

Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment

Babak A. Ardekani^{a,b,*}, Elaine Bermudez^{a,b}, Asim M. Mubeen^a, Alvin H. Bachman^a and for the Alzheimer's Disease Neuroimaging Initiative¹

^a*The Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA*

^b*Department of Psychiatry, New York University School of Medicine, New York, NY, USA*

Handling Associate Editor: Ralph Buchert

Accepted 2 August 2016

Abstract.

Background: Mild cognitive impairment (MCI) is a transitional stage from normal aging to Alzheimer's disease (AD) dementia. It is extremely important to develop criteria that can be used to separate the MCI subjects at imminent risk of conversion to Alzheimer-type dementia from those who would remain stable. We have developed an automatic algorithm for computing a novel measure of hippocampal volumetric integrity (HVI) from structural MRI scans that may be useful for this purpose.

Objective: To determine the utility of HVI in classification between stable and progressive MCI patients using the Random Forest classification algorithm.

Methods: We used a 16-dimensional feature space including bilateral HVI obtained from baseline and one-year follow-up structural MRI, cognitive tests, and genetic and demographic information to train a Random Forest classifier in a sample of 164 MCI subjects categorized into two groups [progressive ($n = 86$) or stable ($n = 78$)] based on future conversion (or lack thereof) of their diagnosis to probable AD.

Results: The overall accuracy of classification was estimated to be 82.3% (86.0% sensitivity, 78.2% specificity). The accuracy in women (89.1%) was considerably higher than that in men (78.9%). The prediction accuracy achieved in women is the highest reported in any previous application of machine learning to AD diagnosis in MCI.

Conclusion: The method presented in this paper can be used to separate stable MCI patients from those who are at early stages of AD dementia with high accuracy. There may be stronger indicators of imminent AD dementia in women with MCI as compared to men.

Keywords: Alzheimer's disease, atrophy, hippocampus, longitudinal analysis, magnetic resonance imaging, mild cognitive impairment, prediction, Random Forest, sex

¹Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<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

*Correspondence to: Babak A. Ardekani, PhD, Nathan Kline Institute, 140 Old Orangeburg Rd, Orangeburg, NY 10962, USA. Tel.: +1 845 398 5490; Fax: +1 845 398 5472; E-mail: ardekani@nki.rfmh.org.

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most prevalent cause of dementia. Cognitive decline in AD starts many years before diagnosis and accelerates with disease progression [1]. Currently there are no available treatments for modifying the course of the disease, although several clinical trials are underway [2–4]. Mild cognitive impairment (MCI) can be a transitional stage from asymptomatic preclinical AD to clinical dementia due to AD. Incidence of progression to AD in individuals with MCI is estimated to be between 10–15% per year [5]. But not all MCI subjects develop AD; some may remain cognitively stable for many years and others may experience improvement in cognition [6]. It is extremely important to develop criteria that can be used to separate the MCI subjects at imminent risk of conversion to an AD diagnosis from those who will remain stable so that disease modifying treatments, when they become available, can be applied as early as possible and also for reducing the cost of future clinical trials by targeting individuals with MCI with AD as the underlying cause.

In assessing the risk of conversion from MCI to AD multiple variables and factors can be considered including age, sex, ethnicity, family history, genetic information, brain imaging data, cerebrospinal fluid (CSF) biomarkers, clinical information, and cognitive test scores. In addition, the progressive nature of AD means that the rate of change of some of these variables with time can be informative for diagnosis. The complex interactions between these variables make it very difficult for humans to process this information for prediction purposes. For this reason, researchers have been developing advanced pattern recognition algorithms to teach computers to predict incipient Alzheimer-type dementia in subjects with MCI with varying degrees of success [5, 7–14]. A recent review of literature with regard to multivariate analysis and machine learning in AD research is given by Falahati et al. [15].

We have recently developed a novel measure of hippocampal volumetric integrity (HVI) based on structural MRI [16]. The HVI is an estimate of the fraction of brain parenchyma in a predefined adaptive hippocampal region of interest (ROI). Lower values of this normalized measure indicate higher hippocampal atrophy. The advantages of HVI are that it takes less than one minute to compute bilaterally and fully automatically, is obtained from raw

MRI volumes without requiring any preprocessing, and does not require correction for intracranial volume. Using HVI and its rate of change with respect to time as features in a linear support vector machine (SVM) algorithm achieved 97% accuracy in separating healthy controls from AD patients in an independent cohort [16]. The main objective of this paper is to determine the efficacy of HVI as a feature in the more clinically interesting problem of classification between MCI subjects who would remain stable (stable MCI or sMCI) and those who would progress to develop AD (progressive MCI or pMCI).

The rate of progression of neurodegeneration in AD can only be measured longitudinally and may enable more specific and sensitive characterizations of the disease. For example, there is evidence that the trajectory of cognitive performance may be a better indicator of an imminent AD diagnosis than the level of cognitive performance at a given time [17]. Despite this, the majority of previous approaches for predicting AD in MCI have been based on static measures, for example, structural MRI at baseline only [5, 7–14]. Therefore, in this paper we use longitudinal structural MRI and neuropsychiatric tests for sMCI versus pMCI classification.

The Random Forest machine learning algorithm [18] is a powerful technique that has not been as widely applied to AD diagnosis as other algorithms [15]. In a comparison of 10 supervised classification algorithms based on 8 different performance measures and application to 11 binary classification problems, the Random Forest method was a close second best performing method [19]. Recent applications of Random Forest classification to the problem of MCI-to-AD prediction have been promising [13, 14].

In the present study, we applied a Random Forest classification algorithm using longitudinal measurement of HVI, a novel index of hippocampal volumetric integrity derived from structural MRI, cognitive test scores, and demographic and genetic information in order to predict whether or not a patient with MCI is at imminent risk of incipient AD. Acronyms used in this paper are defined in Table 1.

MATERIALS AND METHODS

Study subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 as a

Table 1
Acronym definitions

Acronym	Definition
@Baseline	Suffix indicating a quantity measured at the baseline visit
@Followup	Suffix indicating a quantity measured at the followup visit
$\Delta x/\Delta t$	Average rate of change of quantity x with respect to time between baseline and followup visits (e.g., $\Delta\text{MMSE}/\Delta t$)
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's Disease Assessment Scale-cognitive subscale
ADAS11	ADAS-Cog 11-item scale
ADAS13	Modified ADAS-Cog 13-item scale
ADNI	Alzheimer's Disease Neuroimaging Initiative
APOE4	Apolipoprotein E $\epsilon 4$ allele
CDR	Clinical Dementia Rating
CDRSB	Clinical Dementia Rating Sum of Boxes
CSF	Cerebrospinal fluid
HVI	Hippocampal volumetric integrity
	LHVI: left hemisphere HVI
	RHVI: right hemisphere HVI
MCI	Mild cognitive impairment
	pMCI: progressive MCI subjects
	sMCI: stable MCI subjects
MMSE	The Mini Mental State Examination score
MRI	Magnetic Resonance Imaging
OOB	Out-of-bag
PET	Positron emission tomography

public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see <http://www.adni-info.org>.

The participants in this study were MCI subjects in the "Complete year 1 visits" ADNI-1 standardized dataset [20]. The subjects' supplementary data were accessed in October 2013. There are 311 MCI subjects in this cohort. To be classified as MCI in ADNI-1, subjects had Mini-Mental State Examination (MMSE) scores between 24–30 (inclusive), a memory complaint, objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a Clinical Dementia Rating (CDR) of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.

We further classified the MCI subjects into two groups. Those whose diagnoses were MCI at baseline and remained so for the duration of their observation period of at least 3 years were labeled as stable MCI (sMCI). Those whose diagnoses were MCI at base-

line and at one-year follow-up but were subsequently diagnosed as AD some time during their remaining observation period were labeled as progressive MCI (pMCI). At the time of their probable AD diagnosis, the pMCI subjects had MMSE scores between 20–26 (inclusive), CDR of 0.5 or 1.0, and met NINCDS-ADRDA criteria for probable AD. Based on these definitions, our cohort was reduced to 78 sMCI (54 men and 24 women) and 86 pMCI (55 men and 31 women) subjects.

Each of the 164 subjects underwent two structural MRI scans. One scan was conducted at baseline and a second follow-up scan at approximately one year later. In addition, the subjects underwent two neuropsychiatric test batteries, again one set of tests given near the time of the baseline MRI scan and repeated at approximately one year later near the time of the follow-up MRI scan. More information about the MRI scans and the neuropsychiatric test battery is given in the following two subsections.

MRI data

We downloaded baseline and nominal one-year follow-up 3D MRI volumes for each of the 164 subjects in our study (328 volumes in total). The volumes were raw (as acquired) with no pre-processing performed. The mean (\pm SD) baseline to follow-up interval was 1.00 (\pm 0.05) year. All MRI volumes were acquired on 1.5 Tesla scanners using T1-weighted magnetization prepared rapid gradient echo pulse sequences with matrix sizes $192 \times 192 \times 160$ – 170 or $256 \times 256 \times 166$ – 184 , in-plane voxel resolutions 0.94 to 1.25 mm, 1.2 mm slice thickness, repetition times 2300–2400 ms for multi-coil phased array head coils and 3000 ms for birdcage or volume coils, inversion time 1000 ms, and 8° flip angle. More details about the MRI acquisition protocol are given in [21].

Neuropsychiatric tests

We also downloaded results of longitudinal neuropsychiatric tests. A set of tests was given near the time of the baseline MRI scan and repeated approximately one year later near the time of the follow-up MRI scan. The test battery included the following: (1) The MMSE [22] ranging from 0 to 30 where score of 20 to 24 suggests mild dementia, 13 to 20 suggests moderate dementia, and less than 12 indicates severe dementia; (2) The Clinical Dementia Rating Sum of Boxes (CDRSB) [23, 24], the sum of

scores in each of six domains of functioning: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care, where the score in each domain ranges from 0 to 3. Thus, the CDRSB ranges from 0 (no impairment) to 18 (severe impairment in all six domains); (3) The 11-item Alzheimer’s Disease Assessment Scale-cognitive subscale (ADAS-Cog) [25] which ranges from 0 to 70 with higher scores reflecting greater cognitive impairment; and (4) The modified ADAS-Cog 13-item scale [26] which adds to ADAS-Cog a number cancellation task and a delayed free recall task, for a total of 85 points where higher scores indicate greater impairment.

Hippocampal volumetric integrity (HVI)

Details of the algorithm for HVI computation are given in [16]. Briefly, HVI is the fraction of the volume of a region that is expected to encompass the hippocampus in a normal brain that is occupied by tissue (rather than CSF). The fully automated, fast, reliable and robust process is based on 3D T1-weighted structural MRI and involves identification of the mid-sagittal plane [27] and the anterior and posterior commissures [28] on the MRI scan, from which a rigid-body transformation is performed to a standard orientation. Once in standard space, based on *a priori* training, 230 landmarks in the vicinity of the hippocampi are detected by template matching from which two (one for each hemisphere) 12-parameter affine transformations are computed. The composite (rigid-body + affine) transformations are applied to probabilistic left and right hippocampi labels determined based on manual tracings of hippocampi on scans from 65 normal subjects. Thus, a volume is determined (separately for each hemisphere) that is expected to encompass the hippocampus in a normal brain. Finally, an automated histogram analysis method using the expectation maximization (EM) algorithm is used to determine the partial fraction of this region that is occupied by brain tissue (rather than CSF). The ratio is termed the HVI. The histogram analysis method is illustrated in Fig. 1. The purpose of the histogram analysis is to determine a CSF intensity threshold I_{CSF} . For this purpose, a Gaussian mixture model with 5 terms is fitted to the histogram of the voxel intensities in the hippocampus ROI using the EM algorithm. In Fig. 1, the voxel intensities histogram is given by the jagged line where as the EM fit is given by the smooth thicker line. The $(1 - \alpha)$ th percentile value of the histogram is denoted by I_α

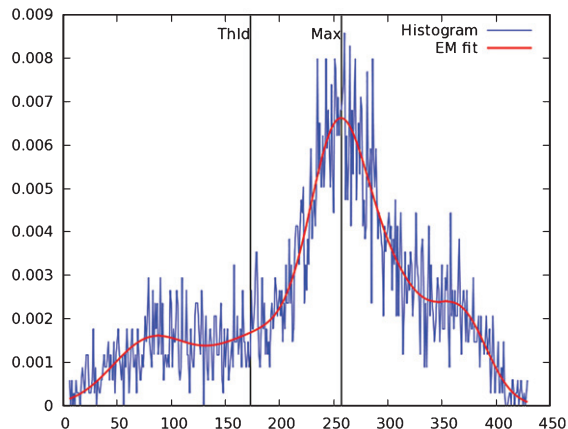


Fig. 1. Histogram analysis using the EM algorithm. The histogram of the voxel intensities comprising the left hippocampus ROI in a subject is shown (thin jagged line) along with a 5-component Gaussian mixture model fit using the EM algorithm (thick smooth line). Smooth fitting allows us to estimate the gray matter peak location I_{gm} (indicated by the “Max” line) from which we determine the location of a CSF intensity threshold I_{CSF} (indicated by the “Thld” line). HVI is defined as the fraction of the supra-threshold voxels, or equivalently, the area under the histogram for intensities above I_{CSF} .

(default $\alpha = 0.25$). A gray matter intensity peak I_{gm} is found as the peak of the EM fit of the histogram in the intensity region $[cI_\alpha, I_\alpha]$ (default $c = 0.4$); and I_{CSF} is defined as $I_{CSF} = I_{gm} - \gamma I_\alpha$ with default $\gamma = 0.2$. The HVI is essentially the area under the histogram curve for voxel intensities above I_{CSF} .

We computed the HVI for the right and left hippocampi on the baseline and followup MRI scans for all 164 subjects. Software (KAIBA) for computing the HVI is available online at www.nitrc.org/projects/art.

Feature space for machine learning

In machine learning, observations obtained from objects or individuals (in this application the MCI patients) to be classified into groups are referred to as features. For each individual, the features are collected into a vector referred to as the feature vector. The dimension of this vector, which we denote by d , equals to the number of observations collected from each subject. The d -dimensional feature vectors can be thought of as single points in a d -dimensional abstract space which is referred to as the *feature space*. The aim of machine learning algorithms is to define boundaries in the feature space, referred to as decision boundaries, that best separate the individuals into the defined classes (in this

Table 2

Mean \pm SD of the 16 features used in the Random Forest classification separated by stable versus progressive MCI groups

Feature	sMCI ($n = 78$)	pMCI ($n = 86$)	p -value [†]
1. Age (y)	74.75 \pm 6.94	74.20 \pm 7.29	0.515
2. Education	15.45 \pm 3.18	16.41 \pm 2.77	0.064
3. APOE4 (0/1/2)	(47/28/3)	(27/44/15)	$<10^{-3}$
4. Sex (M/F)	(54/24)	(55/31)	0.475
5. MMSE	27.76 \pm 1.82	26.49 \pm 1.68	$<10^{-6}$
6. Δ MMSE/ Δ t	0.32 \pm 2.29	-0.59 \pm 1.67	0.001
7. CDRSB	1.31 \pm 0.72	1.93 \pm 0.88	$<10^{-5}$
8. Δ CDRSB/ Δ t	0.27 \pm 0.85	0.57 \pm 0.92	0.018
9. ADAS11	9.41 \pm 4.10	12.93 \pm 3.88	$<10^{-7}$
10. Δ ADAS11/ Δ t	-0.19 \pm 4.16	1.30 \pm 4.00	0.016
11. ADAS13	15.36 \pm 5.85	20.80 \pm 5.10	$<10^{-8}$
12. Δ ADAS13/ Δ t	-0.27 \pm 5.14	1.91 \pm 4.74	0.002
13. LHVI	0.80 \pm 0.09	0.75 \pm 0.10	$<10^{-3}$
14. Δ LHVI/ Δ t	-0.01 \pm 0.02	-0.03 \pm 0.02	$<10^{-5}$
15. RHVI	0.79 \pm 0.09	0.72 \pm 0.10	$<10^{-3}$
16. Δ RHVI/ Δ t	-0.01 \pm 0.02	-0.03 \pm 0.02	$<10^{-5}$

[†]All statistical tests were performed using the Mann-Whitney U tests and the Chi-square tests for APOE4 and Sex distributions. SD, standard deviation; sMCI, stable mild cognitively impaired; pMCI, progressive mild cognitively impaired; APOE4, number of apolipoprotein E ϵ 4 alleles; MMSE, Mini-Mental State Examination; CDRSB, Clinical Dementia Rating Sum of Boxes; ADAS11, 11-item Alzheimer's Disease Assessment Scale cognitive subscale; ADAS13, 13-item Alzheimer's Disease Assessment Scale cognitive subscale; LHVI, Left Hippocampal Volumetric Integrity; RHVI, Right Hippocampal Volumetric Integrity; $\Delta x/\Delta t$, Average rate of change of quantity x during time interval Δt .

application sMCI or pMCI). In this work, we use the notation $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_d^{(i)}\}$ to describe the feature vector for a given subject i . In this section we summarize the d features that comprise our feature vectors.

In total, for classification purposes we used $d = 16$ features measured from each MCI subject as shown in Table 2. Features 1–4 (i.e., $x_1^{(i)}, \dots, x_4^{(i)}$) were subjects' age, years of education, number of apolipoprotein E ϵ 4 alleles (APOE4), and sex. The APOE4 was defined as a categorical variable with three levels (0, 1, and 2) that indicates the number of ϵ 4 alleles carried by the subject. Features 5 and 6 were the average MMSE [i.e., (MMSE@Baseline + MMSE@Followup)/2] and the average rate of change of MMSE with respect to time from baseline to followup, denoted by Δ MMSE/ Δ t and given by:

$$\frac{\Delta \text{MMSE}}{\Delta t} = \frac{\text{MMSE@Followup} - \text{MMSE@Baseline}}{\Delta t} \quad (1)$$

where Δt is the time interval in years between the baseline and followup measurements. Features 7 and 8 were the average CDRSB [i.e., (CDRSB@Baseline

+ CDRSB@Followup)/2] and its rate of change with respect to time, denoted by Δ CDRSB/ Δ t and computed similarly to Equation (1). Features 9 and 10 were the average 11-item ADAS-Cog (denoted by ADAS11) and its rate of change with respect to time (Δ ADAS11/ Δ t). Features 11 and 12 were the 13-item ADAS-Cog (denoted by ADAS13) and its time rate of change: Δ ADAS13/ Δ t. Features 13 and 14 were the average left HVI (denoted by LHVI) and its time rate of change: Δ LHVI/ Δ t. Finally, features 15 and 16 were the corresponding measures for the right hippocampus, that is: RHVI and Δ RHVI/ Δ t.

To summarize, from each of the 164 MCI subjects in the study, we obtained values for the 16 variables described in the previous paragraph. The variables age, years of education, APOE4 status, sex, MMSE, CDRSB, ADAS11, and ADAS13 were downloaded directly from ADNI. The baseline and follow-up values of the left and right HVI were computed by KAIBA using baseline and one-year follow-up MRI scans downloaded from ADNI. Our aim was to train a machine learning algorithm that uses these 16 measures, that is, the feature vector $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_d^{(i)}\}$ to predict whether an MCI subject would remain stable or would convert to AD. For this purpose, we used the Random Forest algorithm described in the following section.

Random Forest algorithm

In this paper, we utilize the Random Forest algorithm [18], implemented in the *randomForest* R package [29], as a supervised binary (sMCI versus pMCI) classification algorithm. In this context, a dataset $D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(n)}, y^{(n)})\}$ is given where n is the number of subjects, $y^{(i)} \in \{0, 1\}$ represents the subject label (e.g., 0 for sMCI and 1 for pMCI), and $x^{(i)}$ are d -dimensional feature vectors measured from each subject. In this study, the dimension of the feature space d is 16. In general, the features $x_j^{(i)}$ can be numerical variables (e.g., age) or categorical (e.g., sex or APOE4 genotype). A supervised classifier essentially uses the training data D to define a function $y = f(x; D)$ which would assign a label y to a subject with a feature vector x for whom the true label is unknown at the time of classification. The decision tree approach is a classical supervised method for defining $f(x; D)$ (i.e., training a classifier using D). However, a single decision tree classifier may have a large variance, that is, a small change in the feature vector x could change the assigned label y .

To overcome this instability, the Random Forest algorithm [18] uses the method of bootstrap aggregation (also known as bagging) where instead of a single decision tree, an ensemble of B trees ($B = 5000$ in our application) are trained, hence the term forest. In this method, the decision function of each tree, denoted by $y_b = f(x; D_b)$ ($b = 1, 2, \dots, B$), is trained using a bootstrap sample D_b from the complete dataset D (i.e., D_b is obtained by sampling from D with replacement), where the size of D_b is also n . Another element of the Random Forest method is that for training a decision tree at any given node of the tree, that is choosing a variable $x_j^{(i)}$ and a corresponding threshold value to guide the decision flow at that node, the entire feature space is not used. Rather, a random subspace of size $m < d$ of the feature space is first selected by sampling without replacement from $\{1, 2, \dots, d\}$ and then the node variable is selected from within this subspace for building the decision tree at the given node. The size of the random subspace (m) is a parameter of the algorithm for training the Random Forest classifier, with the default value of $m = \sqrt{d}$. In summary, the Random Forest method trains an ensemble of B decision trees using *bagging* and *random subspaces*. Given a feature vector x with an unknown label, classifications $\{f(x; D_1), f(x; D_2), \dots, f(x; D_B)\}$ are made by all B trees in the forest and a final label y is estimated by aggregating the results, usually by a majority vote.

Out-of-bag (OOB) estimation of classification accuracy

Since in the Random Forest method, each tree is built using a bootstrap sample D_b from the complete dataset D , on the average about 37% of the training samples are not used in the process of building any given tree. These are referred to as the out-of-bag (OOB) samples and can be used to estimate the generalization error of the Random Forest. Consider a given sample $(x^{(i)}, y^{(i)})$, bootstrap sampling causes this to be an OOB sample in approximately $0.37B$ trees. In OOB estimation of classification accuracy, the $0.37B$ trees for which the sample is considered OOB are used to classify the sample. Let us denote the OOB estimate of the label for the sample by $\hat{y}^{(i)}$. This is then compared with the known label of the sample, that is, $y^{(i)}$. Since every sample in the entire dataset D is OOB for a subset of trees of similar number, the process can be repeated for all n samples in D . Then by comparing the OOB estimates $\{\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(n)}\}$ with the known labels

$\{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}$ one can obtain the OOB estimate of the accuracy, which is simply the percentage of samples for which $\hat{y}^{(i)}$ match $y^{(i)}$. Empirical evidence [30] suggests that the OOB estimate of the true accuracy is as accurate as using an independent test set of size n for estimating the true accuracy. Therefore, using the OOB estimate removes the need for a separate test set.

Assessment of variables for classification

We used the mean reduction of Gini impurity index as a measure of variable importance for classification [30]. During training of a decision tree, whenever a variable is used to split a parent node into two descendent nodes, the sum of the Gini impurity indices of the descendent nodes is smaller than the Gini impurity index of the parent node. The greater the difference the better the variable has performed in splitting the cases into pure classes. The mean amount of reduction over all nodes in the forest where a given variable is used is taken an indication of the importance of that variable for classification.

RESULTS

Observation period

The mean \pm SD of the observation period for the sMCI group ($n = 78$) was 5.88 ± 2.48 years with a minimum of 3 years starting from the time of their baseline MRI scan. By definition, the diagnosis of these subjects remained MCI during this period. We imposed the minimum 3-year observation period requirement for the sMCI group to reduce sample noise, that is, to reduce the chance that some of the subjects included in the sMCI are actually progressive cases. The follow-up periods did not significantly differ between men (5.22 ± 2.68 y) and women (6.17 ± 2.35 y).

In the pMCI group ($n = 86$) the diagnosis changed from MCI to probable AD sometime during their observation period. In this group, the mean \pm SD time interval from their baseline MRI scan to probable AD diagnosis was 2.84 ± 1.50 years with a minimum of 1 year. The conversion times did not significantly differ between men (2.91 ± 1.47 y) and women (2.72 ± 1.57 y).

Feature vector variables

Table 2 shows comparisons between sMCI and pMCI groups separately in each of the 16 variables

comprising the feature space. Statistical comparisons for the categorical variables sex and APOE4 were made using the chi-square test. Significances of group differences in the remaining 14 variables were assessed using the Mann-Whitney U test.

The groups did not differ in age or sex distribution. There was a trend toward the pMCI group being marginally more educated ($p=0.064$) than the sMCI group. There were statistically significant mean differences in all the remaining 13 variables. The distribution of the APOE4 factor was significantly different between groups ($p < 10^{-3}$) indicating that a significantly greater proportion of pMCI subjects carried one or two copies of the apolipoprotein E $\epsilon 4$ allele. In all cognitive measures (MMSE, CDRSB, ADAS11, and ADAS13), the pMCI group had significantly poorer average scores and their scores deteriorated at significantly faster rates as compared to the sMCI group. Also, the HVI bilaterally were significantly lower and declined at significantly faster rates in the pMCI group as compared to the sMCI group.

Since we find sex differences in the accuracy of Random Forest classification with higher accuracy in women (see below), we further examined the 13 variables in Table 2 that were statistically significantly different between sMCI and pMCI groups to determine the effect sizes separately for men and women. The results are given in Table 3. We found that the effect sizes in 11 of the 13 variables were larger in women, which helps to better understand the higher classification accuracy in women compared to men.

Random Forest classification

We trained a Random Forest classifier with $B = 5000$ trees using the default value of $m = \sqrt{d} = 4$ for the dimension of the random subspaces used by the algorithm. To reiterate, the aim of the Random Forest classifier was to classify the MCI subjects into two groups: sMCI and pMCI, in other words, to predict conversion, or lack thereof, from MCI to AD based on the 16 measured/calculated quantities obtained at baseline and one-year follow-up from each MCI subject.

The results of OOB estimations, which is referred to as a confusion matrix, are summarized in Table 4. The estimated classification sensitivity $[TP/(TP+FN)]$ was 86.0%. The estimated classification specificity $[TN/(TN+FP)]$ was 78.2%. The estimated overall accuracy of the classifier $[(TN+TP)/(TN+FP+TP+FN)]$ was 82.3%. The

receiver operating characteristic (ROC) curve for the classifier is shown in Fig. 2 (solid line). The area under the curve (AUC) was 0.83. The performance of the classifier was considerably better for women (sensitivity: 93.6%; specificity: 83.3%; overall accuracy: 89.1%) than for men (sensitivity: 81.8%; specificity: 75.9%; overall accuracy: 78.9%). The relative importance of the 16 variables as measured by the mean reduction in the Gini impurity index when the variable is used as a decision variable at a tree node is shown in Fig. 3. According to this criterion, ADAS13 was the most important variable for classification followed by the rate of change with respect to time of the right HVI.

In order to assess the contribution of longitudinal measurements to classification performance, we retrained the Random Forest classifier after excluding the six features representing the rates of change of variables that could only be obtained longitudinally (Table 2: features 6, 8, 10, 12, 14, and 16). The performance of the classifier reduced considerably yielding an estimated 75.6% sensitivity, 69.2% specificity, and 72.6% overall accuracy. The ROC curve for this analysis is shown in Fig. 2 (dashed line). The AUC reduced to 0.77.

We also retrained the classifier by using only the four features related to the HVI (Table 2: features 13-16). In this case the overall classification accuracy reduced to 68.3% and the AUC of the ROC curve reduced to 0.74 (Fig. 2, dash-dot line). On the other hand, when we retrained the classifier by using all the non-HVI related features (Table 2: features 1-12), we obtained an overall classification accuracy of 71.9% and an AUC of 0.77 (Fig. 2, dotted line). In both cases, the classification performance was considerably lower than the 82.3% achieved using all features. The ability of classification algorithms such as the Random Forest to combine a number of weak learners to obtain a single strong learner is known as boosting.

DISCUSSION

Several biomarkers have shown prediction power for the early detection of AD. Among the most established and clinically validated ones are various neuropsychiatric measures of cognitive function [22-26], measures derived from structural MRI, particularly the hippocampus volume [31-33], CSF total tau (t-tau), tau phosphorylated at threonine 181 (p-tau₁₈₁) and A β ₁₋₄₂, positron emission tomo-

Table 3
Differences between sMCI and pMCI groups within each sex

Feature	Men (<i>n</i> = 109)				Women (<i>n</i> = 55)			
	sMCI (54)	pMCI (55)	<i>p</i>	<i>r</i>	sMCI (24)	pMCI (31)	<i>p</i>	<i>r</i>
APOE4 (0/1/2)	(31/20/3)	(18/28/9)	0.017	0.27 [†]	(16/8/0)	(9/16/6)	0.002	0.42 [†]
MMSE	27.75 ± 1.55	26.79 ± 1.65	0.002	0.29	27.79 ± 2.37	25.97 ± 1.62	0.000	0.51
ΔMMSE/Δt	0.50 ± 1.94	-0.55 ± 1.68	0.014	0.23	-0.09 ± 2.93	-0.67 ± 1.69	0.035	0.28
CDRSB	1.33 ± 0.75	1.90 ± 0.88	0.000	0.35	1.28 ± 0.65	1.98 ± 0.88	0.003	0.40
ΔCDRSB/Δt	0.19 ± 0.79	0.64 ± 0.87	0.002	0.30	0.47 ± 0.94	0.45 ± 1.02	0.830	0.03
ADAS11	9.41 ± 3.78	12.63 ± 3.95	0.000	0.37	9.40 ± 4.84	13.47 ± 3.76	0.000	0.50
ΔADAS11/Δt	-0.59 ± 3.51	0.84 ± 4.24	0.073	0.17	0.71 ± 5.34	2.12 ± 3.46	0.194	0.18
ADAS13	15.66 ± 5.60	21.14 ± 4.92	0.000	0.38	14.69 ± 6.44	21.96 ± 5.30	0.000	0.56
ΔADAS13/Δt	-0.75 ± 4.44	1.66 ± 4.94	0.005	0.27	0.80 ± 6.42	2.35 ± 4.39	0.245	0.16
LHVI	0.772 ± 0.083	0.717 ± 0.095	0.006	0.26	0.852 ± 0.077	0.797 ± 0.086	0.01	0.35
ΔLHVI/Δt	-0.013 ± 0.015	-0.027 ± 0.024	0.001	0.32	-0.015 ± 0.023	-0.027 ± 0.018	0.006	0.37
RHVI	0.765 ± 0.095	0.698 ± 0.096	0.001	0.33	0.832 ± 0.076	0.767 ± 0.098	0.009	0.35
ΔRHVI/Δt	-0.013 ± 0.015	-0.025 ± 0.022	0.005	0.27	-0.013 ± 0.014	-0.029 ± 0.020	0.000	0.53

Quantities are given as mean ± standard deviation, except for APOE4 for which the number of subjects who carry 0/1/2 copies of the apolipoprotein E ε4 allele is given. *r*: Effect size defined as Z/\sqrt{n} . [†]Cramer's *v*.

Table 4
OOB estimation confusion matrix

Prediction	Reference	
	sMCI	pMCI
sMCI	TN = 61 (41 males)	FN = 12 (10 males)
pMCI	FP = 17 (13 males)	TP = 74 (45 males)

TN, true negative; TP, true positive; FP, false positive; FN, false negative; sMCI, stable MCI; pMCI, progressive MCI.

graphy (PET) measures of Aβ plaques assessed by ¹¹C-PiB or ¹⁸F-AV-45 and measures of regional cerebral rates of glucose metabolism derived from ¹⁸F-FDG PET [34]. Measuring CSF biomarkers requires a spinal tap and PET imaging is more expensive and not as widely available as MRI. Training data for machine learning are also limited for CSF biomarkers and PET measures. Therefore, for pragmatic clinical utility, the current study was limited to measures based on longitudinal structural MRI, neuropsychiatric tests, APOE genotype, and simple demographic information for sMCI versus pMCI classification. In all, a 16-dimensional feature vector was used for classification.

We used a novel measure of HVI derived from structural MRI as a biomarker. HVI is computed fully automatically, fast (less than 1 minute), and reliably without requiring any pre-processing of the structural MRI images (e.g., distortion correction). Another advantage is that HVI does not require correction for intracranial volume for two main reasons. Firstly, HVI is a normalized measure (tissue fraction of the ROI). Therefore, if the scale of the ROI changed, the HVI value would remain the same, that is, HVI is scale-invariant. Secondly, before HVI computation, we apply a local 12-parameter affine transformation

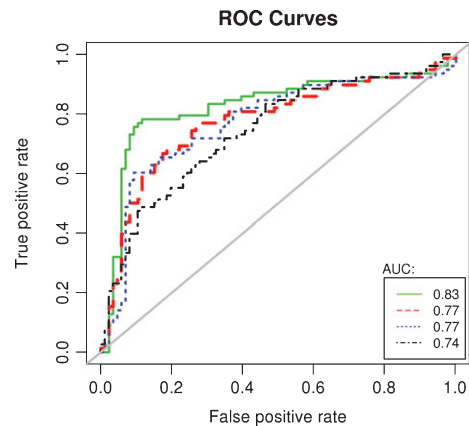


Fig. 2. Receiver operating characteristic (ROC) curves for the cases of using all 16 features (solid line, AUC 0.83); excluding longitudinal data (dashed line, AUC 0.77); excluding features related to HVI (dotted line, AUC 0.77); and including only HVI related features (dash-dot line, AUC 0.74).

computed based on landmark-detection to the MRI which to a large extent normalizes the size of the immediate medial temporal structures surrounding the hippocampus across individuals. Further details of the algorithm for HVI computation are given in [16].

The Random Forest classifier was able to predict conversion from MCI to AD dementia with estimated accuracy of 82.3%. Although it is difficult to precisely compare this result with previous studies because of the different datasets used to train and evaluate the classification algorithms and different methods used for estimating the prediction accuracy, a review of the literature indicates that the typical accuracy of the current algorithms for predicting MCI to AD conversion

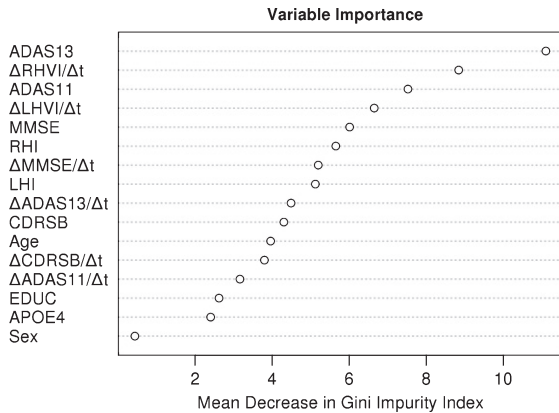


Fig. 3. Order of importance of variables for Random Forest classification as measured by the mean decrease in the Gini impurity index when the variable is used for splitting a tree node in the Random Forest.

is in the low 70th percentiles [5, 7–14]. Therefore, it can be said that the overall estimated accuracy of our Random Forest classifier is high in comparison to previous reports.

A very interesting finding of this research is that the accuracy of classification when broken down by sex (Table 4) was considerably higher in women (89.1%) than in men (78.9%). With the classification accuracy in women, to the best of our knowledge, being the highest reported in any previous application of machine learning to the current problem. A possible explanation for the greater accuracy achieved in predicting progression from MCI to AD in women could simply be sampling bias in the ADNI dataset, and that this result may not be generalizable to the MCI population. On the other hand, multiple studies suggest that the trajectories of structural and functional changes in the brain and the associated clinical and cognitive deficits in the course of AD development may be different between men and women. The apolipoprotein E $\epsilon 4$ allele, the strongest known genetic factor for sporadic AD [35], has been shown to confer greater risk of AD development in women [36–39]. In cross-sectional studies, atrophy of the medial temporal lobe structures in MCI and mild to moderate AD has been shown to be associated with APOE genotype with greater prominence in women compared to men [40, 41]. Longitudinally, sex and age interactions have also been shown in brain atrophy rates obtained by voxelwise analyses [42, 43]. ROI analyses also show sex differences in atrophy rates of several brain structures including the hippocampus and entorhinal cortex in healthy controls, MCI, and AD subjects [16, 43]. Sex differ-

ences have also been observed in the rates of decline of clinical and cognitive measures in MCI subjects with faster deterioration in women [44, 45]. Taken together, these studies suggest that the patterns of AD predictors at baseline and their rates of change with time are different between men and women with MCI. Therefore, it is conceivable that there may be stronger indicators of imminent AD dementia (or lack thereof) in women with MCI as compared to men, and that the results of this paper may be generalized to the MCI population. Future replication studies using independent datasets are necessary to determine whether the results of this paper are simply due to a sampling bias in ADNI or whether in fact it is possible to predict MCI to AD conversion in women with greater accuracy than in men.

We also found that the effect size of HVI difference between sMCI and pMCI groups is larger for the right hippocampus as compared to the left hippocampus (Table 2), and that the HVI and its rate of change on the right hippocampus are more important variables for classification than those of left hippocampus (Fig. 3). These results are consistent with several volumetric studies that have found the right hippocampus to be more affected than the left in AD and MCI [46–48].

A limitation of the current approach is the requirement of longitudinal baseline and one-year follow-up MRI volumes for prediction of MCI to AD conversion. A shorter baseline to follow-up period (e.g., 6 months) would make the current approach a much more practical proposition for both diagnosis and clinical trials. But it would also likely reduce the signal-to-noise ratio of the longitudinal variables used for prediction. Therefore, some loss in the prediction accuracy could be expected when the initial observation period is reduced. It may be possible to compensate for this potential reduction in accuracy by including additional AD biomarkers as features in the classification process. Future studies are required to assess the performance of the Random Forest classification algorithms trained based on shorter observation periods with additional biomarkers.

Additional biomarkers could for example be related to the size and shape of the corpus callosum derived from structural MRI which many studies have shown to be different between mild AD and normal subjects both in their baseline levels and rates of change longitudinally [49–52]. Other potential structural MRI biomarkers are measures of atrophy in the entorhinal cortex [53–55] and posterior cingulate [56, 57]. The feature space could also be augmented by

additional clinical and cognitive measures, such as the Functional Activities Questionnaire (FAQ) [58], a low cost 10-item measure ranging from 0 (normal) to 30 (maximum impairment) which has been shown to have power in predicting a change in diagnostic status from MCI to AD [59–61].

Features used for classification could also include other MRI modalities such as resting-state functional MRI [62], diffusion tensor imaging [63], and arterial spin labeling [64], each of which have shown promise in discriminating between normal aging and AD. Thus a multimodality approach could potentially improve performance. Our use of raw T1-weighted imaging requires the shortest acquisition time and least pre- and post-processing.

It should also be mentioned that the MRI volumes used in the current paper had been acquired using 1.5 Tesla scanners. Intuitively, using higher signal-to-noise ratio volumes that can be obtained from modern 3 Tesla MRI scanners could potentially increase the accuracy of the prediction methods presented in this paper. However, potential improvements need to be demonstrated in future studies.

The mean \pm SD baseline to probable AD diagnosis in the pMCI group was 2.84 ± 1.50 years. This means that after the baseline and 1-year follow-up scans are acquired, if a prediction of imminent conversion to probable AD is made based on the techniques developed in this paper, there is on average a 1.8-year window of treatment aimed at slowing or stopping disease progression. This window of opportunity for therapeutic intervention can be extended by six months if the length of the observation period can be reduced from one year to six months without sacrificing accuracy, e.g., by using 3 Tesla scanners and including corpus callosum and other structural MRI biomarkers.

Another limitation of this study is that even though the accuracies achieved, particularly in women, are among the highest reported in the literature, roughly 80% specificity/sensitivity is still not high enough to be applicable for routine clinical work. Thus, the challenge remains to increase the accuracy in future computer aided diagnosis methods. However, any prediction performance above the level of chance will reduce the cost of clinical trials of therapeutic interventions. It should also be kept in mind that sample noise in both training and testing data (e.g., mislabeling progressive cases as sMCI) will always impose a fundamental limit in achievable accuracy.

In conclusion, we have shown in this paper that using longitudinal HVI measures obtained from

structural MRI along with a number of common cognitive tests and demographic and genetic information, it is possible to accurately predict whether or not an MCI subject would remain stable or progress to develop AD type dementia. We also found that the estimated prediction accuracy in women is considerably higher than that of men. An important feature of this work is that our approach is designed to be clinically practical and viable. The measures used in our prediction are widely available, non-invasive, and measureable without much difficulty or requiring substantial expertise or preprocessing of data. We avoid preprocessing of structural MRI scans (e.g., distortion correction, B1 field inhomogeneity correction, etc.) on the grounds that these procedures may not be easily available for routine clinical work. We take advantage of the longitudinal data based on evidence that biomarkers of AD are not just different at baseline but their rate of change is different between stable and progressive MCI groups. Finally, we use the Random Forest machine learning algorithm, a powerful technique that has not been as widely applied to the current problem as other algorithms. Future studies on independent datasets are necessary to confirm our results of sex differences in prediction accuracy. If confirmed, this finding will have important implications in both clinical practice and in research.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Neu-

roRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (<http://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 Southern California.

Authors' disclosures available online (<http://j-alz.com/manuscript-disclosures/16-0594r1>).

REFERENCES

- [1] Wilson RS, Segawa E, Boyle PA, Anagnos SE, Hizek LP, Bennett DA (2012) The natural history of cognitive decline in Alzheimer's disease. *Psychol Aging* **27**, 1008-1017.
- [2] Remington R, Bechtel C, Larsen D, Samar A, Doshanji L, Fishman P, Luo Y, Smyers K, Page R, Morrell C, Shea TB (2015) A phase II randomized clinical trial of a nutritional formulation for cognition and mood in Alzheimer's disease. *J Alzheimers Dis* **45**, 395-405.
- [3] Wischik CM, Staff RT, Wischik DJ, Bentham P, Murray AD, Storey JM, Kook KA, Harrington CR (2015) Tau aggregation inhibitor therapy: An exploratory phase 2 study in mild or moderate Alzheimer's disease. *J Alzheimers Dis* **44**, 705-720.
- [4] <https://clinicaltrials.gov>
- [5] Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, Londos E, Hansson O (2012) Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: Six-year follow-up study. *PLoS One* **7**, e38639.
- [6] Ganguli M, Snitz BE, Saxton JA, Chang CC, Lee CW, Vander Bilt J, Hughes TF, Loewenstein DA, Unverzagt FW, Petersen RC (2011) Outcomes of mild cognitive impairment by definition: A population study. *Arch Neurol* **68**, 761-767.
- [7] Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging* **32**, 2322.e19-27.
- [8] Hinrichs C, Singh V, Xu G, Johnson SC (2011) Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage* **55**, 574-589.
- [9] Zhang D, Wang Y, Zhou L, Yuan H, Shen D (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856-867.
- [10] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR Jr, Feldman HH, Bokde AL, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* **33**, 1203-1214.
- [11] Cheng B, Liu M, Zhang D, Munsell BC, Shen D (2015) Domain transfer learning for MCI conversion prediction. *IEEE Trans Biomed Eng* **62**, 1805-1817.
- [12] Eskildsen SF, Coupé P, Fonov VS, Pruessner JC, Collins DL (2015) Structural imaging biomarkers of Alzheimer's disease: Predicting disease progression. *Neurobiol Aging* **36**(Suppl 1), S23-S31.
- [13] Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* **104**, 398-412.
- [14] Lebedev AV, Westman E, Van Westen GJ, Kramberger MG, Lundervold A, Aarsland D, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Simmons A (2014) Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *Neuroimage Clin* **6**, 115-125.
- [15] Falahati F, Westman E, Simmons A (2014) Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. *J Alzheimers Dis* **41**, 685-708.
- [16] Ardekani BA, Convit A, Bachman AH (2016) Analysis of the MIRIAD data shows sex differences in hippocampal atrophy progression. *J Alzheimers Dis* **50**, 847-857.
- [17] Monsell SE, Mock C, Hassenstab J, Roe CM, Cairns NJ, Morris JC, Kukull W (2014) Neuropsychological changes in asymptomatic persons with Alzheimer disease neuropathology. *Neurology* **83**, 434-440.
- [18] Breiman L (2001) Random forests. *Machine Learning* **45**, 5-32.
- [19] Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, 161-168.
- [20] Wyman BT, Harvey DJ, Crawford K, Bernstein MA, Carmichael O, Cole PE, Crane PK, DeCarli C, Fox NC, Gunter JL, Hill D, Killiany RJ, Pachai C, Schwarz AJ, Schuff N, Senjem ML, Suhy J, Thompson PM, Weiner M, Jack CR Jr (2013) Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers Dement* **9**, 332-337.
- [21] Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DLG, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbs JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Mag Res Imag* **27**, 685-691.
- [22] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state" A practical method for grading the cognitive state of patients for the clinician. *J Psychiatric Res* **12**, 189-198.
- [23] Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL (1982) A new clinical scale for the staging of dementia. *Br J Psychiatry* **140**, 566-572.
- [24] Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **43**, 2412-2414.
- [25] Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *Am J Psychiatry* **141**, 1356-1364.
- [26] Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, Sano M, Bieliauskas L, Geldmacher D, Clark

- C, Thal LJ (1997) Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* **11**, S13-S21.
- [27] Ardekani BA, Kershaw J, Braun M, Kanno I (1997) Automatic detection of the mid-sagittal plane in 3-D brain images. *IEEE Trans Med Imaging* **16**, 947-952.
- [28] Ardekani BA, Bachman AH (2009) Model-based automatic detection of the anterior and posterior commissures on MRI scans. *Neuroimage* **46**, 677-682.
- [29] Liaw A, Wiener M (2002) Classification and Regression by randomForest. *R News* **2**, 18-22.
- [30] Breiman L (1996) Out-of-bag estimation, <http://www.stat.berkeley.edu/~breiman/OOBestimation.pdf>
- [31] Convit A, De Leon MJ, Tarshish C, De Santi S, Tsui W, Rusinek H, George A (1997) Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiol Aging* **18**, 131-138.
- [32] Jack CR, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ, Tangalos EG, Kokmen E (1998) Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* **51**, 993-999.
- [33] Kehoe EG, McNulty JP, Mullins PG, Bokde AL (2014) Advances in MRI biomarkers for the diagnosis of Alzheimer's disease. *Biomark Med* **8**, 1151-1169.
- [34] Prestia A, Caroli A, Wade SK, van der Flier WM, Ossenkoppele R, Van Berckel B, Barkhof F, Teunissen CE, Wall A, Carter SF, Schöll M, Choo IH, Nordberg A, Scheltens P, Frisoni GB (2015) Prediction of AD dementia by biomarkers following the NIA-AA and IWG diagnostic criteria in MCI patients from three European memory clinics. *Alzheimers Dement* **11**, 1191-1201.
- [35] Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper-MacLachlan DR, Alberts MJ, Hulette C, Crain B, Goldgaber D, Roses AD (1993) Association of apolipoprotein E allele $\epsilon 4$ with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467-1472.
- [36] Altmann A, Tian L, Henderson VW, Greicius MD (2014) Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann Neurol* **75**, 563-573.
- [37] Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349-1356.
- [38] Payami H, Zarepari S, Montee KR, Sexton GJ, Kaye JA, Bird TD, Yu CE, Wijsman EM, Heston LL, Litt M, Schellenberg GD (1996) Gender difference in apolipoprotein E-associated risk for familial Alzheimer disease: A possible clue to the higher incidence of Alzheimer disease in women. *Am J Hum Genet* **58**, 803-811.
- [39] Ungar L, Altmann A, Greicius MD (2014) Apolipoprotein E, gender, and Alzheimer's disease: An overlooked, but potent and promising interaction. *Brain Imaging Behav* **8**, 262-273.
- [40] Fleisher A, Grundman M, Jack CR Jr, Petersen RC, Taylor C, Kim HT, Schiller DH, Bagwell V, Sencakova D, Weiner MF, DeCarli C, DeKosky ST, van Dyck CH, Thal LJ (2005) Sex, apolipoprotein E $\epsilon 4$ status, and hippocampal volume in mild cognitive impairment. *Arch Neurol* **62**, 953-957.
- [41] Juottonen K, Lehtovirta M, Helisalme S, Riekkinen PJ Sr, Soininen H (1998) Major decrease in the volume of the entorhinal cortex in patients with Alzheimer's disease carrying the apolipoprotein E $\epsilon 4$ allele. *J Neurol Neurosurg Psychiatry* **65**, 322-327.
- [42] Hua X, Hibar DP, Lee S, Toga AW, Jack CR, Weiner MW, Thompson PM (2010) Sex and age differences in atrophic rates: An ADNI study with $n=1368$ MRI scans. *Neurobiol Aging* **31**, 1463-1480.
- [43] Skup M, Zhu H, Wang Y, Giovanello KS, Lin J, Shen D, Shi F, Gao W, Lin W, Fan Y, Zhang H (2011) Sex differences in grey matter atrophy patterns among AD and aMCI patients: Results from ADNI. *Neuroimage* **56**, 890-906.
- [44] Holland D, Desikan RS, Dale AM, McEvoy LK (2013) Higher rates of decline for women and apolipoprotein E $\epsilon 4$ carriers. *AJNR Am J Neuroradiol* **34**, 2287-2293.
- [45] Lin KA, Choudhury KR, Rathakrishnan BG, Marks DM, Petrella JR, Doraiswamy PM (2015) Marked gender differences in progression of mild cognitive impairment over 8 years. *Alzheimers Dement Transl Res Clin Interv* **1**, 103-110.
- [46] Barnes J, Schill RI, Schott JM, Frost C, Rossor MN, Fox NC (2005) Does Alzheimer's disease affect hippocampal asymmetry? Evidence from a cross-sectional and longitudinal volumetric MRI study. *Dement Geriatr Cogn Disord* **19**, 338-344.
- [47] Teipel SJ, Pruessner JC, Faltraco F, Born C, Rocha-Unold M, Evans A, Möller HJ, Hampel H (2006) Comprehensive dissection of the medial temporal lobe in AD: Measurement of hippocampus, amygdala, entorhinal, perirhinal and parahippocampal cortices using MRI. *J Neurol* **253**, 794-800.
- [48] Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM (2006) Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Arch Neurol* **63**, 693-699.
- [49] Elahi S, Bachman AH, Lee SH, Sidtis JJ, Ardekani BA (2015) Corpus callosum atrophy rate in mild cognitive impairment and prodromal Alzheimer's disease. *J Alzheimers Dis* **45**, 921-931.
- [50] Bachman AH, Lee SH, Sidtis JJ, Ardekani BA (2014) Corpus callosum shape and size changes in early Alzheimer's disease: A longitudinal MRI study using the OASIS brain database. *J Alzheimers Dis* **39**, 71-78.
- [51] Ardekani BA, Bachman AH, Figarsky K, Sidtis JJ (2014) Corpus callosum shape changes in early Alzheimer's disease: An MRI study using the OASIS brain database. *Brain Struct Funct* **219**, 343-352.
- [52] Lee SH, Yu D, Bachman AH, Lim J, Ardekani BA (2014) Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease. *J Neurosci Methods* **15**, 78-84.
- [53] Dickerson BC, Goncharova I, Sullivan MP, Forchetti C, Wilson RS, Bennett DA, Beckett LA, deToledo-Morrell L (2001) MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiol Aging* **22**, 747-754.
- [54] Devanand DP, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, Rusinek H, Pelton GH, Honig LS, Mayeux R, Stern Y, Tabert MH, de Leon MJ (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology* **68**, 828-836.
- [55] Juottonen K, Laakso MP, Insausti R, Lehtovirta M, Pitkänen A, Partanen K, Soininen H (1998) Volumes of the entorhinal

- and perirhinal cortices in Alzheimer's disease. *Neurobiol Aging* **19**, 15-22.
- [56] Jones BF, Barnes J, Uylings HB, Fox NC, Frost C, Witter MP, Scheltens P (2006) Differential regional atrophy of the cingulate gyrus in Alzheimer disease: A volumetric MRI study. *Cereb Cortex* **16**, 1701-1708.
- [57] Pengas G, Hodges JR, Watson P, Nestor PJ (2010) Focal posterior cingulate atrophy in incipient Alzheimer's disease. *Neurobiol Aging* **31**, 25-33.
- [58] Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community (1982) *J Gerontol* **37**, 323-329
- [59] Tabert MH, Albert SM, Borukhova-Milov L, Camacho Y, Pelton G, Liu X, Stern Y, Devanand DP (2002) Functional deficits in patients with mild cognitive impairment: Prediction of AD. *Neurology* **58**, 758-764.
- [60] Teng E, Becker BW, Woo E, Knopman DS, Cummings JL, Lu PH (2010) Utility of the functional activities questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer disease. *Alzheimer Dis Assoc Disord* **24**, 348-353.
- [61] Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE (2011) Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. *Arch Gen Psychiatry* **68**, 961-969.
- [62] Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, Jiang T (2007) Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. *Hum Brain Mapp* **28**, 967-978.
- [63] Nir TM, Jahansad N, Villalon-Reina JE, Toga AW, Jack CR, Weiner MW, Thompson PM (2013) Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *Neuroimage Clin* **3**, 180-195.
- [64] Collij LE, Heeman F, Kuijjer JPA, Ossenkoppele R, Benedictus MR, Moller C, Verfaillie SCJ, Sanz-Arigitia EJ, van Berckel BNM, van der Flier WM, Scheltens P, Barkhof F, Wink AM (2016) Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. *Radiology*. doi: 10.1148/radiol.2016152703