrary,

and based on subjective criteria that depend on the purpose or the task. If feature selection for AD prediction is done based on an algorithm, no free lunch theorems (Wolpert, 1996) may also hold true. This implies that prior domain knowledge may be important for AD prediction. Although the sample size was limited, we consider atlas-based analysis to be a useful tool for finding promising biomarkers because of its simplicity and ease of interpreting the results. In the future, the combination of atlas-based analysis with unsupervised feature representation learning methods may be a useful tool for finding novel biomarkers. Longitudinal analysis using our method will be useful for further investigation of MCI classification.

## Acknowledgments

This work was supported by a Health Labor Sciences Research Grant from the Ministry of Health, Labor, and Welfare of Japan (H17-Tyojyu-023) and the Research Funding for Longevity Sciences from the National Center for Geriatrics and Gerontology, Japan (20-1). This work was partly supported by the following: Grant-in-Aid for Young Scientists (B) 24791296 from the Japan Society for the Promotion of Science (JSPS) (to N.O.) and Development of BMI Technologies for Clinical Application under the Strategic Research Program for Brain Sciences by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (to N.O.). The authors appreciate the efforts and contributions of those who engaged in the subjects' care and the collection of FDG-PET and MRI images and clinical reports.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jneumeth.2015.08.020>.

## References

- Aisen PS, Petersen RC, Donohue MC, Gamst A, Raman R, Thomas RG, et al. *Clinical core of the Alzheimer's disease neuroimaging initiative: progress and plans*. *Alzheimers Dement* 2010;6:239–46.
- 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;7:270–9.
- Alzheimer's Association. *2013 Alzheimer's disease facts and figures*. *Alzheimers Dement* 2013;9:208–45.
- Bohland JW, Bokil H, Allen CB, Mitra PP. *The brain atlas concordance problem: quantitative comparison of anatomical parcellations*. *PLoS One* 2009;4:e7200.
- Braak H, Braak E. *Neuropathological staging of Alzheimer-related changes*. *Acta Neuropathol* 1991;82:239–59.
- Breiman L. *Bagging predictors*. *Mach Learn* 1996;24:123–40.
- Chang C-C, Lin C-J. *LIBSVM: a library for support vector machines*. *ACM Trans Intell Syst Technol* 2011;2(27):1–27 [27].
- Chao LL, Buckley ST, Kornak J, Schuff N, Madison C, Yaffe K, et al. *ASL perfusion MRI predicts cognitive decline and conversion from MCI to dementia*. *Alzheimer Dis Assoc Disord* 2010;24:19–27.
- Chételat G, Landeau B, Eustache F, Mézenge F, Viader F, la Sayette de V, et al. *Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study*. *Neuroimage* 2005;27:934–46.
- 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.
- Cohen J. *Eta-squared and partial eta-squared in fixed factor Anova designs*. *Educ Psychol Meas* 1973;33:107–12.
- Cohen J. *A power primer*. *Psychol Bull* 1992;112:155–9.
- 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.
- DeKosky ST, Scheff SW, Styren SD. *Structural correlates of cognition in dementia: quantification and assessment of synapse change*. *Neurodegeneration* 1996;5:417–21.

- Desikan RS, Cabral HJ, Settecase F, Hess CP, Dillon WP, Glastonbury CM, et al. Automated MRI measures predict progression to Alzheimer's disease. *Neurobiol Aging* 2010;31:1364–74.
- Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A, Misc functions of the department of statistics (e1071), TU Wien, 2005. Available from: <http://cran.r-project.org/web/packages/e1071/index.html>.
- Dosenbach NUF, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, et al. Prediction of individual brain maturity using fMRI. *Science* 2010;329:1358–61.
- Drachman DA. The amyloid hypothesis, time to move on: amyloid is the downstream result, not cause, of Alzheimer's disease. *Alzheimers Dement* 2014;10:372–80.
- Duan K-B, Rajapakse JC, Wang H, Azuaje F. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans Nanobiosci* 2005;4:228–34.
- Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol* 2014;13:614–29.
- Duda RO, Hart PE, Stork DG. *Pattern classification*. 2nd ed. New York: John Wiley & Sons, Inc.; 2001.
- Dukart J, Mueller K, Horstmann A, Vogt B, Frisch S, Barthel H, et al. Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. *Neuroimage* 2010;49:1490–5.
- 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.
- Eskildsen SF, Coupé P, Fonov VS, Pruessner JC, Collins DL. Alzheimer's disease neuroimaging initiative: structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiol Aging* 2015;36(Suppl. 1):S23–31.
- 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.
- Fawcett T. ROC graphs: notes and practical considerations for researchers. *Mach Learn* 2004;31:1–38.
- 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.
- Fox J, Weisberg S. Car: companion to applied regression. Available at (<http://cran.r-project.org/web/packages/car/index.html>); 2009.
- Frisoni GB, Bocchetta M, Chételat G, Rabinovici GD, de Leon MJ, Kaye J, et al. Imaging markers for Alzheimer disease: which vs how. *Neurology* 2013;81:487–500.
- Furst AJ, Rabinovici GD, Rostomian AH, Steed T, Alkalay A, Racine C, et al. Cognition, glucose metabolism and amyloid burden in Alzheimer's disease. *Neurobiol Aging* 2012;33:215–25.
- Fushimi Y, Miki Y, Urayama S-I, Okada T, Mori N, Hanakawa T, et al. Gray matter-white matter contrast on spin-echo T1-weighted images at 3-T and 1.5-T: a quantitative comparison study. *Eur Radiol* 2007;17:2921–5.
- Giacobini E, Gold G. Alzheimer disease therapy—moving from amyloid- $\beta$  to tau. *Nat Rev Neurol* 2013;9:677–86.
- Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;46:389–422.
- Henry JD, Crawford JR, Phillips LH. Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia* 2004;42:1212–22.
- Herholz K, Salmon E, Perani D, Baron J-C, Holthoff V, Frölich L, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage* 2002;17:302–16.
- Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015;520:224–9.
- Hirono N, Mori E, Ishii K, Imamura T, Tanimukai S, Kazui H, et al. Neuronal substrates for semantic memory: a positron emission tomography study in Alzheimer's disease. *Dement Geriatr Cogn Disord* 2001;12:15–21.
- 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 Psychiatry* 1992;3:647–55.
- Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. *Knowl Data Eng, IEEE Trans* 2005;17:299–310.
- Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* 2012;8:1–13.
- Ito K, Mori E, Fukuyama H, Ishii K, Washimi Y, Asada T, et al. Prediction of outcomes in MCI with (123)I-IMP-CBF SPECT: a multicenter prospective cohort study. *Ann Nucl Med* 2013;27:898–906.
- Ito K, Fukuyama H, Senda M, Ishii K, Maeda K, Yamamoto Y, et al. Prediction of outcomes in mild cognitive impairment by using 18F-FDG-PET: a multicenter study. *J Alzheimers Dis* 2015;45:543–52.
- 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;7:257–62.
- Jack CR, Holtzman DM. Biomarker modeling of Alzheimer's disease. *Neuron* 2013;80:1347–58.
- Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 2010;9:119–28.
- Jagust W, Gitcho A, Sun F, Kuczynski B, Mungas D, Haan M. Brain imaging evidence of preclinical Alzheimer's disease in normal aging. *Ann Neurol* 2006;59:673–81.
- Jessen F, Wolfsgruber S, Wiese B, Bickel H, Mösch E, Kaduszkiewicz H, et al. AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimers Dement* 2014;10:76–83.
- Karow DS, McEvoy LK, Fennema-Notestine C, Hagler DJ, Jennings RG, Brewer JB, et al. Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. *Radiology* 2010;256:932–42.
- Kawashima S, Ito K, Kato T. SEAD-J study group inclusion criteria provide heterogeneity in baseline profiles of patients with mild cognitive impairment: comparison of two prospective cohort studies. *Br Med J Open* 2012;2:e000773.
- Kerr WT, Douglas PK, Anderson A, Cohen MS. The utility of data-driven feature selection: re: Chu et al. (2012). *Neuroimage* 2014;84:1107–10.
- Kurth F, Luders E, Gaser C. VBM8-Toolbox manual [Internet]. Available from: (<http://dbm.neuro.uni-jena.de/vbm/>); 2010.
- Li H-J, Hou X-H, Liu H-H, Yue C-L, He Y, Zuo X-N. Toward systems neuroscience in mild cognitive impairment and Alzheimer's disease: a meta-analysis of 75 fMRI studies. *Hum Brain Mapp* 2015;36:1217–32.
- Langsrud Ø. ANOVA for unbalanced data: use type II instead of type III sums of squares. *Stat Comput* 2003;13:163–7.
- McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:263–9.
- Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283–98.
- Mével K, Desgranges B, Baron J-C, Landeau B, la Sayette de V, Viader F, et al. Detecting hippocampal hypometabolism in mild cognitive impairment using automatic voxel-based approaches. *Neuroimage* 2007;37:18–25.
- 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.
- Mosconi L, Sorbi S, De Leon MJ, Li Y, Nacmias B, Myoung PS, et al. Hypometabolism exceeds atrophy in presymptomatic early-onset familial Alzheimer's disease. *J Nucl Med* 2006;47:1778–86.
- Mufson EJ, Binder L, Counts SE, DeKosky ST, de Toledo-Morrell L, Ginsberg SD, et al. Mild cognitive impairment: pathology and mechanisms. *Acta Neuropathol* 2012;123:13–30.
- Nettiksimmons J, DeCarli C, Landau S, Beckett L. Alzheimer's disease neuroimaging initiative. Biological heterogeneity in ADNI amnesic mild cognitive impairment. *Alzheimers Dement* 2014;10:511–21.e1.
- Nijijima S, Kuhara S. Recursive gene selection based on maximum margin criterion: a comparison with SVM-RFE. *BMC Bioinform* 2006;7:543.
- 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.
- Oishi K, Akhter K, Mielke M, Ceritoglu C, Zhang J, Jiang H, et al. Multi-modal MRI analysis with disease-specific spatial filtering: initial testing to predict mild cognitive impairment patients who convert to Alzheimer's disease. *Front Neurol* 2011;2:54.
- Ota K, Oishi N, Ito K, Fukuyama H. SEAD-J study group. A comparison of three brain atlases for MCI prediction. *J Neurosci Methods* 2014;221:139–50.
- Petersen RC, Smith GE, Wang LC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 1999;56:303–8.
- Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, et al. Current concepts in mild cognitive impairment. *Arch Neurol* 2001;58:1985–92.
- Price JC. Molecular brain imaging in the multimodality era. *J Cereb Blood Flow Metab* 2012;32:1377–92.
- Provenzano FA, Muraskin J, Tosto G, Narkhede A, Wasserman BT, Griffith EY, et al. White matter hyperintensities and cerebral amyloidosis: necessary and sufficient for clinical expression of Alzheimer disease? *J Am Med Assoc Neurol* 2013;70:455–61.
- Querbes O, Aubry F, Pariante J, Lotterie JA, Demonet JF, Duret V, et al. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* 2009;132:2036–47.
- R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013.
- 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 Bioinform* 2011;12:77.
- Rocher AB, Chapon F, Blaizot X, Baron J-C, Chavoix C. Resting-state brain glucose utilization as measured by PET is directly related to regional synaptophysin levels: a study in baboons. *Neuroimage* 2003;20:1894–8.
- Scheff SW, Price DA, Schmitt FA, Mufson EJ. Hippocampal synaptic loss in early Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging* 2006;27:1372–84.
- Scheff SW, Price DA, Schmitt FA, Roberts KN, Ikonovic MD, Mufson EJ. Synapse stability in the precuneus early in the progression of Alzheimer's disease. *J Alzheimers Dis*: JAD 2013;35:599–609.
- Scheff SW, Price DA, Schmitt FA, Scheff MA, Mufson EJ. Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis*: JAD 2011;24:547–57.
- Scheff SW, Price DA. Alzheimer's disease-related alterations in synaptic density: neocortex and hippocampus. *J Alzheimers Dis*: JAD 2006;9:101–15.
- Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med* 2011;1:a006189.

- 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.
- Shokouhi S, Claassen D, Kang H, Ding Z, Rogers B, Mishra A, et al. Longitudinal progression of cognitive decline correlates with changes in the spatial pattern of brain 18F-FDG PET. *J Nucl Med* 2013;54:1564–9.
- Sima C, Dougherty ER. What should be expected from feature selection in small-sample settings. *Bioinformatics* 2006;22:2430–6.
- Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:280–92.
- 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.
- Suzuki WA, Amaral DG. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J Comp Neurol* 1994;350:497–533.
- Terry RD, Masliah E, Salmon DP, Butters N, DeTeresa R, Hill R, et al. Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann Neurol* 1991;30:572–80.
- 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.
- Valentini G, Muselli M, Ruffino F. Cancer recognition with bagged ensembles of support vector machines. *Neurocomputing* 2004;56:461–6.
- Vapnik VN. *Statistical learning theory*. New York: Wiley; 1998.
- Watanabe S. *Knowing and guessing: A quantitative study of inference and information*. New York: Wiley; 1969.
- Wolpert DH. The existence of a priori distinctions between learning algorithms. *Neural Comput* 1996;8:1391–420.
- Xia M, Wang J, He Y. BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS One* 2013;8:e68910.
- 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.
- Young J, Modat M, Cardoso MJ, Mendelson A, Cash D, Ourselin S, et al. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin* 2013;2:735–45.
- Zadrozny B. Learning and evaluating classifiers under sample selection bias. In: *Proceedings of the twenty-first international conference on machine learning*. New York: Association for Computing Machinery; 2004. p. 903–10.
- Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 2011;55:856–67.