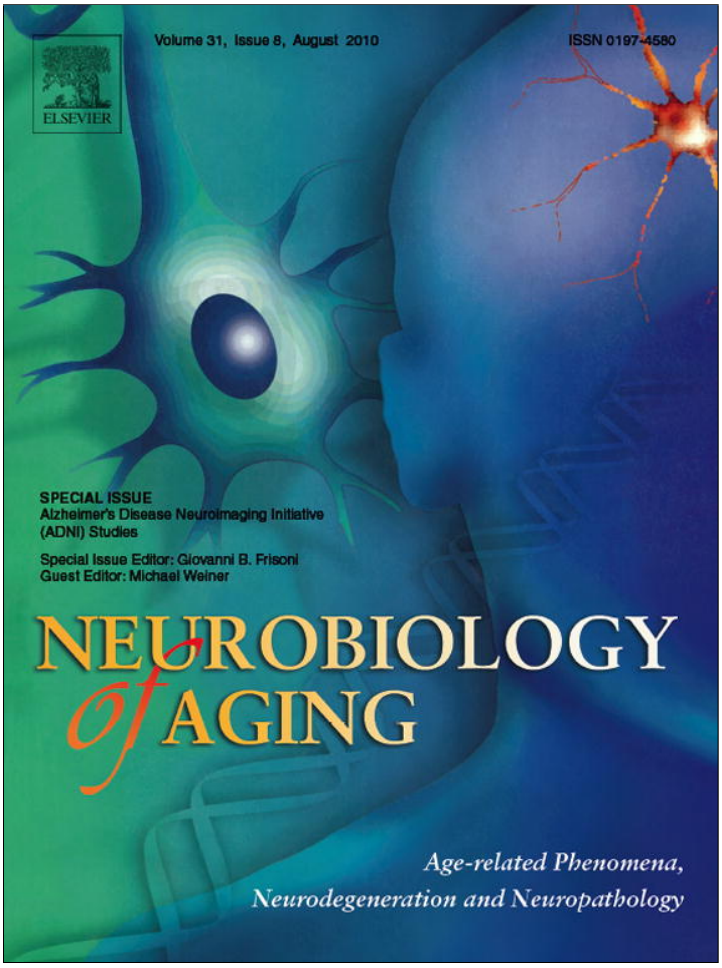
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.


ELSEVIER
Volume 31, Issue 8, August 2010
ISSN 0197-4580
SPECIAL ISSUE
Alzheimer's Disease Neuroimaging Initiative (ADNI) Studies
Special Issue Editor: Giovanni B. Frisoni
Guest Editor: Michael Weiner
NEUROBIOLOGY of AGING
Age-related Phenomena, Neurodegeneration and Neuropathology

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

ELSEVIER

Neurobiology of Aging 31 (2010) 1429–1442

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

Boosting power for clinical trials using classifiers based on multiple biomarkers

Omid Kohannim[a], Xue Hua[a], Derrek P. Hibar[a], Suh Lee[a], Yi-Yu Chou[a], Arthur W. Toga[a], Clifford R. Jack, Jr.[b], Michael W. Weiner[c,d,e], Paul M. Thompson[a,*], The Alzheimer's Disease Neuroimaging Initiative[†]

[a] *Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA, USA*
[b] *Department of Radiology, Mayo Clinic, Rochester, MN, USA*
[c] *Department of Radiology and Biomedical Imaging, UCSF, San Francisco, CA, USA*
[d] *Department of Medicine, UCSF, San Francisco, CA, USA*
[e] *Department of Psychiatry, UCSF, San Francisco, CA, USA*

Received 3 February, 2010; received in revised form 6 April 2010; accepted 22 April 2010

Abstract

Machine learning methods pool diverse information to perform computer-assisted diagnosis and predict future clinical decline. We introduce a machine learning method to boost power in clinical trials. We created a Support Vector Machine algorithm that combines brain imaging and other biomarkers to classify 737 Alzheimer's disease Neuroimaging initiative (ADNI) subjects as having Alzheimer's disease (AD), mild cognitive impairment (MCI), or normal controls. We trained our classifiers based on example data including: MRI measures of hippocampal, ventricular, and temporal lobe volumes, a PET-FDG numerical summary, CSF biomarkers (t-tau, p-tau, and Aβ_{42}), ApoE genotype, age, sex, and body mass index. MRI measures contributed most to Alzheimer's disease (AD) classification; PET-FDG and CSF biomarkers, particularly Aβ_{42}, contributed more to MCI classification. Using all biomarkers jointly, we used our classifier to select the one-third of the subjects most likely to decline. In this subsample, fewer than 40 AD and MCI subjects would be needed to detect a 25% slowing in temporal lobe atrophy rates with 80% power—a substantial boosting of power relative to standard imaging measures.
© 2010 Elsevier Inc. All rights reserved.

Keywords: Clinical trial enrichment; Alzheimer's disease; Mild cognitive impairment; Magnetic resonance imaging; Neuroimaging; Biomarkers; Classification; Support vector machines

Alzheimer's disease (AD), the most common form of dementia, affects approximately 5.3 million people in the USA alone, and its prevalence continues to rise (Alzheimer's Association, 2009). Research and therapeutic efforts also focus on subjects with Mild Cognitive Impairment (MCI) – an intermediate condition between healthy aging and AD – as they convert to AD at a heightened rate of 10–15% per year (Petersen et al., 1999). Multiple imaging biomarkers have been used for quantifying disease progression and measuring various aspects of AD pathology, such as amyloid and tau deposition, measured by Positron Emission Tomography (PET) and radiotracers that bind to the plaques and tangles in the brain (Klunk et al., 2004; Protas et al., 2010), metabolic decline or perfusion deficits assessed

[*] Correspondence: Paul Thompson, PhD, Professor of Neurology. Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Neuroscience Research Building 225E, 635 Charles Young Drive, Los Angeles, CA 90095-1769, USA. Tel.: (310) 206 2101; fax: (310) 206 5518.

E-mail address: thompson@loni.ucla.edu.

[†] Data used in the preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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. Complete listing of ADNI investigators available at: www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf.

0197-4580/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neurobiolaging.2010.04.022

Author's personal copy

by fluoro-deoxyglucose PET (PET-FDG), brain atrophy on MRI, and risk factors that influence these measures (e.g. ApoE, cardiovascular risks, etc.) (Frisoni et al., 2010; Jack et al., 2010; Petersen, 2010).

Although the disease can be tracked in many ways, methods are also needed to integrate these multiple measures to achieve greater power in diagnosis and prognosis. Machine learning algorithms such as linear discriminant analysis, support vector machines, and boosting have recently been proposed to combine multiple AD features derived from brain imaging and other biomarkers, for AD and MCI classification. Several studies have performed diagnostic classification based on MRI scans, using measures such as whole-brain patterns of atrophy (Davatzikos et al., 2009; Mesrob et al., 2008), tissue densities from voxel-based morphometry (Vemuri et al., 2008), and cortical thickness (Lerch et al., 2008). Vemuri et al. (2008) assigned overall "scores" for each subject's MRI – called the Structural Abnormality Index (STAND) − based on gray and white matter voxels that best differentiated AD patients from controls. In related work, Davatzikos et al. (2009) assigned "scores" to each subject's MRI scan based on a minimal set of brain regions that best discriminated AD from normal controls in a training sample; their approach is termed Spatial Pattern of Abnormality for Recognition of Early Alzheimer's disease, or SPARE-ED.

Researchers have also explored adding other predictors to improve the accuracy of MRI for computer-assisted diagnosis of AD and MCI, and for predicting whether a person will convert from MCI to AD soon. PET, for example, offers metabolic or perfusion-based information that complements measures of structural atrophy on MRI (Fan et al., 2008; Hinrichs et al., 2009). Vemuri et al. (2009) adjusted their STAND scores by incorporating demographic variables such as age, sex, and ApoE genotype, and this improved their classification accuracy. Additionally, MRI-based STAND scores were shown to improve the accuracy of CSF biomarkers for predicting cognitive decline, including total tau (t-tau), phosphorylated tau (p-tau) and the beta-amyloid isoform, Aβ_{42} (Vemuri et al., 2009).

It is worth noting that MRI-based machine learning has been used widely for classification not only for AD, but also for predicting changes in patients with brain tumors (Lukas et al., 2004), aphasia (Wilson et al., 2009), autism (Ecker et al., 2010), psychosis (Koutsouleris et al., 2009), and even for classifying patterns of brain activation in functional MRI (Mourão-Miranda et al., 2005). Similar algorithms have been implemented to distinguish AD from other types of dementia such as frontotemporal dementia (Davatzikos et al., 2008; Klöppel et al., 2008). Support vector machines (SVMs) are one of the most widely used and effective tools for classification of AD and other neurological disorders, and are used in many of the reports listed above. We therefore set out to test how well SVMs would perform for classifying patients as having AD and MCI based on multiple imaging and biological measures in the Alzheimer's Disease Neuroimaging Initiative (ADNI), as well as for predicting imminent cognitive decline.

A second goal of this paper was to make a conceptual connection between sample size requirements for clinical trials and the power of classifiers to predict future decline. Using our classifiers to predict those most likely to decline, we tested the hypothesis that this subset might experience atrophic rates with greater effect sizes. This concept is termed clinical trial enrichment, as it seeks out a subsample of subjects who might be better candidates for demonstrating therapeutic effects, at least from a statistical standpoint (see Discussion for assumptions of this approach).

We found that regional numerical summaries derived from tensor-based morphometry of longitudinal MRI (over a 1 year interval) can reduce the estimated sample size requirements to 48 AD and 88 MCI subjects per arm of a hypothetical clinical trial (treatment v. placebo), for detecting a 25% reduction in the mean annual temporal lobe atrophy rate with 80% power (Hua et al., 2009). Power was similar when 3 Tesla or 1.5 Tesla MRI scans were used (Ho et al., 2009); still higher power was possible for trials with longer follow-up intervals (Hua et al., 2010b). Other groups report comparable power for measures based on hippocampal volumes (Schuff et al., 2009). Through the use of multimodality classifiers, these and other similar sample size estimates can presumably be reduced still further.

In this report, our goals were (1) to statistically combine baseline MRI measures of hippocampal, temporal and ventricular volumes with age, sex, ApoE genotype, and body mass index (BMI) for AD and MCI classification; (2) to examine how the best-performing predictors would be further enhanced by using information on CSF biomarkers and PET-FDG; (3) to evaluate this multimodality approach for predicting cognitive decline in MCI and, most importantly; (4) to assess whether we could expect to reduce clinical trial sample size estimates by using our classifiers to target those most likely to decline. Numerous structural MRI-based measures, including hippocampal and ventricular volumes, as well as other temporal lobe summaries, have already been validated as indicators of AD progression, particularly after the MCI stage (Frisoni et al., 2010). We hypothesized that using multiple MRI summaries (rather than choosing one) might offer complementary information to classify patients into the correct diagnostic categories and predict cognitive decline, thereby providing a new way to boost the power of clinical trials.

1. Methods

1.1. Subjects

Baseline neuroimaging and biomarker data were downloaded from the ADNI public database (www.loni.ucla.edu/ADNI/Data/) on or before 20 November 2009 and reflect the status of the database at that point; as data collection is

Author's personal copy

Table 1
ADNI subjects and biomarkers included in each study Here we outline the subject samples analyzed for different classification tests

Study	Biomarkers	Number of subjects (training + testing)		
		AD	MCI	CN
1	MRI, age, ApoE, sex, BMI	158 (118 + 40)	264 (184 + 80)	213 (163 + 50)
2a	MRI, age, ApoE, CSF	77 (57 + 20)	158 (118 + 40)	93 (68 + 25)
2b	MRI, age, ApoE, PET-FDG	79 (59 + 20)	191 (146 + 45)	94 (74 + 20)
2c	MRI, age, ApoE, CSF, PET-FDG	40 (20 + 20)	83 (43 + 40)	43 (23 + 20)
3a	MRI, age, ApoE, CSF, PET-FDG	—	64 (41 + 23)	—
3b	MRI, ApoE, PET-FDG	—	129 (67 + 62)	—

Here we outline the subject samples analyzed for different classification tests. Subjects are split into independent training and testing samples approximately in a 3 :1 ratio, except for the smaller studies, to ensure correct evaluation of classifier performance. The 3 : 1 ratio is used in several machine learning studies such as Vemuri et al. (2008a). MRI denotes that a 1.5 T MRI scan was available; BMI denotes body mass index. CSF denotes that CSF-derived biomarkers were available.

ongoing. ADNI is a large 5 year study launched in 2004 with the primary goal of testing whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments at multiple sites (as in a typical clinical trial), can replicate results from smaller single site studies measuring the progression of MCI and early AD. More sensitive and specific markers of early AD progression will help monitor the effectiveness of new treatments, and lessen the time and cost of clinical trials.

1.1.1. Available data for baseline subjects

In what follows, sample sizes for analyses using different predictors are slightly different, as the study is ongoing and not all measures could be collected from all ADNI subjects. For our classification study based on baseline MRI numerical summaries, ApoE, age, sex, and BMI, data were available from 737 ADNI subjects (158 AD: 75.4 ± 7.4 years of age, 366 MCI: 74.8 ± 7.3 years of age, and 213 controls: 76.0 ± 5.1 years of age). To equalize the sex distribution, we reduced the MCI subject set to a group of 264 sex-matched subjects. As there were 102 more men than women in the MCI group, we ranked the MCI males based on numbers assigned to them via a computerized random number generator and removed the first 102 to ensure that the elimination process was random and unbiased. For our next classification study, we were limited by the availability of PET-FDG and CSF data, so our studies included subsets of the subjects considered above (328 subjects after adding only CSF, 364 subjects after adding only PET-FDG, and 166 subjects after adding both CSF and PET-FDG). For the first part of the cognitive decline prediction study, we considered 64 sex-matched MCI subjects, of whom 12 converted to AD in 12 months. Sixty-four subjects remained after selecting MCI subjects who had all biomarker information and equalizing the distribution of sex. The fraction of converters here (18.75%) is a slightly higher than the previously estimated rate of conversion in ADNI (13% according to Petersen, 2010); the rate is marginally higher as a subgroup of male nonconverters was excluded to allow sex matching. A larger sample of 129 sex matched MCI subjects with a reduced number of biomarkers was considered for the second part of the same study, 22 of whom (17.05%) converted to AD within 12 months. Sex matching was performed through a random elimination process as described above. The subjects and biomarkers included in each study are summarized in Table 1.

1.2. Biomarkers

For each subject, the biomarkers we considered included three MRI-derived numerical summaries, a PET-FDG numerical summary, and three CSF biomarkers (t-tau, p-tau and Aβ_{42}). In addition to MRI, PET-FDG and CSF can provide important functional and pathological information on AD progression (Jack et al., 2010). We also considered ApoE genotype (coded as 0, 1, or 2 for the number of ϵ4 alleles), age, sex and BMI, as each can influence AD risk (Azad et al., 2007; Buchman et al., 2005; Corder et al., 1993; Lindsay et al., 2002). BMI was included as several recent studies found that higher BMI is associated with greater brain atrophy in normal elderly subjects (Raji et al., 2010), and in MCI and AD (Ho et al., 2010b). This effect still holds true after accounting for the effects of hypertension, diabetes, and the level of white matter hyperintensities (Ho et al., 2010b) on the brain. In addition, a commonly carried risk gene for obesity, FTO, was recently reported to be associated with the level of brain atrophy in the ADNI cohort (Ho et al., 2010a), so we included BMI as it is a cardiovascular risk factor associated with brain atrophy. Clinical biomarkers that were used in ADNI to determine diagnosis, such as the sum-of-boxes Clinical Dementia Rating (CDR-SB) and other similar measures are used by physicians for making diagnoses and were therefore not used as features for classification to avoid circular inference. In fact, using CDR-SB alone for classification led to almost perfect classification accuracy, as accuracy here is judged in terms of agreement with clinical diagnosis, the best available proxy when *post mortem* neuropathological data are not yet available. Instead, the annual rate of change in CDR-SB was used as an outcome measure of cognitive decline to help define conversion from MCI to AD.

The MRI features included numerical summaries from the hippocampus, lateral ventricles and a tensor-based morphometry (TBM)-derived measure of atrophy in the tempo-

Author's personal copy

ral lobes. The hippocampal summaries were volumes generated from an automatic segmentation method that we developed based on machine learning; we recently validated this method against manual gold standards (Morra et al., 2008; Morra et al., 2009; Morra et al., 2010). The ventricular summaries were volumes acquired from a semiautomated, multiatlas segmentation technique that we developed (multiatlas fluid image alignment or MAFIA; (Chou et al., 2008)). The temporal lobe summaries were obtained from an anatomically defined region-of-interest (ROI) on three-dimensional atrophy maps generated with tensor-based morphometry (Hua et al., 2008a; Hua et al., 2008b). PET-FDG numerical summaries were based on a predefined temporal lobe ROI (Landau et al., 2009). All imaging summaries were averaged for the lobes in the left and right brain hemispheres.

CSF samples were obtained through lumbar puncture, after an overnight fast. Samples from various sites were transferred, on dry ice, to the ADNI Biomarker Core Laboratory at the University of Pennsylvania Medical Center, where the levels of t-tau, p-tau and Aβ_{42} are measured with a multiplex immunoassay platform under the direction of Drs Leslie Shaw and John Trojanowski. ApoE genotyping was performed on DNA samples from subjects' blood. Genomic DNA samples were analyzed using the Human610-Quad BeadChip (Illumina, Inc, San Diego, CA) at the University of Pennsylvania. Demographic data were obtained from https://www.loni.ucla.edu/ADNI/Data/. It should be emphasized that only baseline values of the biomarkers were used for prediction.

1.3. Support vector machines

SVMs are a type of machine learning or pattern recognition method that can be used to classify a dataset into different groups, based on multiple features, or measures, available for each subject (see, e.g. Morra et al., 2009b). As with linear discriminant analysis, some observations about a subject (here the imaging and other measures) may be assembled into a vector, with as many components as there are measures. Then a mathematical function is estimated (or "learned") that best combines these features to give an output that indicates, as accurately as possible, which group the individual belongs to. For an introduction to SVM − comparing it to simpler methods such as linear discriminant analysis (LDA) − please see our tutorial (Morra et al., 2009b). As mentioned in the introduction, SVM was chosen as a machine learning algorithm for this report due to its successful performance in the previous AD literature (Davatzikos et al., 2009; Fan et al., 2008; Mesrob et al., 2008; Vemuri et al., 2008), and for other neurobiological applications (Ecker et al., 2010; Koutsouleris et al., 2009; Wilson et al., 2009). SVMs may be considered as generalizations of linear regression, which use a supervised learning method to fit a classification function to the data in a training set of labeled observations. Other types of classifiers, such as adaptive boosting (Freund and Schapire, 1999; Morra et al., 2010), may also be useful for subject classification based on multiple biomarker measures, as they optimally combine predictors that perform weakly individually, but strongly in combination.

SVM is formulated as an optimization problem. Given a set of training data with corresponding class labels, a hyperplane is sought that maximizes the margin (a measure of the ability to differentiate) between different classes. This hyperplane, computed from a training set of example data, can then be used to classify newly presented (independent) testing datasets. Data consist of a set of vectors ($x_1 \ldots x_n$) where each vector contains several features and the class labels are scalars ($y_1 \ldots y_n$) where y_i is either 1 or −1 in a 2-class problem. The optimization problem for a linear SVM is written as:

$$min\frac{1}{2}\|w\|^2 \ subject \ to \ y_i\left(x_i \cdot w + b\right) \geq 1,$$

where w and b represent the normal vector to and the intercept of the hyperplane respectively. For cases where a linear surface (hyperplane) cannot effectively separate the data, nonlinear kernels, such as radial basis functions (RBFs), are incorporated into the optimization problem. Additionally, "slack variables" may be introduced with a tunable parameter, C, to allow a balance between misclassifications and the width of the margin. With this modification, the optimization problem may be restated as:

$$min\frac{1}{2}\|w\|^2 + C\sum_i \zeta_i \ subjected \ to \ y_i\left(x_i \cdot w + b\right) \geq 1 - \xi_i \ ,$$

where ξ_i is the slack variable for each i (Burges, 1998; Vapnik, 1995). SVMs may also be used for regression, where instead of a binary output, it would predict a continuous output for each subject's input vector, x. We performed our experiments using the LS-SVM package for classification and regression (Suykens and Vandewalle, 1999) in Matlab (MathWorks, Natick, MA).

1.4. Training and testing

We divided AD, MCI, and control subjects randomly into training and testing sets as shown in Table 1. The training sets were used for parameter optimization (regularization parameter C for a linear kernel; C and kernel-specific parameter, σ, for an RBF kernel) and for leave-one-out cross-validation. The SVM models were tested on independent testing sets to ensure generalizability. Receiver operating characteristic (ROC) curves were obtained to demonstrate the trade-off between sensitivity and specificity. ROC curves were compared, to evaluate different classifiers, using a statistical method developed for ROC analysis (Hanley and McNeil, 1983) in the MedCalc Statistical Software (MedCalc, Mariakerke, Belgium). When SVM was implemented for prediction instead of classification, mean

Author's personal copy

squared errors were used for comparison, instead of misclassification errors.

1.5. Power analysis

A power analysis was defined by the ADNI biostatistics Core to estimate the sample size required to detect a 25% reduction in the mean annual rate of atrophy, using a two-sided test and standard significance level (α = 0.05) for a hypothetical two-arm study (treatment v. placebo), with 80% power (this number is referred to as n80, and smaller numbers are better). The formula is:

$$n = \frac{2\sigma^2\left(z_{1-\alpha/2} + z_{power}\right)^2}{\left(0.25\beta\right)^2},$$

where σ and β refer to the mean and standard deviation in the atrophic rates respectively, α is set to be 0.05, and the desired power is 80%. Atrophic rates were determined based on a statistically defined ROI by training on 22 AD subjects, as described more fully in (Hua et al., 2009). Brain atrophy rates measured by MRI correlate with the progression of Alzheimer's disease, and offer baseline and transitional predictive power for diagnosis, making them clinically relevant endpoints for power analysis (Duara et al., 2008; Fox et al., 2000; Jack et al., 2004).

2. Results

2.1. AD and MCI classification based on MRI markers, ApoE genotype and demographic information

We first used the three MRI-derived summaries, ApoE genotype and demographic variables (age, sex and BMI) for AD and MCI classification with 635 ADNI subjects. SVM training was performed with all seven features using a linear kernel with $C = 1$, and the contributions of the different biomarkers were put into a rank order (best to worst) based on their SVM weights, assessed by w_i^2 in the notation of SVM described in the methods. The rank orders are shown in Table 2.

We then aimed to find the top N (N ranging from 1 to 7) features that yielded the highest leave-one-out accuracy in the training set, using an RBF kernel with parameter optimization. Both linear and RBF kernels identified the same set of top features, but the RBF kernel gave better performance, so we only present those results here. For AD vs. control, the best combination included the top four features (baseline hippocampal and ventricular volumes, as well as ApoE and age); this joint classifier yielded a leave-one-out accuracy of 82.21% correct classification, with a corresponding area under the ROC curve (AUC) of 0.945, which is relatively high. For classifying MCI vs. control, the best feature combination consisted of the top 3 (baseline hippocampal volume, ApoE and age), which gave 70.89% accurate classification, with a corresponding area under the ROC curve of 0.860. As expected, MCI classification accuracy was slightly poorer than AD classification, as there is substantial overlap on all known measures, between MCI and normal aging. The best biomarker sets for each classification are highlighted in Table 2. Figure 1 shows the ROC curves. In Table 2, only a subset of features was actually used: the best classifiers did not include BMI, sex, and the TBM-derived numeric summary. Also in Table 2, it is interesting that ventricular volume was helpful for the AD classification problem but not for distinguishing MCI from controls. This is reasonable given past findings by ourselves and others that ventricular expansion in MCI is relatively mild; there is also substantial cross-subject variation in ventricular volume, even in healthy subjects (Chou et al., 2009b), and this may throw off a classifier's accuracy unless the disease effect outweighs this natural variation (Chou et al., 2008; Chou et al., 2009a; Chou et al., 2009b).

In our next study, our goal was to compare the predictive power from the best combination of features obtained above, which included MRI, ApoE, and age, to that obtained when also including the PET-FDG temporal summary and CSF biomarkers. This may seem like an artificial distinction between two lists of biomarkers, but, from a practical point of view, the first classifier could be applied to a study that only used MRI, while the extended classifier would also need PET scans and lumbar puncture to be

Table 2
Rank order list with relative SVM weights for MRI, ApoE, Age, Sex, and BMI in AD and MCI classification

Rank	Biomarker	
	AD v. control (Weight/W_i^2)	MCI v. control (Weight/W_i^2)
1	MRI hip[a] 0.1664	MRI hip 0.1045
2	ApoE 0.1063	ApoE 0.0938
3	Age 0.0369	Age 0.0188
4	MRI Vent[b] 0.0349	MRI Vent 0.0103
5	MRI Temp[c] 0.0210	MRI Temp 0.0045
6	BMI 0.0147	BMI 0.0019
7	Sex 0.0013	Sex 0.0009

Hippocampal volumes were the most influential feature for differentiating AD from controls, closely followed by ApoE genotype, which outperformed all the other MRI-derived markers. For classifying subjects as either MCI or controls, the exact same features were useful, in the same order of priority. This is somewhat in line with expectation, as hippocampal volume is so widely used and is perhaps the most well-validated MRI measure in AD studies. This rank order refers to a situation in which all measures are used jointly for classification. Also, the gray highlighted measures are the ones that, when used jointly, gave the best classification accuracy in our independent test datasets (see Figure 1 for ROC curves).

[a] Hippocampal volume summary.

[b] Ventricular volume summary.

[c] Temporal lobe summary from TBM.

Author's personal copy

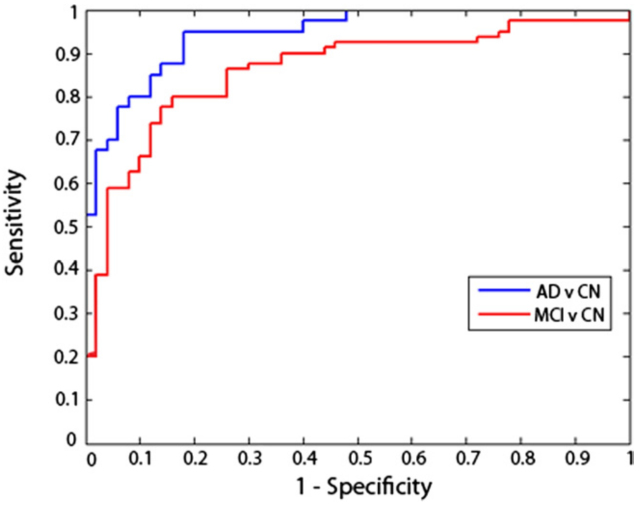


Fig. 1. ROC curves for AD and MCI classification. These curves show the trade-off between specificity and sensitivity for classifiers that best distinguished MCI from controls (*red curve*) and AD from controls (*blue curve*). The AD classifier used four measures and the MCI classifier only used three. These evaluations are based on finding the top set of features that yielded the highest leave-one-out accuracies on the training set. The curves gradually rise, meaning that there is a natural trade-off: the parameters of the classifier's decision boundary can be adjusted to be stricter or more lenient. For stricter classification settings, false positive classifications will decrease but so will the rate of true positives. Curves are slightly jagged and not perfectly smooth as they are based on a finite set of test data; with more data, they would be smoother.

performed. Although using more features is almost certainly better statistically, we wanted to assess how much difference it made, given the added expense, logistics, and possible attrition effects of performing multiple assessments.

Here, we considered three subsets of the ADNI subjects ($N = 328$ when adding CSF alone, $N = 364$ when adding PET-FDG alone, and $N = 166$ when adding both CSF and PET-FDG) of the ADNI subjects, for whom the data from these additional diagnostic modalities were available. We applied the same ranking algorithm based on SVM weights and obtained rank orders for the biomarkers, with CSF and PET-FDG taken into account. We found the top set of biomarkers yielding the highest leave-one-out accuracies on the training set for each classification. The rank orders and best sets of biomarkers are displayed in Table 3. CSF t-tau and $A\beta_{42}$ were included in the best set of biomarkers for both AD and MCI classification. PET-FDG also contributed substantially to AD and MCI classification. The remaining top biomarkers were essentially the same as the ones identified in the above study.

It may seem paradoxical that when we list the biomarkers in order of priority (Table 3) some of them are listed even though they are not ultimately used in the best-performing classifier (only the lists of features in gray are used in the best classifier). The reason this occurs is that when all features are included, some features are given nonzero weights, which means that they are useful for classifying the training set. Even so, these features may give no detectable improvement in classifying the test set, so they were dropped from the final classifier. This does not mean that they are not useful predictors under any circumstances; it just means that in this sample, they did not improve classification accuracy on the independent evaluation data.

We then compared the performance of AD and MCI classifiers trained with the top biomarkers from the previous ($N = 635$) study to those trained with the top biomarkers that included either CSF or the PET-FDG temporal summary or with both combined. Comparison of leave-one-out accuracies on the training set improved classification, implying that PET-FDG and CSF provide complementary information to MRI, ApoE and age. Leave-one-out accuracies for AD vs. control improved by 6.4%, 3.8%, and 11.6% by adding CSF alone, PET alone and both CSF and PET respectively. The corresponding improvements for MCI vs. control were 2.3%, 2.7%, and 4.6%.

When we compared the ROC curve AUCs, however, the improvement obtained by adding CSF, PET-FDG or both measures to the MRI measures was not statistically significant (p values > 0.05; Table 4). This lack of statistical significance may be due to the small size of the testing sets. If, however, this lack of significance is verified in even larger studies, it could have considerable implications for clinical trials in terms of total cost, efficiency and adverse effects.

2.2. *Boosting power for clinical trials*

A novel use of classifiers is to identify subjects who are more likely to decline. Under some reasonable assumptions (see Discussion), this can lead to larger effect sizes for detecting changes in biomarkers over time; this may also be useful for reducing sample size requirements for clinical trials of potential disease-modifying therapies. In the past, several authors have suggested that people in the lowest 50% (or some other quantile) of hippocampal volume are more likely to show future decline, both clinically (e.g., conversion from MCI to AD) and on imaging (see, e.g., Frisoni et al., 2010). Of course, this idea could be generalized to defining a sample based on the *k*% of subjects that a classifier declares as most likely to decline clinically in the future. Such a classifier could include not just MRI but any biomarker relevant for improving prediction.

As such, we computed minimum sample size estimates (n80) for the top *k* percent of subjects (for different values of *k* noted below) classified as *most likely to have AD* with our best AD classifier, using MRI hippocampal and ventricular summaries, ApoE and age as features. This *k*% of people are subjects in the independent test datasets (not used to train the classifier) who are assigned by the classifier to the AD class; they are those classified as AD who are farthest from the "SVM classifier decision boundary". We did not include PET-FDG and CSF biomarkers here, since adding these covariates limited our sample size and, as

Author's personal copy

Table 3
Rank order list with relative SVM weights for MRI, ApoE, Age, Sex, BMI, and either (a) CSF or (b) PET-FDG, for AD and MCI classification

Rank	Biomarker			
	A. MRI + CSF		B. MRI + PET-FDG	
	AD v. control (Weight/W_i^2)	MCI v. control (Weight/W_i^2)	AD v. control (Weight/W_i^2)	MCI v. control (Weight/W_i^2)
1	MRI hip[a] 0.0794	MRI hip 0.0519	ApoE 0.1529	ApoE 0.0929
2	CSF t-tau 0.0614	CSF Aβ_{42} 0.0313	PET-FDG 0.1022	MRI hip 0.0354
3	CSF Aβ_{42} 0.0505	Age 0.0308	MRI hip 0.0846	PET-FDG 0.0289
4	ApoE 0.0268	ApoE 0.0292	MRI Vent 0.0181	Age 0.0161
5	MRI Vent[b] 0.0238	CSF t-tau 0.0231	Age 0.0080	MRI Temp 0.0075
6	Age 0.0210	Sex 0.0157	MRI Temp 0.0057	Sex 0.0036
7	MRI Temp[c] 0.0163	MRI Temp 0.0085	BMI 0.0010	MRI Vent 0.0021
8	BMI 0.0077	BMI 0.0017	Sex 0.0004	BMI 0.0020
9	CSF p-tau 0.0003	CSF p-tau 0.0014		
10	Sex 0.0001	MRI Vent 0.0013		

Biomarkers are ranked according to their relative weights (contributions) in an SVM classifier that includes them all. A secondary question is which subset of these gives best classification accuracy, and this sublist is shown in gray. In these sublists, some features are omitted as adding them does not improve classification accuracy. Of the CSF markers, p-tau is relatively unhelpful but both t-tau and aβ_{42} provide independent predictive value. PET-FDG is a useful feature; whether it ranks above MRI hippocampal measures or not depends on whether the task is MCI or AD classification (hippocampal volume is slightly more useful than PET for MCI). PET measures are also somewhat correlated with MRI measures, so that when they are both included, each absorbs some of the variance; this may explain why ApoE genotype rises to the top of the predictors in terms of its independent contribution when MRI and PET are both included (*last two columns*).
Sets of biomarkers yielding the highest leave-one-out accuracy are highlighted.

[a] Hippocampal volume summary.

[b] Ventricular volume summary.

[c] Temporal lobe summary from TBM.

shown above, did not significantly improve classification in our tests. The subjects were ranked based on the SVM classifier output, the arithmetic sign of which determines the class assigned to each subject. A few AD subjects were excluded from the training and testing sets to avoid any overlap with the training set used in our prior report (Hua et

Table 4
Comparison of AD and MCI classification accuracy and false positive/false negative trade-offs (ROC analyses) for classifiers that use different types of information: MRI, MRI+ CSF, MRI+ PET-FDG, and MRI+ CSF+ PET-FDG

Biomarkers	AD v. control			MCI v. control		
	LOOCV accuracy	ROC AUC ± SE	Δ AUC[a] (*p* value)	LOOCV accuracy	ROC AUC ± SE	Δ AUC (*p* value)
Top MRI[b]	0.8160	0.8940 ± 0.0499	—	0.8421	0.8350 ± 0.0632	—
Top MRI+CSF[c]	0.8800	0.9560 ± 0.0273	0.0620 (0.191)	0.8647	0.8125 ± 0.0672	0.0225 (0.722)
Top MRI[b]	0.7634	0.7760 ± 0.0585	—	0.7227	0.7067 ± 0.0696	—
Top MRI+PET-FDG[d]	0.8011	0.7820 ± 0.0580	0.0060 (0.906)	0.7500	0.7444 ± 0.0672	0.0377 (0.382)
Top MRI[b]	0.7907	0.8850 ± 0.0501	—	0.7121	0.7488 ± 0.0649	—
Top MRI+CSF+PET-FDG[e]	0.9070	0.9175 ± 0.0413	0.0325 (0.357)	0.7576	0.7688 ± 0.0669	0.0200 (0.709)

Information for the top MRI classifier is listed twice, because MRI data were available for all ADNI subjects, but CSF and PET data were available only for a subset of those who had MRI. So it is only fair to report the classification accuracy on the full sample of MRIs, as well as on the subsamples in which head-to-head comparisons could be made with classifiers that also included the available CSF and PET data. The classifiers include ApoE and age, but not sex or BMI as the latter two did not contribute to the classification accuracies.
LOOCV: leave-one-out cross-validation.

[a] AUC difference relative to using the top MRI-based classifier only.

[b] Top biomarkers identified in the N = 635 study with MRI.

[c] Top biomarkers identified in the N = 328 study with MRI and CSF.

[d] Top biomarkers identified in the N = 364 study with MRI and PET-FDG.

[e] Top biomarkers identified in the N = 166 study with MRI, CSF and PET-FDG.

Author's personal copy

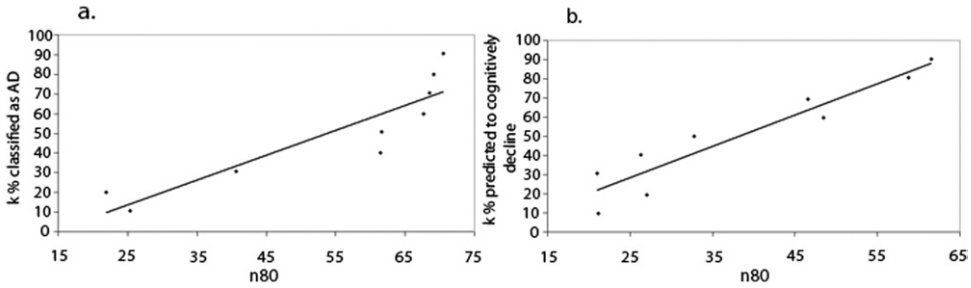


Fig. 2. n80 estimates (i.e., sample sizes required to detect a 25% slowing of the rate of atrophy with 80% power) as a function of restricting the sample to likely decliners. (a) Samples are based on the top *k*% classified, based on all biomarkers, as most likely to have AD (lower *k* gives smaller samples). (b) Here samples are based on the top *k*% of MCI subjects predicted by the classifier as most likely to decline (again lower values of *k* give dramatically lower samples). If only one-third of the most likely decliners were kept, in a subanalysis based on the classifier's predictions, then the sample size needed (n80) for an MCI trial would only be around 30 subjects per arm (see Discussion for caveats of this approach).

al., 2009) for creating the statistical ROIs. The results are shown in Figure 2a. When *k* is less than about 33%, the power estimates for AD subjects are improved compared with the minimal sample size of AD 48 subjects reported by Hua et al. (2009). There is a drop in the sample sizes needed to show a specific slowing effect, as the more AD-like subjects are selected. This has to be weighed against other factors (see Discussion), but it is interesting that the changes in these subjects have a greater effect size. It is also by no means obvious in advance that these subjects would give greater effect sizes. For the effect size to be greater, the changes have to be large and their variance has to be small; restricting the sample did not lead to an increase in the variability of the change measures sufficient to deplete effect sizes.

We could use the classifiers in many different ways to define a subsample – the *diagnostic* classifiers single out those who are most likely, based on all their imaging measures, to fall into a specific diagnostic category (e.g. AD). We also tested the benefit of defining a subsample of subjects with a classifier trained to identify likely decliners, based on all their imaging measures and other biomarkers, all at baseline.

To obtain similar n80 estimates for MCI subjects predicted to undergo cognitive decline, we considered 64 MCI subjects using MRI measures (three features), PET-FDG, CSF biomarkers (three features), ApoE and age. Here, the output of the SVM algorithm was set to be the 12-month rate of change in CDR-SB, instead of a binary output for the classification approach used in the studies above. Training with all possible 2^9-1 feature combinations using a linear kernel (parameter $C = 1$) revealed PET-FDG, MRI ventricular and temporal summaries, and ApoE as the best set of features, with the lowest mean squared error on the testing set.

To increase our sample size for evaluating this classifier, we considered a larger group of 129 MCI subjects with only the four features identified above and trained a model that predicted the rate of CDR-SB change in a novel testing set. We ranked the testing MCI subjects in order of predicted cognitive decline and computed n80 estimates for the top *k*% percent (for different values of *k*) of MCI subjects who the classifier predicted to be most likely to decline within a year (Figure 2b). The n80 values were even lower than the 88 MCI individuals we reported before as the minimal sample size for MCI (Hua et al., 2009). In addition to sample size estimates reported by Hua et al. (2009), similar estimates have also been made by other investigators such as Fox et al. (2000); Jack et al. (2004) and Schuff et al. (2009), and we were able to improve upon these too with our approach.

In this report, we have considered AD and MCI classification as well as prediction of MCI conversion. Classifiers can also be trained to distinguish MCI converters from diagnostic groups other than MCI. For instance, when we performed classification with a small group of 12 MCI converters versus 12 healthy controls using all features in our study, we obtained a reasonably promising 71% accuracy, as this discrimination is more challenging than separating AD patients from controls.

In general, however, we do not want to discriminate MCI decliners from groups other than MCI for the prediction of later decline. We assumed here that MCI diagnosis was given, and we aimed to predict who would decline within that group. Predicting decline in a mixed group of controls and MCIs is a little easier, as the knowledge that a person is MCI is already fairly good evidence that future decline is likely. Because of that, we wanted to assess the specific additive value of neuroimaging markers once a person is diagnosed as MCI (and it is reasonably helpful).

3. Discussion

We explored the power of several baseline biomarkers for AD and MCI, used jointly for diagnostic classification and for predicting future (1 year) cognitive decline in MCI. We also showed how to apply the multimodality classifiers to choose subsamples of subjects for boosting power in clinical trials. We determined combinations of regional MRI numerical summaries with demographic variables and

Author's personal copy

ApoE that best classified AD vs. control and MCI vs. control. The top set of complementary biomarkers for AD classification (when used together) were the MRI hippocampal volume summary (measured with the method of Morra et al., 2008), ApoE genotype, age and the MRI ventricular summary (measured with the method of Chou et al., 2009) in that order, resulting in an 82.21% accuracy, and a ROC AUC of 0.945, which is quite strong. Biologically, hippocampal atrophy and ventricular enlargement are established manifestations of AD pathology, and the two structures are routinely monitored via MRI for AD clinical trials (Frisoni et al., 2010). ApoE and advancing age are also well-known risk factors for AD (Carlsson et al., 2009), and age is associated with atrophic rates in ADNI (Hua et al., 2010a). The best set of features identified agrees with the AD literature. The one exception is the MRI temporal lobe summary, which did not improve classification power. This is not entirely surprising as it is quite highly correlated with the other two measures of atrophy (hippocampal and ventricular volume), so it may not add very much independent information for diagnostic classification. As expected, MCI classification was less accurate, and ventricular summaries were not as helpful; the best MCI diagnostic classifier only used hippocampal volume, ApoE genotype and age (Frisoni et al., 2010; Petersen, 2010).

When compared with accuracy results reported by groups such as Klöppel et al. (2008); Vemuri et al. (2008) and Fan et al. (2008), our accuracies may seem a bit low. Perhaps, the main reason the accuracy values are not so high is that we are using numerical summary measures (single values for each imaging modality) as opposed to voxel-wise maps (which are implemented in papers that report higher accuracies). Even so, it is difficult to compare the results across papers as different subject samples are used. For example, ADNI considers only AD patients with relatively mild AD, and classification of AD is clearly easier in cohorts with a greater proportion of more severely affected patients. Nonetheless, if some future classifier performs better, it could also be used to boost power using the same subpopulation selection method shown here.

By separately adding CSF biomarkers and PET-FDG as covariates for classification, where available, we obtained new rank order lists. These demonstrated how much the additional diagnostic measures contributed to AD and MCI classification, at least with this type of classifier. Different classes of AD biomarkers have dynamic trajectories that are thought to be temporally ordered with respect to the progression of the disease; in general, markers of amyloid deposition are thought to rise earlier than markers of neurodegeneration detectable on MRI, and these in turn become abnormal before tests of clinical function (Braskie et al., 2008; Jack et al., 2010; Petersen, 2010; Protas et al., 2010).

It is therefore plausible to expect classifiers to perform best with biomarkers that are maximally dynamic during the stages of disease being considered; measurement reproducibility and precision are important. The top feature lists are generally consistent with this hypothesis, as MRI contributes more strongly to AD classification, whereas PET-FDG and CSF biomarkers, particularly $A\beta_{42}$, play more important roles in MCI classification. The observation that CSF tau levels were more important for AD classification, and CSF $A\beta_{42}$ more contributory to MCI classification is also consistent with Jack et al.'s model, in which the dynamic range of $A\beta_{42}$ precedes that of tau in the progression of AD. ApoE is consistently included among the best biomarkers for both AD and MCI classification, which agrees with another component of the Jack et al. (2010) hypothesis, stating that carrying $\epsilon4$ alleles may shift the sequence of biomarker activities to earlier time points relative to the onset of overtly detectable clinical symptoms.

Predicting future decline in MCI subjects is more challenging than AD and MCI classification, as differences among MCI subjects are subtle. Instead of approaching this problem with a binary classifier, we adapted the algorithm to predict a continuous cognitive outcome, which is the 12-month change in CDR-SB. The baseline PET-FDG temporal summary, MRI temporal and ventricular summaries, and ApoE, were the best predictors of future cognitive decline in MCI (assessed over a 1 year follow-up interval). The combination of PET-FDG and ApoE genotype has been previously shown to provide good accuracy for predicting MCI conversion (Mosconi et al., 2004). MRI-based temporal and ventricular volumes have also been reported for their predictive power in MCI subjects (Fleisher et al., 2008; Korf et al., 2004). It is mechanistically reasonable for this combination of structural, functional and genetic information to supply complementary predictive power. By using a multimodality regression approach to predicting cognitive decline in ADNI subjects, a very recent study found that a linear combination of MRI and PET-FDG was a better predictor of cognitive decline than CSF biomarkers (Walhovd et al., 2010), consistent with our best set of biomarkers. Unexpectedly, however, the MRI hippocampal summaries were not incorporated into our predictive model, which is surprising as hippocampal volume can be useful for prediction of MCI progression to AD (Apostolova et al., 2006a; Apostolova et al., 2006b; Apostolova et al., 2007; Frisoni et al., 2010). The presence of detectable extra-hippocampal atrophy (e.g. in the ventricles and white matter) may also be good predictors of whether an MCI patient is deteriorating.

Our choice of brain regions and imaging measures to analyze was based on discussions among the ADNI Clinical, MRI and PET Cores. We chose imaging measures that had been used successfully in the past for disease classification or to monitor disease progression, preferring those measures that could be derived efficiently from a large dataset, without substantial manual interaction with the images. Clinical ratings were based on those widely used in clinical trials – CDR and mini-mental status examination

Author's personal copy

(MMSE) – and the CSF biomarker measures were those found to be most promising in pilot studies (Shaw et al., 2009). Needless to say, more brain regions or alternative cognitive tests could be proposed, and could be added to those analyzed here to boost performance even further. Specifically, in conference abstracts, Alexander et al. (2008) and Zhang et al. (2008) have advocated a *multivariate network analysis* in which a very large number of regional brain volumes are jointly used as predictors, in an SVM model. Other groups have parcellated the brain into a large number of subregions, but found that temporal lobe regions showed the greatest disease-related changes and significantly outperformed any of the clinical or cognitive measures examined for both AD and MCI (Holland et al., 2009). To single out brain regions that are most promising for analysis of disease-related brain change, we also focused on preselecting voxels in maps of brain change that show greatest effect sizes in independent samples. We and others have found that a classifier can be given an entire brain image, and from it can derive the voxels whose signals are most promising for group classification (Sun et al., 2009). By comparing different imaging measures (voxel-based, ROI-based, or surface-based; Gutman et al., 2008), and different classifiers (SVM v. others), future studies may be able to gauge which aspects of the classifier (its mathematical design or the features used) are most relevant for boosting performance.

In addition to scanning all the subjects with MRI at 1.5 T field strength, one quarter of ADNI's subjects also received 3-T scans. In prior work (Ho et al., 2009), we studied 110 ADNI subjects scanned longitudinally at both 3 and 1.5 T, across a 1-year interval. Our power analyses found that 37 AD and 108 MCI subjects would be needed at 1.5 T versus 49 AD and 166 MCI subjects at 3 T, to detect a 25% slowing of atrophy with 80% power, but these estimates did not differ significantly with field strengths. At both field strengths, temporal lobe atrophy rates were highly correlated with interval decline in Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog), MMSE and CDR-SB scores. To avoid modeling the effects of scanner field strength as a confound, here we used the 1.5 T ADNI data only. Some additional work may be needed to show that 3T scanners perform equally well for all biomarkers assessed here. The few ADNI studies that have assessed the field strength effect (Ho et al., 2009; Kruggel et al., 2010) suggest that 1.5 and 3 T scanners did not significantly differ in their power to detect neurodegenerative changes over a year.

Some clinical measures, such as CDR-SB, were not used as features for classification to avoid circular inference. Because these measures are used in making a diagnosis, it would be circular to incorporate them into our diagnostic classifiers and then test their empirical accuracy relative to the diagnosis given by physicians in the clinic. Even so, if used in practice to assist diagnosis, a classifier could use more cognitive measures – including those conventionally used for diagnosis and any other relevant information. However, the diagnostic accuracy of such a classifier could not then be "independently" validated in the same way as we did here. Doing so would require some other form of independent diagnostic ground truth, not used by the classifier, such as autopsy confirmation of characteristic signs of AD neuropathology. This could in principle be done, but neuropathology is not available in large numbers for the ADNI cohort.

A major clinical application of disease classifiers is for boosting power for clinical trials by reducing sample size estimates required to observe therapeutic effects. The idea of targeting a subgroup for analysis of treatment effects is not new (Frisoni et al., 2010). In fact, a drug trial for prodromal AD is currently recruiting subjects, with an inclusion criterion based on CSF Aβ_{42} and t-tau (clinicaltrials.gov/ct2/show/NCT00890890?term=bms+alzheimer%27s&rank=2). It appears new, however, to base the selection on a machine learning-based classifier that combines numerous biomarkers, which include neuroimaging measures. Combinations of disease markers are more likely to achieve sample size reductions than using single measures, such as subpopulation selection based on hippocampal volume only (of course statistical power must be traded off against the logistical complexity and cost of collecting and analyzing multiple biomarker assessments). When we considered the subset of subjects classified as most likely to have AD by our multifeature AD classifier, and the most likely decliners in MCI, we were able to reduce the n80 estimates to fewer than 40 subjects for both AD and MCI, improving on those estimates we reported before (Ho et al., 2009; Hua et al., 2009; Hua et al., 2010b). This result supports the concept of clinical trial enrichment, which has been previously advocated (Cummings et al., 2007; Frisoni et al., 2010; Hampel and Broich, 2009). Our enrichment strategy works because the subpopulation of subjects who are more likely to decline are selected based on disease classifiers and outcome predictors that integrate information from a number of complementary biomarkers.

We chose to compute sample sizes needed to detect a 25% slowing of atrophy with 80% power. While 25% is a reasonable target for a treatment that aims to slow atrophy, the exact number chosen is arbitrary. It is simple to compute sample size estimates for other percentage reductions in the atrophic rate, such as 5% or 50%, for example. As we noted in Hua et al. (2010b), treatments may slow atrophy to different degrees, which may be denoted by k%, for different k. The sample size estimates required to detect a k% slowing of atrophy may be easily derived by multiplying the sample size estimates (n80) in this paper by $(25/k)^2$, as the numbers follow an inverse-square law. For example, four times as many subjects would be needed to detect a 12.5% slowing of atrophy (half of 25%), versus a 25% slowing of atrophy (Ho et al., 2009). The quadratic relationship between the sample size estimates and the percentage atrophic rate is illustrated in (Hua et al., 2010b). Similarly, the results

Author's personal copy

of this paper can be easily translated to studies aiming to detect a different level of treatment effect, and our findings remain unaffected as multiplying the variables by a constant $(25/k)^2$ does not alter the ranking of the effect sizes in the statistical tests (it is a monotone transformation, i.e. it preserves the rank order).

As a caveat, the n80 "minimal sample size" measure is practical but has limitations: first, it is based on changes in the patient groups only, and not their difference from controls; second, it assumes that a treatment would slow atrophy in the same places as it normally occurs, with the same clinical outcome as observing an untreated sample with less atrophy. Finally, any treatment effects in a subanalysis might only apply to people who fit the selection criteria for that subanalysis; even so, evidence of an effect in a subanalysis might suffice to initiate a broader study.

The approach and results reported here are relevant to future work in the neuroimaging of AD in several ways. First, several authors advocate "enrichment" in clinical trials by trying to select those most likely to decline, based on clinical criteria, or occasionally based on imaging criteria. This can be done by applying thresholds or cut-offs to volumetric measures on MRI scans, such as hippocampal volume, but here we advocate using the full armory of imaging and CSF measures to classify subjects first, and then use the classifier's output to select subpopulations for later statistical testing.

Although this may seem like basing the statistical approach in part on the data collected, rather than specifying it all in advance of the study, this approach would identify subjects whose imaging data made them most likely to show treatment effects, regardless of the treatment. A similar approach to boost the power of imaging biomarkers is voxel-set preselection, which substantially boosts power to detect the slowing of atrophy (Chen et al., 2010; Hua et al., 2010).

For these statistically guided measures to be widely adopted as outcome measures in clinical trials, there needs to be some flexibility on the part of regulatory bodies that some features of the data collected may play a role in establishing which measures or subjects are evaluated. The analysis strategy can then adapt to the incoming data, and can exploit the power of Bayesian statistics and machine learning to obtain more powerful measures. It is quite defensible − and even advisable − for these machine learning approaches to be used, so long as the independence of statistical training and test samples is rigorously maintained.

A limitation of our study is that sample sizes become small when multiple imaging modalities and biomarkers are considered. In longitudinal studies especially, assessments of many kinds bring added costs, complexity, logistical difficulty, subject burden, and subject attrition (although in ADNI, attrition rates are only around 7% per year). Larger cohorts of subjects with available data from multiple biomarkers would allow more powerful classifiers and predictors to be developed, incorporating the best combinations of available diagnostic tools. More accurate ranking of biomarkers for verifying the details of Jack et al.'s temporal sequence hypothesis would become feasible. In addition, future studies will include additional diagnostic modalities such as Pittsburgh Compound B (PiB), diffusion tensor imaging (DTI), arterial spin labeling (ASL) and resting state functional MRI for disease classification. PiB has been collected in a small subsample of ADNI subjects, but we did not evaluate it here as requiring all biomarkers would have further limited our sample sizes. Another future direction would be to employ machine learning algorithms other than SVM (e.g. boosting; Morra et al., 2009b), or classifiers based on features in voxel-based maps (Sun et al., 2009), to improve classification and prediction accuracy. More powerful classifiers may then be implemented to improve upon our clinical trial boosting results. Furthermore, machine learning can perhaps be used to discover genetic (Stein et al., 2010a; Stein et al., 2010b), epidemiological and physiological factors that influence the progression of AD.

Disclosure statement

The authors have no potential financial or personal conflicts of interest including relationships with other people or organizations within 3 years of beginning the work submitted that could inappropriately influence their work.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's disease Neuroimaging Initiative (ADNI) (National Institutes of Health, Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly, and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer, Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., and Wyeth, as well as nonprofit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the US Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH Grants P30 AG010129, K01 AG030514, and the Dana Foundation.

Author's personal copy

References

Alexander, G.E., Hanson, K.D., Chen, K., Reiman, E.M., Bernstein, M.A., Kornak, J., Schuff, N.W., Fox, N.C., Thompson, P.M., Weiner, M.W., Jack, C.R., Jr., 2008. Six month MRI gray matter declines in Alzheimer's dementia evaluated by voxel-based morphometry with multivariate network analysis: Preliminary findings from the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement, 4, T273.

Alzheimer's Association, 2009. Alzheimer's disease facts and figures. Alzheimers Dement. 5, 234–270.

Apostolova, L.G., Dinov, I.D., Dutton, R.A., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M., 2006a. 3D comparison of hippocampal atrophy in amnestic mild cognitive impairment and Alzheimer's disease. Brain 129, 2867–2873.

Apostolova, L.G., Dutton, R.A., Dinov, I.D., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M., 2006b. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. Arch Neurol 63, 693–699.

Apostolova, L.G., Steiner, C.A., Akopyan, G.G., Dutton, R.A., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M., 2007. Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer disease. Arch Neurol 64, 1489–1495.

Azad, N.A., Al Bugami, M., Loy-English, I., 2007. Gender differences in dementia risk factors. Gend Med 4, 121–129.

Braskie, M.N., Klunder, A.D., Hayashi, K.M., Protas, H., Kepe, V., Miller, K.J., Huang, S.C., Barrio, J.R., Ercoli, L.M., Siddarth, P., Satyamurthy, N., Liu, J., Toga, A.W., Bookheimer, S.Y., Small, G.W., Thompson, P.M., 2008. Plaque and tangle imaging and cognition in normal aging and Alzheimer's disease. Neurobiol Aging Nov 10 [Epub ahead of print].

Buchman, A.S., Wilson, R.S., Bienias, J.L., Shah, R.C., Evans, D.A., Bennett, D.A., 2005. Change in body mass index and risk of incident Alzheimer disease. Neurology 65, 892–897.

Burges, C.J.C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Min Knowl Discov 2, 121–167.

Carlsson, C.M., Gleason, C.E., Puglielli, L., Asthana, S., 2009. Dementia including Alzheimer's disease. In: J.B. Halter, J.G. Ouslander, M.E. Tinetti, S. Studenski, K.P. High, S. Asthana, editors. Hazzard's Geriatric Medicine and Gerontology, 6th edition. Ch 65. The McGraw-Hill Companies. (Available at: www.accessmedicine.com/content.aspx?aID=5122625).

Chen, K., Langbaum, J.B., Fleisher, A.S., Ayutyanont, N., Reschke, C., Lee, W., Liu, X., Bandy, D., Alexander, G.E., Thompson, P.M., Foster, N.L., Harvey, D.J., de Leon, M.J., Koeppe, R.A., Jagust, W.J., Weiner, M.W., Reiman, E.M., 2010. Twelve-month metabolic declines in probable Alzheimer's disease and amnestic mild cognitive impairment Assessed using an empirically pre-defined statistical region-of-interest: Findings from the Alzheimer's Disease Neuroimaging Initiative. Neuroimage 51, 654–664.

Chou, Y.Y., Lepore, N., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Toga, A.W., Thompson, P.M., 2009a. Mapping correlations between ventricular expansion and CSF amyloid and tau biomarkers in 240 subjects with Alzheimer's disease, mild cognitive impairment and elderly controls. Neuroimage 46, 394–410.

Chou, Y.Y., Lepore, N., Chiang, M.C., Avedissian, C., Barysheva, M., McMahon, K.L., de Zubicaray, G.I., Meredith, M., Wright, M.J., Toga, A.W., Thompson, P.M., 2009b. Mapping genetic influences on ventricular structure in twins. Neuroimage 44, 1312–1323.

Chou, Y.Y., Lepore, N., de Zubicaray, G.I., Carmichael, O.T., Becker, J.T., Toga, A.W., Thompson, P.M., 2008. Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. Neuroimage 40, 615–630.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., Pericak-Vance, M.A., 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261, 921–923.

Cummings, J.L., Doody, R., Clark, C., 2007. Disease-modifying therapies for Alzheimer disease: challenges to early intervention. Neurology 69, 1622–1634.

Davatzikos, C., Resnick, S.M., Wu, X., Parmpi, P., Clark, C.M., 2008. Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 41, 1220–1227.

Davatzikos, C., Xu, F., Yang, A., Yong, F., Resnick, S.M., 2009. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain 132, 2026–2035.

Duara, R., Loewenstein, D.A., Potter, E., Appel, J., Greig, M.T., Urs, R., Shen, Q., Raj, A., Small, B., Barker, W., Schofield, E., Wu, Y., Potter, H., 2008. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. Neurology 71, 1986–1992.

Ecker, C., Rocha-Rego, V., Johnston, P., Mourao-Miranda, J., Marquand, A., Daly, E.M., Brammer, M.J., Murphy, C., Murphy, D.G., 2010. Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. Neuroimage 49, 44–56.

Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C., 2008. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage 41, 277–285.

Fleisher, A.S., Sun, S., Taylor, C., Ward, C.P., Gamst, A.C., Petersen, R.C., Jack, C.R., Jr, Aisen, P.S., Thal, L.J., 2008. Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment. Neurology 70, 191–199.

Fox, N.C., Cousens, S., Scahill, R., Harvey, R.J., Rossor, M.N., 2000. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease. Arch Neurol 57, 339–344.

Freund, Y., Schapire, R.E., 1999. Short introduction to boosting. J Jap Soc Artif Intell 14, 771–780.

Frisoni, G.B., Fox, N.C., Jack, C.R., Jr, Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nature Reviews. Neurology 6, 1–11.

Gutman, B., Wang, Y.L., Morra, J.H., Tu, Z., Jack, C.R., Weiner, M.W., Toga, A.W., Thompson, P.M., 2008. Disease Classification with Hippocampal Surface Invariants. MICCAI Workshop on Hippocampal Mapping.

Hampel, H., Broich, K., 2009. Enrichment of MCI and early Alzheimer's disease treatment trials using neurochemical and imaging candidate biomarkers. J Nutr Health Aging 13, 373–375.

Hanley, J.A., McNeil, B.J., 1983. Method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148, 839–843.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., 2009. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48, 138–149.

Ho, A.J., Stein, J.L., Hua, X., Lee, S., Hibar, D.P., Leow, A.D., Dinov, I.D., Toga, A.W., Saykin, A.J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M.J., Craig, D.W., Gerber, J.D., Allen, A., Corneveaux, J., Stephan, D.A., Webster, J., DeChairo, B.M., Potkin, S.G., Jack, C.R., Weiner, M.W., Raji, C.A., Lopez, O.L., Becker, J.T., Thompson, P.M., 2010a. A commonly carried allele of the obesity-related FTO gene is associated with reduced brain volume in the healthy elderly. Proc Natl Academy Sci U S A 107, 8404-8409.

Ho, A.J., Raji, C.A., Becker, J.T., Lopez, O.L., Kuller, L.H., Hua, X., Lee, S., Hibar, D., Dinov, I.D., Stein, J.L., Jack, C.R., Weiner, M.W., Toga, A. W., Thompson, P.M., 2010b. Obesity and brain structure in 700 MCI and AD patients. Neurobiology of Aging. In press [Accepted Apr 5, 2010].

Ho, A.J., Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Gutman, B., Dinov, I.D., Toga, A.W., Jack, C.R., Jr., Bernstein, M.A., Reiman, E.M., Harvey, D., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2009. Comparing 3T and 1.5T MRI for tracking AD progression with tensor-based morphometry. Human Brain Mapping. 31, 499 -514.

Author's personal copy

Holland, D., Brewer, J.B., Hagler, D.J., Fenema-Notestine, C., Dale, A.M., 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc Natl Acad Sci USA 106, 20954–20959.

Hua, X., Hibar, D.P., Lee, S., Toga, A.W., Jack, C.R., Jr., Weiner, M.W., Thompson, P.M., 2010a. Sex and age differences in atrophic rates: an ADNI study with N=1368 MRI scans. Neurobiology of Aging. In press [Accepted Apr 28, 2010].

Hua, X., Lee, S., Hibar, D.P., Yanovsky, I., Leow, A.D., Toga, A.W., Jack, C.R., Jr, Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2010b. Mapping Alzheimer's disease progression in 1309 MRI scans: power estimates for different inter-scan intervals. Neuroimage 51(1), 63–75.

Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Chou, Y.Y., Ho, A.J., Gutman, B., Toga, A.W., Jack, C.R., Jr, Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2009. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. Neuroimage 48, 668–681.

Hua, X., Leow, A.D., Lee, S., Klunder, A.D., Toga, A.W., Lepore, N., Chou, Y.Y., Brun, C., Chiang, M.C., Barysheva, M., Jack, C.R., Jr Bernstein, M.A., Britson, P.J., Ward, C.P., Whitwell, J.L., Borowski, B., Fleisher, A.S., Fox, N.C., Boyes, R.G., Barnes, J., Harvey, D., Kornak, J., Schuff, N., Boreta, L., Alexander, G.E., Weiner, M.W., Thompson, P.M., 2008a. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. Neuroimage 41, 19–34.

Hua, X., Leow, A.D., Parikshak, N., Lee, S., Chiang, M.C., Toga, A.W., Jack, C.R., Jr, Weiner, M.W., Thompson, P.M., 2008b. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects. Neuroimage 43, 458–469.

Jack, C.R., Jr, Knopman, D.S.K., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9, 119–128.

Jack, C.R., Jr, Shiung, M.M., Gunter, J.L., O'Brien, P.C., Weigand, S.D., Knopman, D.S., boeve, B.F., Ivnik, R.J., Smith, G.E., Cha, R.H., Tangalos, E.G., Petersen, R.C., 2004. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology 62,591–600.

Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Jr, Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689.

Klunk, W.E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D.P., Bergstrom, M., Savitcheva, I., Huang, G.F., Estrada, S., Ausen, B., Debnath, M.L., Barletta, J., Price, J.C., Sandell, J., Lopresti, B.J., Wall, A., Koivisto, P., Antoni, G., Mathis, C.A., Langstrom, B., 2004. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound B. Ann Neurol 55, 306–319.

Korf, E.S.C., Wahlund, L.O., Visser, P.J., Scheltens, P., 2004. Medial temporal lobe atrophy on MRI predicts dementia in patients with mild cognitive impairment. Neurology 63, 94–100.

Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Möller, H.J., Gaser, C., 2009. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch Gen Psychiatry 66, 700–712.

Kruggel, F., Turner, J., Muftuler, L.T., 2010. Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49, 2123–2133.

Landau, S.M., Harvey, D., Madison, C.M., Koeppe, R.A., Reiman, E.M., Foster, N.L., Weiner, M.W., Jagust, W.J., 2009. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol Aging Aug 4 [Epub ahead of print].

Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A.C., 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol Aging 29, 23–30.

Lindsay, J., Laurin, D., Verreault, R., Hébert, R., Helliwell, B., Hill, G.B., McDowell, I., 2002. Risk Factors for Alzheimer's Disease: A Prospective Analysis from the Canadian Study of Health and Aging. Am J Epidemiol 156, 445–453.

Lukas, L., Devos, A., Suykens, J.A.K., Vanhamme, L., Howe, F.A., Majós, C., Moreno-Torres, A., Van Der Graaf, M., Tate, A.R., Arús, C., Van Huffel, S., 2004. Brain tumor classification based on long echo proton MRS signals. Artif Intell Med 31, 73–89.

Mesrob, L., Magnin, B., Colliot, O., Sarazin, M., Hahn-Barma, V., Dubois, B., Gallinari, P., Lehéricy, S., Kinkingnéhun, S., Benali, H., 2008. Identification of Atrophy Patterns in Alzheimer's Disease Based on SVM Feature Selection and Anatomical Parcellation. Lecture Notes Comput Sci 5128, 124–132.

Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Toga, A.W., Jack, C.R., Jr, Weiner, M.W., Thompson, P.M., 2008. Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. Neuroimage 43, 59–68.

Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Toga, A.W., Jack, C.R., Jr, Schuff, N., Weiner, M.W., Thompson, P.M., 2009. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Neuroimage 45 suppl, S3–15.

Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Toga, A.W., Thompson, P.M., 2010. Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. IEEE Trans. Med. Imaging 29, 30–43.

Mosconi, L., Perani, D., Sorbi, S., Herholz, K., Nacmias, B., Holthoff, V., Salmon, E., Baron, J.-C., De Cristofaro, M.T.R., Padovani, A., Borroni, B., Franceschi, M., Bracco, L., Pupi, A., 2004. MCI conversion to dementia and the APOE genotype, A prediction study with FDG-PET. Neurology 63, 2332–2340.

Mourão-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. Neuroimage 28, 980–995.

Petersen, R.C., 2010. Alzheimer's disease: progress in prediction. Lancet Neurol. 9, 4–5.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303–308.

Protas, H.D., Huang, S.C., Kepe, V., Hayashi, K., Klunder, A., Braskie, M.N., Ercoli, L., Bookheimer, S., Thompson, P.M., Small, G.W., Barrio, J.R., 2010. FDDNP binding using MR derived cortical surface maps. Neuroimage 49, 240–248.

Raji, C.A., Ho, A.J., Parikshak, N., Becker, J.T., Lopez, O.L., Kuller, L.H., Hua, X., Leow, A.D., Toga, A.W., Thompson, P.M., 2010. Brain structure and obesity. Hum Brain Mapp 3, 353–364.

Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L.M., Trojanowski, J.Q., Thompson, P.M., Jack, C.R., Jr, Weiner, M.W., 2009. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132, 1067–1077.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M., Trojanowski, J.Q., 2009. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann. Neurol. 65, 403–413.

Stein, J.L., Hua, X., Lee, S., Ho, A.J., Leow, A.D., Toga, A.W., Saykin, A.J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M.J., Craig, D.W., Gerber, J.D., Allen, A., Corneveaux, J., Stephan, D.A., Webster,

Author's personal copy

J., deChairo, B.M., Potkin, S.G., Jack, C.R., Jr, Weiner, M.W., Thompson, P.M., 2010a. Voxelwise Genome-Wide Association Study: vGWAS. Neuroimage Feb 17 [Epub ahead of print].

Stein, J.L., Hua, X., Morra, J.H., Lee, S., Hibar, D.P., Ho, A.J., Leow, A.D., Toga, A.W., Sul, J.H., Kang, H., Eskin, E., Saykin, A.J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M.J., Craig, D.W., Gerber, J.D., Allen, A.N., Corneveaux, J.J., Stephan, D.A., Webster, J., deChairo, B.M., Potkin, S.G., Jack, C.R., Jr, Weiner, M.W., Thompson, P.M., 2010b. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. Neuroimage 51(2):542–554.

Sun, D., van Erp, T.G.M., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K., Toga, A.W., Cannon, T.D., 2009. Elucidating an MRI-based biomarker for psychosis: classification using probabilistic brain atlas and machine learning algorithms. Biol Psychiatry 66, 1055–1060.

Suykens, J.A.K., Vandewalle, J., 1999. Least Squares Support Vector Machine Classifiers. Neural Process Lett 9, 293–300.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R., Jr, 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 39, 1186–1197.

Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack, C.R., Jr, 2009. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology 73, 294–301.

Walhovd, K.B., Fjell, A.M., Brewer, J., McEvoy, L.K., Fennema-Notestine, C., Hagler, D.J., Jr, Jennings, R.G., Karow, D., Dale, A.M., 2010. Combining MR Imaging, Positron-Emission Tomography, and CSF Biomarkers in the Diagnosis and Prognosis of Alzheimer Disease. AJNR Am J Neuroradiol 31, 347–354.

Wilson, S.M., Ogar, J.M., Laluz, V., Growdon, M., Jang, J., Glenn, S., Miller, B.L., Weiner, M.W., Gorno-Tempini, M.L., 2009. Automated MRI-based classification of primary progressive aphasia variants. Neuroimage 47, 1558–1567.

Zhang, H., Wu, T., Bae, M., Reiman, E.M., Alexander, G.E., Jack, C.R., Jr, Thompson, P.M., Chen, K., 2008. Use of the Support Vector Machine and Sensitivity of an AD-Related Region-of-Interest Gray Matter Classifier in Identifying Amnestic MCI Subjects Who Convert t6o AD: Preliminary Findings From the AD Neuroimaging Initiative. Presented at the International Conference on Alzheimer's Disease, Chicago, July 26–31, 2008.