Neurobiology of Aging 34 (2013) 2815–2826

Contents lists available at SciVerse ScienceDirect

Neurobiology of Aging

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

How early can we predict Alzheimer's disease using computational anatomy?

Stanisław Adaszewski [a,b,1], Juergen Dukart [a,*,1], Ferath Kherif [a], Richard Frackowiak [a], Bogdan Draganski [a,c], for the Alzheimer's Disease Neuroimaging Initiative[2]

[a] Département des Neurosciences Cliniques, Laboratoire de Recherche en Neuroimagerie, Centre Hospitalier Universitaire Vaudois, Université de Lausanne, Lausanne, Switzerland
[b] Department of Neurology, Faculty of Electronics and Information Technology, Warsaw University of Technology, Warsaw, Poland
[c] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:
Received 2 April 2013
Received in revised form 24 May 2013
Accepted 20 June 2013
Available online 26 July 2013

Keywords:
Structural magnetic resonance imaging
Alzheimer's disease
Mild cognitive impairment
Biomarker

ABSTRACT

Computational anatomy with magnetic resonance imaging (MRI) is well established as a noninvasive biomarker of Alzheimer's disease (AD); however, there is less certainty about its dependency on the staging of AD. We use classical group analyses and automated machine learning classification of standard structural MRI scans to investigate AD diagnostic accuracy from the preclinical phase to clinical dementia. Longitudinal data from the Alzheimer's Disease Neuroimaging Initiative were stratified into 4 groups according to the clinical status—(1) AD patients; (2) mild cognitive impairment (MCI) converters; (3) MCI nonconverters; and (4) healthy controls—and submitted to a support vector machine. The obtained classifier was significantly above the chance level (62%) for detecting AD already 4 years before conversion from MCI. Voxel-based univariate tests confirmed the plausibility of our findings detecting a distributed network of hippocampal-temporoparietal atrophy in AD patients. We also identified a subgroup of control subjects with brain structure and cognitive changes highly similar to those observed in AD. Our results indicate that computational anatomy can detect AD substantially earlier than suggested by current models. The demonstrated differential spatial pattern of atrophy between correctly and incorrectly classified AD patients challenges the assumption of a uniform pathophysiological process underlying clinically identified AD.

1. Introduction

Recent advances in computer-based diagnosis making use of structural magnetic resonance imaging (sMRI) and machine learning methods provide evidence of sufficient accuracy in discriminating Alzheimer's disease (AD) patients not only from healthy controls but also from other common types of dementia (Davatzikos et al., 2008b; Dukart et al., 2011a, 2012; Fan et al., 2008; Kloppel et al., 2008a, 2008b). For the clinically and neuroscientifically pertinent case of early AD detection, support vector machine (SVM) classification and other machine learning studies tapping into the preclinical phase of AD convincingly demonstrate the potential for reliable early diagnosis (Casanova and Hsu, 2012; Davatzikos et al., 2008a; Devanand et al., 2007; McEvoy et al., 2009; Misra et al., 2009; Modrego, 2006). Two limitations applying to most of the previous studies are the tuning of the classifiers to specific cohorts and with respect to cross-validation. Both restrict their generalizability to the general population. Tuning of a classifier to achieve a high accuracy for detection of mild cognitive impairment (MCI) patients might result in a substantial drop in accuracy when applying the same classifier to AD patients. Similarly, the tuning of a classifier to achieve high cross-validation accuracies might substantially increase the risk of overfitting the classification model to the particular dataset used in the study therewith providing an overoptimistic estimation for the accuracy of the method when applied to a general population.

Despite the progress in the field of computer-based AD detection, our knowledge about the capability of sMRI for early diagnosis even before the first manifestation of clinical signs is still very limited. The most recent model of brain anatomy–derived biomarker in AD (Jack et al., 2010) suggested a protracted progression of atrophy compared with functional changes as observed by [F18]fluorodeoxyglucose positron emission tomography. In contrast, other prospective studies provided evidence that sMRI measurements may contain information of ongoing disease-related process already before clinical

* Corresponding author at: Département des Neurosciences Cliniques, LREN, CHUV, Université de Lausanne, Rue de Bugnon 21, 1011 Lausanne, Switzerland. Tel.: +41 21 3149593; fax: +41 21 3140916.
E-mail address: dukart@cbs.mpg.de (J. Dukart).

[1] These authors contributed equally to this work.

[2] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.ucla.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.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

0197-4580/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2013.06.015

Table 1
Subject group characteristics

	Training set		Testing set				ANOVA
	AD	Control subjects	AD	Control subjects	ncMCI	cMCI	*F*, df, *p*
n	54	54	54	83	61	142	—
Age (y) (mean ± SD)	74.4 ± 8.3	74.6 ± 5.5	75.8 ± 6.9	75.3 ± 5.0	74.0 ± 7.6	74.0 ± 7.0	0.8, 5, 0.544
Gender (M/F)	31/23	29/25	25/29	40/43	45/16	93/49	—
MMSE (mean ± SD)	23.1 ± 1.9	29.0 ± 1.1	23.4 ± 2.1	29.3 ± 0.8	27.5 ± 1.9	26.7 ± 1.7	166.5, 5, <0.001
Follow-up time (y) (mean ± SD)	2.1 ± 0.2	3.3 ± 0.7	2.8 ± 0.4	3.8 ± 1.0	3.4 ± 0.9	3.5 ± 0.9	—

Key: AD, Alzheimer's disease; ANOVA, analysis of variance; cMCI, mild cognitive impairment (converters); df, degree of freedom; F, female; M, male; MMSE, mini-mental state examination; ncMCI, mild cognitive impairment (nonconverters); SD, standard deviation.

manifestation of cognitive decline (Dickerson and Wolk, 2012; Quiroz et al., 2012; Smith et al., 2008). However, these studies did not test the predictive power of sMRI information to detect AD in untested subjects or cross-validation. Therefore, there is a pressing need to investigate the timescale of disease-related structural brain changes in AD not only to advance our understanding of the disorder but also to provide better tools for early diagnosis when neuroprotection is possible. Another related aspect is that the application of multivariate pattern classification techniques for early AD detection produces a high proportion of erroneous predictions for conversion from MCI to AD (Ewers et al., 2010; Misra et al., 2009). Thus, the secondary aim of our study is to investigate if false predictions are because of random noise or deterministic atrophy pattern.

We systematically address the questions of timescale of disease detection and potential causes of erroneous prediction while aiming to overcome the aforementioned limitations. We first adopt a pragmatic strategy testing whether AD-related atrophy is already detectable several years before conversion followed by in-depth investigation of atrophy patterns comparing incorrectly and correctly diagnosed AD, MCI converting to AD during the follow-up (MCI converters [cMCI]), MCI nonconverters (ncMCI), and healthy control subjects. To this end, we apply classical mass-univariate voxel-based analysis paralleled by machine learning classification using SVMs.

2. Methods

2.1. Subjects

To evaluate temporal sensitivity of sMRI data for early detection of AD, we used 1.5-T T1-weighted images from the Alzheimer's Disease Neuroimaging Initiative (ADNI, http://www.adni-info.org/) of all available AD, cMCI, and ncMCI patients and healthy controls who had baseline and at least 2 years of follow-up MRI scans.

The AD patient and control subject data were split into a dataset used for SVM classifier training and another for diagnosis (Tables 1 and 2). Critically, the cMCI (Table 3) and ncMCI (Table 2) data were used for diagnosis only. Baseline and follow-up scans after 6, 12, 24, 36, 48, and 60 months, if available, were downloaded from the ADNI database along with the corresponding clinical information. All the data available in ADNI1 and ADNI-GO studies were used for subsequent evaluation. The diagnosis of AD was based on NINCDS/ADRDA (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association) criteria (McKhann et al., 1984). Exclusion criteria were the presence of any significant neurologic disease other than AD, history of head trauma followed by persistent neurologic deficits or structural brain abnormalities, psychotic features, agitation or behavioral problems within the previous 3 months, or history of alcohol or substance abuse. The study was conducted according to the Declaration of Helsinki. Written informed consent was obtained from all participants before protocol-specific procedures were performed.

The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether sMRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments, monitor their effectiveness, and lessen the time and cost of clinical trials. The Principal Investigator of this initiative

Table 2
Follow-up testing group characteristics

	Baseline	1 y	2 y	3 y	4 y
Control subjects					
n	42	42	37	38	14
Age (mean ± SD)	76.5 ± 4.7	77.1 ± 4.7	77.7 ± 5.1	79.2 ± 4.4	81.8 ± 4.4
Gender (M/F)	18/24	18/24	16/21	17/21	8/6
MMSE (mean ± SD)	29.2 ± 0.8	29.3 ± 1.0	29.4 ± 0.9	29.1 ± 1.3	29.5 ± 1.1
AD					
n	54	53	54	—	—
Age (mean ± SD)	76.7 ± 7.0	77.1 ± 6.7	78.8 ± 7.0	—	—
Gender (M/F)	25/29	25/28	25/29	—	—
MMSE (mean ± SD)	23.4 ± 2.1	22.4 ± 3.7	19.2 ± 5.8	—	—
ncMCI					
n	61	61	61	44	—
Age (mean ± SD)	75.1 ± 7.7	75.7 ± 7.7	76.8 ± 7.6	79.0 ± 7.4	—
Gender (M/F)	45/16	45/16	45/16	34/10	—
MMSE (mean ± SD)	27.5 ± 1.9	27.4 ± 2.3	27.0 ± 3.1	27.3 ± 2.1	—

Key: AD, Alzheimer's disease; cMCI, mild cognitive impairment (converters); F, female; M, male; MMSE, mini-mental state examination; ncMCI, mild cognitive impairment (nonconverters); SD, standard deviation.

Table 3
Follow-up cMCI characteristics

	TTC, 4 y	TTC, 3 y	TTC, 2 y	TTC, 1 y	ToC	ToC, +1 y	ToC+, 2 y
cMCI							
n	17	38	83	128	129	70	45
Age (mean ± SD)	74.0 ± 6.5	75.2 ± 6.4	75.6 ± 7.0	75.8 ± 6.7	76.7 ± 6.7	77.5 ± 6.4	78.3 ± 7.0
Gender (M/F)	14/3	24/14	54/29	85/43	83/46	47/23	30/15
MMSE (mean ± SD)	26.7 ± 1.6	26.7 ± 2.0	26.7 ± 1.8	26.0 ± 2.3	24.6 ± 2.6	21.7 ± 4.1	20.4 ± 5.3
cMCI (same subgroup)							
n	—	17	17	17	17	—	—
Age (mean ± SD)	—	73.8 ± 5.6	74.7 ± 5.5	75.7 ± 5.5	76.9 ± 5.6	—	—
Gender (M/F)	—	11/6	11/6	11/6	11/6	—	—
MMSE (mean ± SD)	—	26.9 ± 2.0	26.6 ± 2.0	26.5 ± 2.2	23.3 ± 2.5	—	—

"Same subgroup" indicates that the same cMCI patients were used for all time points.
Key: AD, Alzheimer's disease; cMCI, mild cognitive impairment (converters); F, female; M, male; MMSE, mini-mental state examination; SD, standard deviation; ToC, time point of conversion; TTC, time to conversion.

is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from >50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55–90, to participate in the research, 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information, see http://www.adni-info.org/.

2.2. MRI data

The MRI dataset included standard T1-weighted images obtained with different 1.5-T scanner types using a three-dimensional magnetization-prepared rapid gradient-echo sequence varying in repetition time and echo time with an in-plane resolution of 1.25 × 1.25 mm and 1.2 mm slice thickness. All images were modified as described on the ADNI Web site (http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml) including correction for distortions and B1-field nonuniformity.

2.3. Image preprocessing

Structural MRI data preprocessing was carried out using the SPM8 software package (Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab 7.11 (MathWorks, Inc, Sherborn, MA, USA). For automated tissue classification in gray matter (GM), white matter, and cerebrospinal fluid, we used the NewSegment option. Aiming at optimum anatomic precision, we then applied a diffeomorphic registration algorithm DARTEL (Ashburner, 2007) for spatial registration to Montreal Neurological Institute space using a study-specific template created from a random representative subsample of 400 scans. The obtained GM images were additionally modulated to preserve the total amount of signal from each region. The data were smoothed with a Gaussian kernel of 8-mm full width at half maximum and masked by applying a probability threshold of 0.2 to the first and the last DARTEL templates (Dukart et al., 2011a). The obtained mask contains all relevant GM structures whereas excluding smoothing-related artifacts at the edges and signal coming from periventricular GM. Overall, 384,065 voxels were left after threshold masking restricting all subsequent SVM and voxel-based morphometry (VBM) analyses to the GM structures of the brain. Additionally, to dissociate disease-related effects from changes associated with healthy aging, we used a linear "age-detrending" model (Dukart et al., 2011b). This method estimated an age model with voxel-wise linear fit in a group of healthy controls ($n = 41$) randomly selected from the ADNI dataset and not used for further analyses. The obtained model parameters are then applied to the sMRI data accounting for any variance attributable to healthy aging in all subjects and patients.

2.4. SVM classification

For SVM classification, we used the freely available LIBSVM software (Chang and Lin, 2001). The data of AD patients and healthy controls were split randomly into a dataset used for training of the SVM classifier ($n = 54/54$, AD/healthy controls) and a separate dataset used for classification-based diagnosis (Tables 1 and 2). The training dataset was used to build a linear soft-margin SVM classifier using voxel-wise whole-brain GM information without any further parameter optimization and default cost value of 1 (Kloppel et al., 2008b). The independent diagnostic set included data of AD, cMCI, and ncMCI patients and healthy controls (Tables 2 and 3).

To obtain the precision of classification accuracy for the diagnostic cohorts, we apply bootstrapping with 100 permutations using whole-brain information from a randomly selected 2/3 of the training set without replacement and keeping the original proportion of AD and controls (36 AD/36 healthy subjects) from the whole training dataset to build a classifier to predict test data of healthy controls and cMCI, ncMCI, and AD patients at each data acquisition time point. Aiming at an unbiased and precise estimation of classification/prediction accuracy, we used bootstrapping, which was shown in the literature to provide a more accurate estimation of the classification accuracy compared with other approaches such as leave-one-out or other cross-validation modes (Dougherty et al., 2010; Jain et al., 1987).

The results are presented using 2 different timescales—for healthy controls, AD, and ncMCI, the zero point on time axis signifies the baseline scan, whereas for cMCI, it signifies the time point of conversion (time to conversion [TTC]). The cMCI data were ordered on the timescale relative to the TTC as this ordering indicates that all subjects reached a predefined level of cognitive decline. For ncMCI, classification performance was evaluated as assignment to the healthy control group, whereas for cMCI, correct classification was considered as assignment to the AD group. To enable longitudinal estimates of accuracy in the same group of cMCI patients, all classifications were repeated for a subset ($n = 17$) with imaging data for each of 3 years up to time point of conversion.

In an additional step, we performed feature selection within the training dataset using weights provided by the classifier for each voxel as a measurement of its contribution to the separation between AD patients and healthy controls (weights close to zero indicate a low contribution). The absolute values of the weights

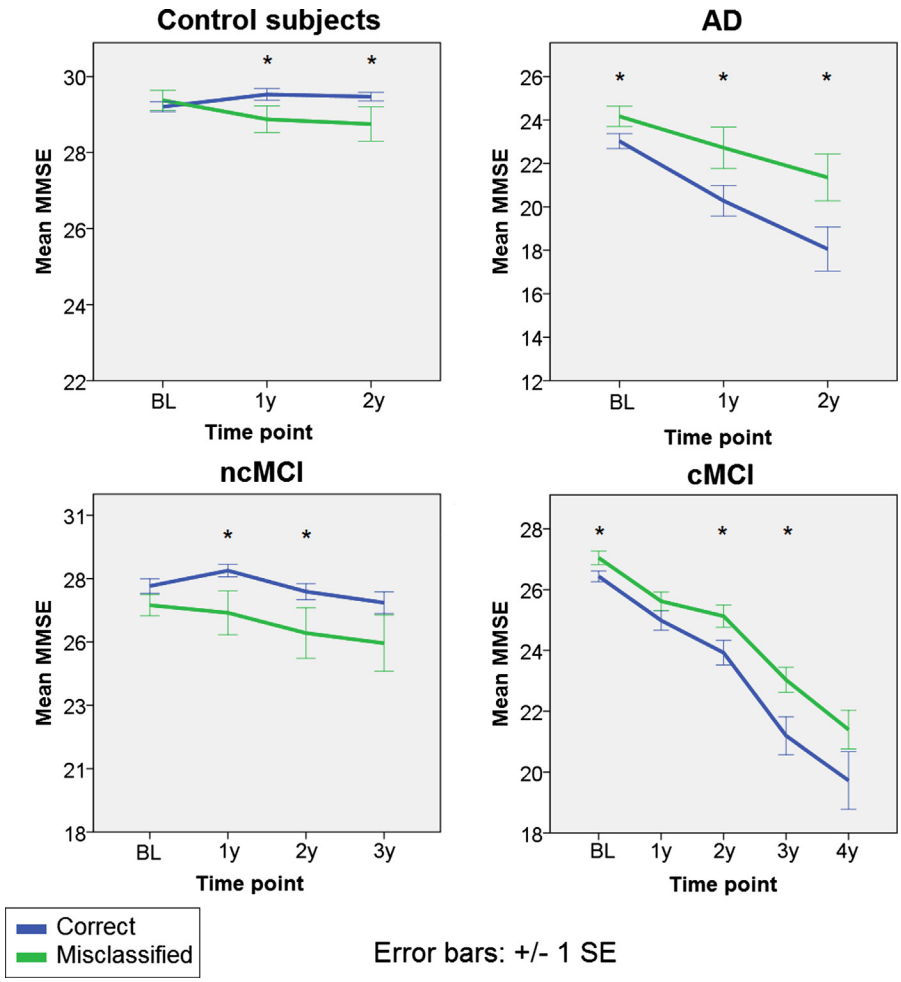

Fig. 1. Mean MMSE values of correctly and incorrectly classified subjects within each clinical group at different time points. For comparability reasons and to facilitate interpretation, MMSE values for cMCI subjects are also displayed as starting from the baseline examination. * Significant differences (one sided, $p < 0.05$) in MMSE between correctly and incorrectly classified subjects. Abbreviations: AD, Alzheimer's disease; BL, baseline; cMCI, mild cognitive impairment (converters); MMSE, mini-mental state examination; ncMCI, mild cognitive impairment (nonconverters); SE, standard error.

were first sorted from lowest to highest. In a second step, we sequentially removed features with lowest weights as provided by the whole-brain classification. Similar to optimization of SVM parameters to a particular training dataset, iterative feature selection procedures often carry the risk of overfitting a classifier to the training dataset. This might be the case when the weights are redetermined after each selection step, and the newly obtained weights are used to guide further feature selection steps. This procedure leads to establishment of numerous feature maps each having different weights assigned to the same features. Selection of the best of them without accounting for the total number of calculated weight maps may therefore lead to overoptimistic accuracy estimations. To avoid overfitting the classifier by multiple testing, the distribution of weights was calculated only once, using the classifier trained on whole-brain information in the training cohort. The computed maximum weight was set as 100% (with a weight of zero corresponding to 0%). Subsequently, all features with weights <5% of the maximum were removed from classification. A classifier based on weights exceeding this threshold was used on all cMCI data at each time point. This procedure was repeated in 5% steps for the whole range of weights by consecutively removing features with lowest weights in the initial whole-brain classification, training a classifier on the remaining features, and reclassifying the cMCI data.

All classification accuracies with and without feature selection were compared with each other, between consecutive time points, and with chance-level distributions (obtained from the same data by randomly shuffling diagnostic labels within the training dataset) by using 2-sample t tests with a significance threshold of 0.05 (1 tailed, Bonferroni corrected for multiple comparisons).

2.5. Voxel-based morphometry

We performed analyses of covariance in the test cohort VBM data comparing correctly and incorrectly classified AD, cMCI, and ncMCI patients to each other and to healthy controls. For each patient, we used the first scan misclassified using whole-brain information in more than 50% of all permutations for VBM comparisons. As group sizes of correctly and incorrectly classified subjects differed within each clinical group, a random sample was drawn from the bigger group to match the smaller group in size. Age, gender, and total intracranial volume were included as covariates for all comparisons. A statistical threshold of $p < 0.001$ uncorrected at voxel level and an extent threshold of 200 voxels at cluster level were applied for all statistical analyses. We used the automated anatomic labeling atlas (Tzourio-Mazoyer et al., 2002) implemented in the Matlab-based WFU PickAtlas 2.3 toolbox

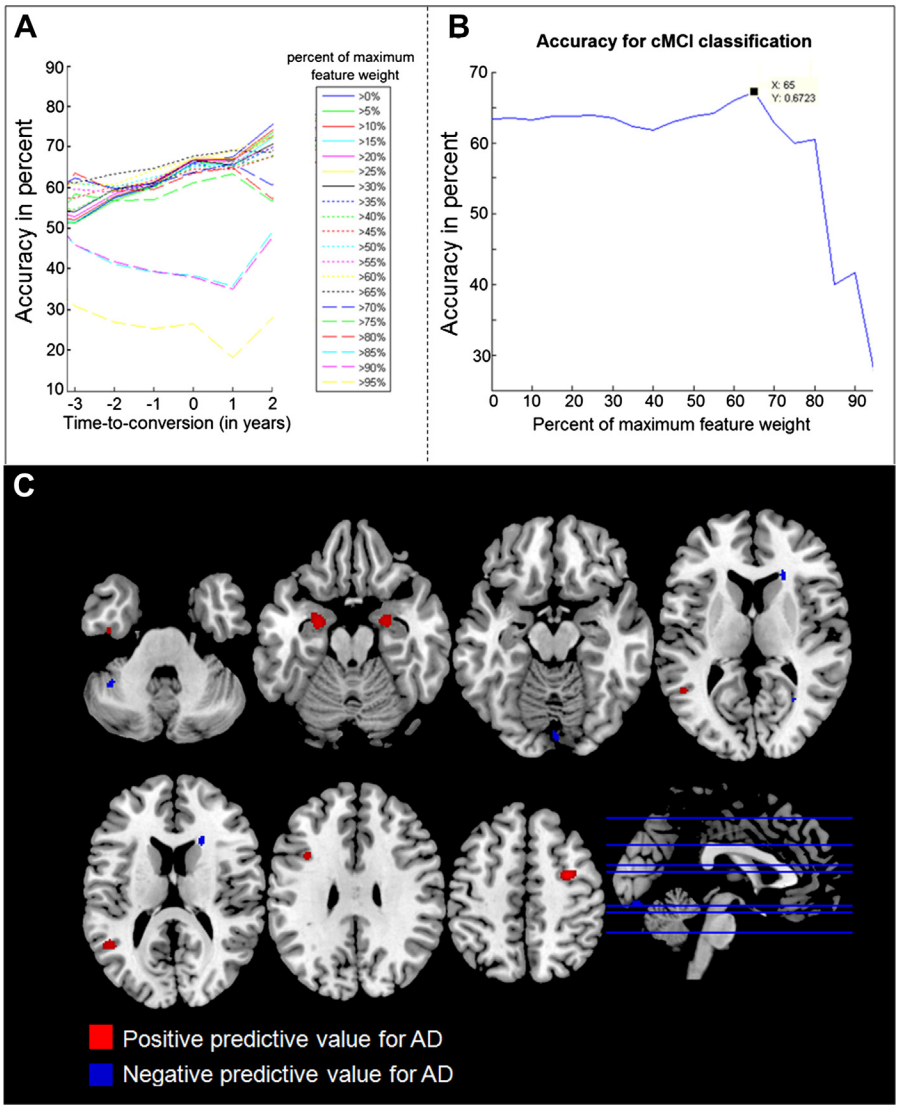

Fig. 2. (A) Performance of the SVM, classifier with different numbers of features for classification of the cMCI group at different time points. Features are selected with their weight exceeding a predefined percentage threshold of the maximum feature weight obtained using whole-brain classification. (B) Performance of the feature set providing the highest average accuracy for correct classification of cMCI subjects using scans from all time points in the feature selection procedure. (C) Features providing highest classification accuracy for correct prediction of conversion from MCI to AD (thresholded at 65% of maximum feature weight). Abbreviations: AD, Alzheimer's disease; cMCI, mild cognitive impairment (converters); SVM, support vector machine.

(Maldjian et al., 2003, 2004) to assign imaging results to anatomic labels.

2.6. Statistical analysis of behavioral and demographic data

We performed analyses of variance (ANOVAs) to compare all testing and training groups at baseline in terms of age and mini-mental state examination (MMSE, Folstein et al., 1975). For all ANOVAs that revealed significant between-group differences, we conducted post hoc Bonferroni *t* tests with a significance threshold of $p < 0.05$ for pairwise group comparisons. Group differences regarding sex were evaluated using a chi-square test for independent samples. The statistical analyses were performed with the software package SPSS 17.0 (http://www.spss.com/statistics/).

To investigate potential differences in the development of cognitive profiles (as measured by MMSE) between misclassified and correctly classified subjects (as classified by SVM) within each clinical group, we calculated repeated-measures ANOVAs with factor time as a within-group factor and classification label (correct vs. misclassified) as a between-group factor. Only time points for which MMSE was available for a sufficient number of subjects in both correctly and misclassified groups (number of subjects per condition >5) were used for these analyses. If an ANOVA revealed significant group or group-by-time differences, post hoc *t* tests were performed to compare the groups at each time point. As we expected that subjects misclassified as AD in both the control and ncMCI groups would show greater cognitive deficits and vice versa subjects classified as controls would express fewer cognitive deficits, a 1-sided significance threshold of $p < 0.05$ was applied for these analyses.

3. Results

3.1. Behavioral and demographic results

Groups used for testing and training did not differ in terms of age (Table 1). As expected, we observed significant differences in

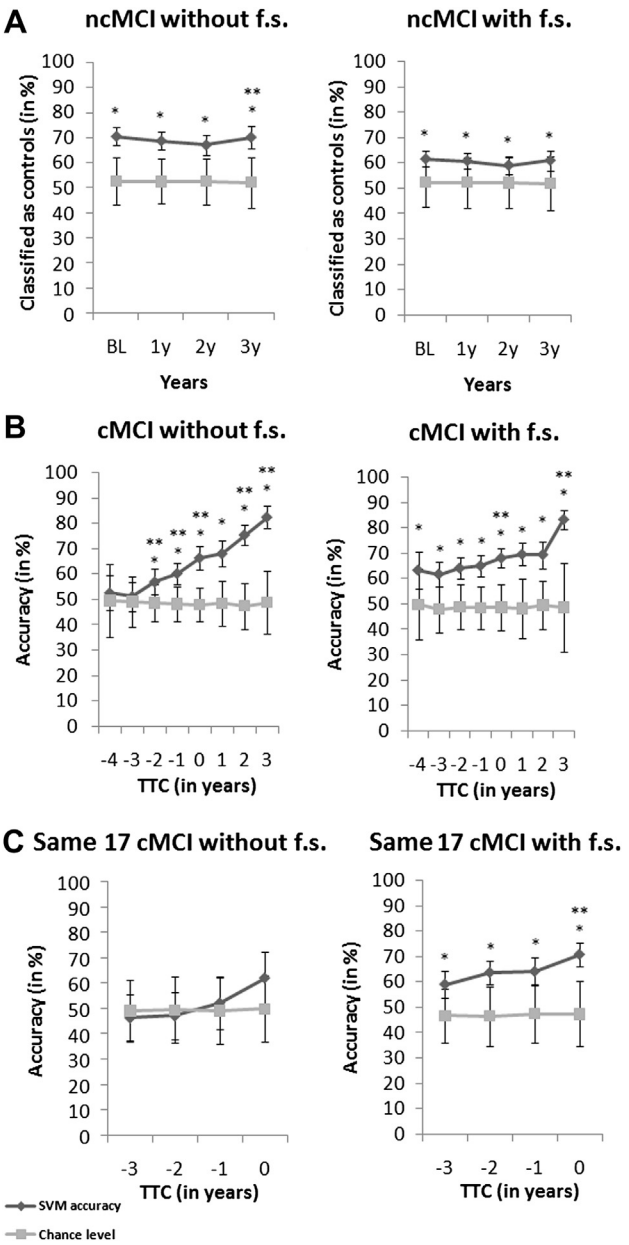

Fig. 3. (A) Classification accuracies for ncMCI subjects at different MRI scanning time points with and without feature selection. (B) Classification accuracies for all cMCI subjects at different MRI scanning time points with and without feature selection. (C) Classification accuracies for the 17 cMCI subjects for which the data were available for all 4 time points with and without feature selection. * Significant difference to chance level ($p < 0.05$, 1 tailed, Bonferroni corrected for multiple comparisons) and ** significant difference to the preceding time point ($p < 0.05$, 1 tailed, Bonferroni corrected for multiple comparisons). Abbreviations: cMCI, mild cognitive impairment (converters); f.s., feature selection; MRI, magnetic resonance imaging; ncMCI, mild cognitive impairment (nonconverters); SVM, support vector machine, TTC, time to conversion.

MMSE scores between all groups. Gender differed significantly between AD and both MCI groups and between controls and both MCI groups but not between AD and controls and not between the MCI groups. Post hoc t tests revealed significant differences in MMSE between healthy controls and all other clinical groups for both the test (healthy controls vs. AD: $t(94) = 17.12$, $p < 0.05$; healthy controls vs. cMCI: $t(182) = 9.41$, $p < 0.05$; healthy controls vs. ncMCI: $t(113) = 5.87$, $p < 0.05$) and training cohorts (healthy controls vs. AD: $t(106) = 19.51$, $p < 0.05$). cMCI and ncMCI differed significantly from each other ($t(201) = 2.81$, $p < 0.05$) and from AD in the test cohort (cMCI vs. AD: $t(194) = 11.40$, $p < 0.05$; ncMCI vs. AD: $t(113) = 10.97$, $p < 0.05$). No significant differences in MMSE were observed between test and training sets for AD patients (AD training vs. AD testing: $t(106) = 0.73$, $p > 0.99$) and for healthy controls ($t(94) = 1.41$, $p > 0.99$). The overall comparison revealed a significant difference in gender between groups ($\chi^2(5) = 16.80$, $p = 0.005$). However, no significant differences were observed for gender between AD patients and healthy controls in the training ($\chi^2(1) = 0.150$, $p = 0.699$) and testing cohorts ($\chi^2(1) = 0.113$, $p = 0.737$). Further, no differences were observed between cMCI and ncMCI ($\chi^2(1) = 1.343$, $p = 0.246$). Both cMCI and ncMCI differed significantly in gender distribution from healthy controls (cMCI vs. healthy controls: $\chi^2(1) = 6.94$, $p < 0.05$; ncMCI vs. healthy controls: $\chi^2(1) = 10.01$, $p < 0.05$) and AD patients (cMCI vs. AD: $\chi^2(1) = 6.02$, $p < 0.05$; ncMCI vs. AD: $\chi^2(1) = 9.08$, $p < 0.05$).

The ANOVAs investigating differences in MMSE between correctly and incorrectly classified subjects revealed significant differences in all clinical groups. Thus, both in the control ($F(1,40) = 4.1$, $p = 0.049$) and ncMCI ($F(1,59) = 4.7$, $p = 0.035$) groups, subjects classified as controls had higher MMSE values compared with those who were classified as AD. The opposite pattern was observed in cMCI ($F(1,158) = 6.0$, $p = 0.015$) and AD ($F(1,51) = 5.0$, $p = 0.03$), with subjects correctly classified as AD showing significantly lower MMSE values. Group-by-time interactions failed to reach the significance threshold in any group though showed a trend ($p < 0.1$) in control and cMCI groups. However, in the post hoc t tests, differences between correctly and incorrectly classified subjects in control and ncMCI groups were not significant at baseline but became significant on follow-up examinations (Fig. 1). In the AD and cMCI groups, differences between correctly and incorrectly classified subjects were significant at all time points except for time points 2 and 5 for cMCI.

3.2. SVM classification

The results of the feature selection procedure are displayed in Fig. 2. All results described as with feature selection are referring to the feature set providing the maximum classification accuracy in this feature selection procedure.

Whole-brain SVM classification yielded average (over all time points) diagnostic accuracies of 80.3%, 73.5%, and 63.7% for healthy controls, AD, and cMCI, respectively. In the ncMCI group, 69.0% have been classified as control subjects. At a TTC of 4 and 3 years, cMCI patients were classified at chance level (Fig. 3). At a TTC of 2 years, the accuracy increased above the chance level and kept increasing annually. In all other groups, accuracy was significantly above the chance level at all time points (Figs. 3 and 4).

Feature selection resulted in a feature set consisting of 814 voxels, corresponding to 0.21% of the initial brain volume, in bilateral hippocampus, amygdala, precentral gyrus, middle frontal gyrus, left inferior temporal gyrus, crus I, supramarginal gyrus, right calcarine sulcus, caudate nucleus, and cerebellum VI (Fig. 2C). Average classification accuracy over all time points using this feature set was significantly increased compared with classification without feature selection—82.0%, 76.6%, and 67.2% for healthy controls ($t(344) = 3.21$, $p < 0.05$), AD ($t(320) = 7.23$, $p < 0.05$), and cMCI ($t(1068) = 7.36$, $p < 0.05$), respectively. Mean assignment for ncMCI patients as control subjects significantly decreased to 60.6% ($t(452) = 23.64$, $p < 0.05$). With feature selection, the accuracy for cMCI at 62% was higher than chance level at a TTC of 4 years and was mostly stable at consecutive time points (Fig. 3). The classification accuracy for the 17 cMCI patients in the feature selection set was similar to that in the whole cMCI group.

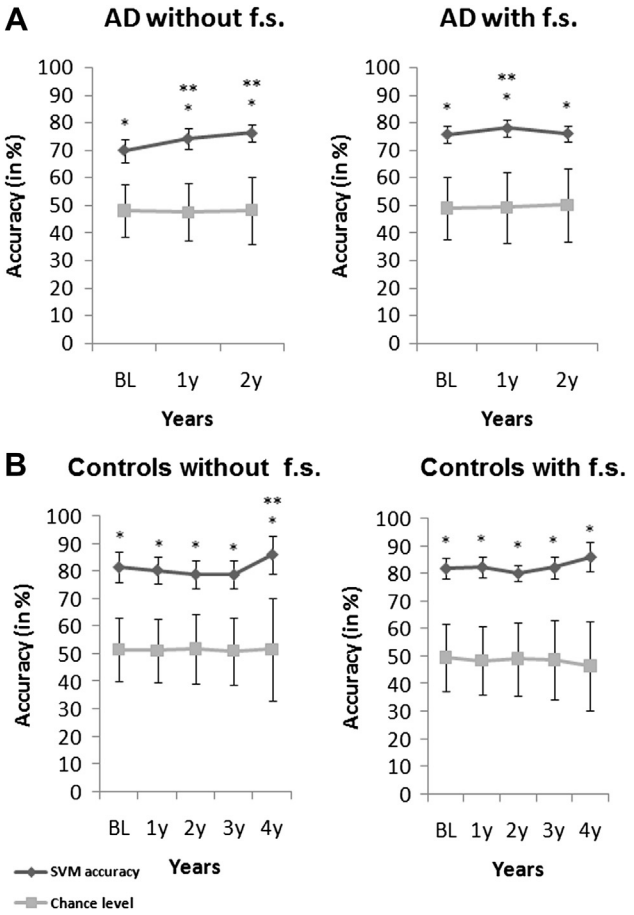

Fig. 4. (A) Classification accuracies for AD subjects at different MRI scanning time points with and without feature selection. (B) Classification accuracies for control subjects at different MRI scanning time points with and without feature selection. * Significant difference to chance level ($p < 0.05$, 1 tailed, Bonferroni corrected for multiple comparisons) and ** significant difference to the preceding time point ($p < 0.05$, 1 tailed, Bonferroni corrected for multiple comparisons). Abbreviations: AD, Alzheimer's disease; f.s., feature selection; MRI, magnetic resonance imaging.

After feature selection, diagnostic accuracy for AD significantly increased at baseline ($t(106) = 8.41$, $p < 0.05$) and 1-year follow-up ($t(104) = 5.61$, $p < 0.05$) compared with classification without feature selection (Fig. 4). For ncMCI, the percentage of assignment to the control group significantly decreased at all time points—baseline ($t(120) = 14.96$, $p < 0.05$), 1-year follow-up ($t(120) = 12.86$, $p < 0.05$), 2-year follow-up ($t(120) = 12.00$, $p < 0.05$), and 3-year follow-up ($t(86) = 10.20$, $p < 0.05$). For healthy controls, accuracy significantly increased only at 3-year follow-up ($t(74) = 3.26$, $p < 0.05$). For cMCI, significant increases in accuracy were observed at TTC of 4 years ($t(32) = 4.39$, $p < 0.05$), 3 years ($t(74) = 8.43$, $p < 0.05$), 2 years ($t(164) = 9.99$, $p < 0.05$), 1 year ($t(254) = 8.96$, $p < 0.05$), and at time point of conversion ($t(256) = 3.24$, $p < 0.05$). For the cMCI subgroup consisting of the same subjects at all time points, accuracy increased at all time points (TTC of 3 years: $t(32) = 4.92$, $p < 0.05$; TTC of 2 years: $t(32) = 6.60$, $p < 0.05$; TTC of 1 year: $t(32) = 4.16$, $p < 0.05$; time point of conversion: $t(32) = 3.22$, $p < 0.05$).

3.3. Voxel-based morphometry

The comparison of incorrectly classified cMCI to healthy controls revealed atrophy restricted to bilateral hippocampus, amygdala, ventral putamen, left thalamus, and right parahippocampal gyrus (Fig. 5A, Table 4). Correctly classified cMCI showed extensive atrophy extending to bilateral hippocampus, fusiform gyrus, superior, middle, and inferior temporal gyri, inferior parietal lobule, supramarginal gyrus, parahippocampal gyrus, insula, angular gyrus, amygdala, and putamen. Left-sided atrophy was observed in the postcentral gyrus. A direct comparison of correctly and incorrectly classified cMCI revealed significant atrophy in correctly classified cMCI in a temporoparietal-hippocampal network (Fig. 5B, Table 5). The opposite contrast revealed atrophy in misclassified cMCI in the cerebellum bilaterally.

In ncMCI classified as AD, we observed significant atrophy compared with healthy controls in bilateral hippocampus, amygdala, and parahippocampal gyrus and also atrophy restricted to the left hemisphere in fusiform gyrus and middle and inferior temporal gyri. ncMCI classified as controls showed no differences relative to healthy controls. A direct comparison in the ncMCI cohort revealed atrophy in ncMCI classified as AD in temporal and hippocampal regions. In the opposite contrast, we found calcarine sulcal atrophy in ncMCI classified as controls.

In the comparison of correctly classified AD to healthy controls, we observed significant atrophy in AD in bilateral hippocampus, parahippocampal gyrus, fusiform gyrus, superior and middle temporal gyri, supramarginal gyrus, superior and inferior parietal lobules, angular gyrus, and amygdala. Only left-hemispheric decreases were observed in the postcentral gyrus and ventral putamen. Right-hemispheric changes were restricted to the inferior temporal gyrus. Correctly classified AD showed greater atrophy compared with misclassified AD in the right inferior temporal and left lingual gyri. Misclassified AD showed atrophy in the right gyrus rectus only.

The comparison of correctly and incorrectly classified control subjects revealed an AD-typical pattern of atrophy in misclassified control subjects in a temporoparietal-hippocampal network. No other significant differences were observed.

4. Discussion

Here, we investigate the longitudinal changes in predictive value of computer-based diagnosis of AD. We detect systematic differences in classification accuracy of cMCI that were dependent to closeness to conversion time. Whole-brain SVM classification 3–4 years before conversion provides chance-level accuracy for AD detection. After feature selection restricting the analysis to structures of the medial temporal lobe and parietal cortex, we achieve accuracies significantly higher than chance level already 4 years before conversion. Our findings not only provide a novel perspective on the time- and state-dependent changes of diagnostic accuracy within the preconversion period from MCI to AD but also bring new evidence for potentially differential pathophysiological processes underlying neurodegeneration in clinically defined AD.

Our findings bring strong evidence that sMRI-based information can be used to predict conversion from cMCI to AD at disease stages earlier than previously suggested (Jack et al., 2010). On the one side, this demonstrates that already at early disease stages, sMRI data contain information on AD-related pathologic changes, which can be used for prediction of conversion from MCI to AD. On the other side, the relatively low accuracy of 62% obtained using an unbiased classifier, which was not optimized for MCI subjects or cross-validation, also suggests that interindividual variability in the sMRI measurements might be the impeding factor in our study preventing high diagnostic accuracies at this early disease stage. Using optimized preprocessing and classification methods and integration of longitudinal trajectories into the algorithms might provide remedy.

Additionally, we find strong evidence that AD-related structural changes can be observed before the occurrence of any significant

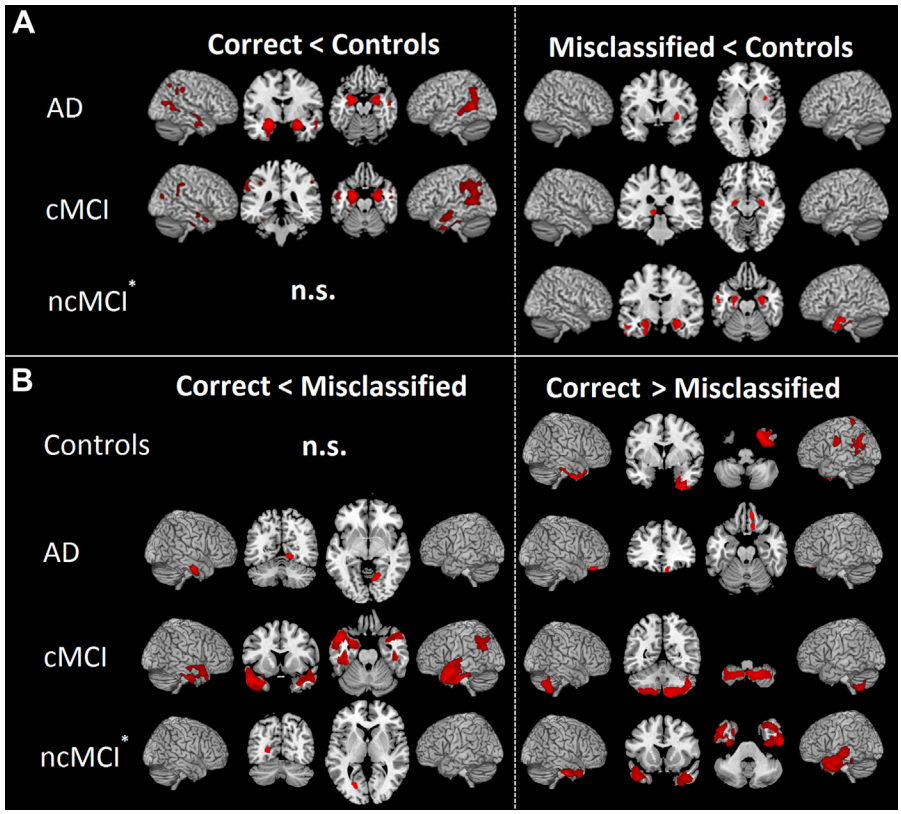

Fig. 5. Voxel-based morphometry results for comparison of correctly and incorrectly classified subjects from each clinical group to control subjects (A) and to each other (B). Results are displayed at a significance threshold of 0.001 uncorrected at voxel level with a cluster threshold of 100 voxels. * for ncMCI, the term "correct" refers to their classification as control subjects; however, the terms "correct" and "misclassified" are misleading for this cohort as the etiology of the symptoms is largely unknown. Abbreviations: AD, Alzheimer's disease; cMCI, mild cognitive impairment (converters); ncMCI, mild cognitive impairment (nonconverters); n.s., not significant.

cognitive deficit. This finding is in line with recent studies demonstrating an increased atrophy rate in cognitively normal subjects with high beta-amyloid deposition and greater atrophy in the right medial temporal lobe in cognitively normal subjects with future cognitive impairment (Chételat et al., 2012; Tondelli et al., 2012). Control subjects and ncMCI classified as AD demonstrate a pattern of structural GM abnormalities in a temporoparietal-hippocampal network that is very similar to the one previously reported in AD patients (Devanand et al., 2007; Frisoni et al., 2002; Schroeter et al., 2009). Corroborating these structural observations, we find statistically significant yet subclinical cognitive deficits in control and ncMCI subjects classified as AD compared with those classified as controls. These differences increase over time without reaching the significance threshold of a formal interaction analysis. However, the observation of AD-typical structural changes combined with a decline in cognitive performance in control subjects and in ncMCI patients strongly supports the conjecture that AD-related pathology can be detectable at much earlier time points than proposed in current models of AD progression (Jack et al., 2010). The finding of AD-like atrophy patterns linked with faster cognitive decline in control subjects and ncMCI patients classified as AD is in line with previous studies performed by Davatzikos et al. (2009, 2011), who demonstrated faster cognitive decline in control and ncMCI subjects with an AD-like atrophy pattern.

We further demonstrate that detection accuracy of conversion from MCI to AD significantly increases with closeness to the time point of conversion. Interestingly, classification after feature selection provides accuracies higher than chance level 4 years before conversion opposed to classification of whole-brain data reaching similar level of accuracy at earliest at the time point of clinical conversion. The localization and spatial extent of the anatomic feature set are in line with previous research showing the precedence of pathologic changes in hippocampus and parietal cortex and relative sparing of cerebellum (Braak and Braak, 1991; Jack et al., 2010; Schroeter et al., 2009). These studies have shown that beta-amyloid and tau pathologies start focally in AD and spread with time to other areas of GM whereas the further accumulation of histopathologic changes in already affected regions slows with time. Atrophy is assumed to follow this spreading pattern of beta-amyloid and tau depositions with a delay of several years. Based on these findings, one would expect that atrophy is restricted to a few brain regions in earlier AD stages. Other parts of the brain are not expected to contain relevant information for disease detection at this early stage. Given that the standard SVM assigns a weight to all—including noninformative—features, the inclusion of a large number of features (as performed in a whole-brain approach) may decrease the signal-to-noise ratio and therewith lower the possible classification accuracy. Feature selection at this early AD stage will remove noisy features and so increase classification accuracy. In contrast, at later AD stages, atrophy involves large parts of cortex and so all brain regions carry some information that are relevant for discrimination between AD patients and healthy control subjects. Performing feature selection at this stage will lead to removal of informative features and may decrease classification accuracy. Both effects are observed in our study: the gain achieved in cMCI using feature selection rapidly decreases with closeness to the conversion time point. Correspondingly, feature selection in manifest AD and also in healthy control subjects adds little, giving a similar accuracy to that with a whole-brain approach. This observation is consistent with the literature evaluating the effect of feature selection on

Table 4
Voxel-based morphometry results for comparisons with control subjects

Region	Side	MNI coordinates			Cluster size	Peak *T* value
		x	*y*	*z*		
AD correct < controls						
Amygdala, fusiform gyrus, hippocampus, parahippocampal gyrus, putamen, middle and superior temporal poles	L	−24	−5	−16	2099	5.65
Amygdala, fusiform gyrus, hippocampus, parahippocampal gyrus, middle and superior temporal poles	R	28	−8	−18	1611	5.30
Angular gyrus, supramarginal gyrus, middle temporal gyrus, superior temporal gyrus	L	−58	−54	9	3437	5.01
Middle occipital and middle temporal gyri	R	40	−74	16	715	4.61
AD misclassified < controls						
Putamen	R	31	3	−6	270	4.07
cMCI correct < controls						
Amygdala, fusiform gyrus, hippocampus, insula, middle occipital, olfactory cortex, parahippocampal gyrus	L	−26	−8	18	4486	6.52
Amygdala, hippocampus, parahippocampal gyrus	R	27	−8	18	2793	6.12
Angular gyrus, supramarginal gyrus	L	−52	−59	29	3537	5.00
Middle temporal gyrus	L	−54	−12	−21	940	4.21
Middle occipital gyrus	R	44	−77	18	109	3.92
Fusiform gyrus, inferior temporal gyrus	R	42	−17	−31	364	3.87
Superior temporal pole	R	54	−9	−15	383	3.60
cMCI misclassified < controls						
Amygdala, hippocampus, parahippocampal gyrus	R	28	−11	−18	701	4.36
Thalamus	L	−12	−30	0	332	4.00
Amygdala, hippocampus	L	−27	−11	−10	293	3.86
ncMCI misclassified < controls						
Inferior temporal gyrus, middle temporal gyrus	L	−44	1	−45	1140	4.28
Amygdala, hippocampus, parahippocampal gyrus	R	30	−9	−19	602	4.13
Amygdala, fusiform gyrus, hippocampus, parahippocampal gyrus	L	−28	−12	−18	1016	4.09

Key: AD, Alzheimer's disease; cMCI, mild cognitive impairment (converters); L, left; MNI, Montreal Neurological Institute; ncMCI, mild cognitive impairment (nonconverters); R, right.

classification accuracy (Chu et al., 2012) and studies reporting similar accuracies for differentiation between AD patients and healthy controls using sMRI with and without feature selection (Dukart et al., 2011a; Gerardin et al., 2009; Hinrichs et al., 2009; Kloppel et al., 2008a, 2008b). Our result therefore emphasizes the importance of feature selection procedures at the MCI stage of early incipient AD using sMRI.

Another finding is that there is a significant reduction in assignment of ncMCI to healthy controls using feature selection. In general, all MCI patients from the ADNI cohort used in our study correspond to the amnestic MCI type. Previous studies with a follow-up of up to 10 years have shown that at longer follow-up, most of the amnestic MCI convert to AD (Fischer et al., 2007; Visser et al., 2006). It is therefore likely that a 4-year follow-up is insufficient to detect all cases of premanifest AD in the group. Given the strong boost in accuracy we observe after feature selection at the MCI stage of early incipient AD, we would expect the classifier trained on selected features to be much more sensitive to potential AD also in the ncMCI cohort. Supportive of this reasoning is our VBM result comparing ncMCI classified as AD with healthy controls. This reveals a GM atrophy pattern that is very similar to that observed in correctly classified AD and cMCI. An evaluation of assignment of ncMCI patients with a short follow-up as controls is therefore highly problematic in terms of interpretation.

We further detect substantial differences between correctly and incorrectly classified control subjects and MCI and AD patients using the VBM approach. Both correctly classified AD and cMCI show extensive atrophy in a temporoparietal-hippocampal network, whereas only minor changes restricted to the hippocampus are detected in misclassified patients. The finding of a temporoparietal-hippocampal atrophic network is consistent with a similar set of regions reported for cMCI and AD in the literature (Karow et al., 2010; Schroeter et al., 2009). However, cumulative evidence shows that current diagnostic selection criteria do not discriminate accurately between AD and other types of dementia (Dubois et al., 2007; Varma et al., 1999) and that hippocampal atrophy is also found in other dementia syndromes such as frontotemporal lobar degeneration (Barnes et al., 2006, 2007; van de Pol et al., 2006). Taking both findings into account questions the reliability of selection criteria used in the ADNI cohort and of suggested diagnostic algorithms that use only hippocampal atrophy as a biomarker for AD (Devanand et al., 2007; Laakso et al., 1995; Morra et al., 2009a, 2009b). Indeed, the diagnostic accuracy we achieve for AD of 80.3% is compatible with results published in pathologic studies of patients dying with a clinical diagnosis of AD and differ from the SVM-based classification accuracy of ∼95% when scans from pathologically verified cases of AD were used (Kloppel et al., 2008b). Given very recent pathologic-clinical diagnostic results, this is clearly an important area for further research (Beach et al., 2012).

As previously shown, sMRI can be used to predict conversion from MCI to AD (Devanand et al., 2007; McEvoy et al., 2009; Plant et al., 2010) and even to distinguish between cMCI and ncMCI with reasonable accuracy (Ewers et al., 2010; Hinrichs et al., 2011). However, in these studies, cMCI patients were pooled over all time points and so obscuring the influence of TTC on diagnostic accuracy. Additionally, disease-related processes have been shown to influence the estimation of healthy aging and vice versa—healthy aging has been shown to interact with disease detection (Dukart et al., 2011b; Franke et al., 2010). For these reasons, we removed the voxel-wise variance explained by healthy aging from all imaging data based on estimates from an independent healthy cohort (Dukart et al., 2011b). All observed changes in accuracy estimates over time can therefore be attributed to disease-related pathologic processes.

The average accuracies obtained in our study for detection of AD and healthy controls and of conversion from MCI to AD are comparable with those reported in the literature. However, most previous studies applied feature selection algorithms optimizing the cross-validation performance within their own datasets. These types of algorithms carry the risk of overestimating the conversion from MCI to AD detection accuracy because of multiple testing and

Table 5
Voxel-based morphometry results for comparisons between correctly and incorrectly classified subjects

Region	Side	MNI coordinates			Cluster size	Peak *T* value
		x	*y*	*z*		
Controls correct > controls misclassified						
Hippocampus, parahippocampal gyrus, amygdala, fusiform gyrus, superior and middle temporal poles, inferior temporal gyrus	R	23	8	−47	3409	4.9
Middle occipital gyrus, inferior parietal lobule, supramarginal gyrus, angular gyrus, superior and middle temporal gyri	L	−45	−60	38	1196	4.5
Postcentral gyrus, superior parietal lobule, precuneus, paracentral lobule	L	−15	−47	66	367	4.3
Precentral gyrus, postcentral gyrus	L	−63	−11	32	346	3.9
Postcentral gyrus, superior and inferior parietal lobules	L	−36	−39	45	254	4.2
Fusiform gyrus, inferior temporal gyrus	R	60	−27	−30	242	3.8
cMCI correct < cMCI misclassified						
Rolandic operculum, olfactory cortex, insula, hippocampus, parahippocampal gyrus, amygdala, fusiform gyrus, superior, middle, and inferior temporal gyri, superior and middle temporal poles	L	−36	−15	−35	10,064	5.4
Insula, parahippocampal gyrus, amygdala, superior, middle, and inferior temporal gyri, superior and middle temporal poles	R	35	15	−32	2355	4.4
Fusiform gyrus, middle temporal gyrus, middle temporal pole, inferior temporal gyrus	R	48	−21	−24	1481	4.6
Inferior parietal lobule, supramarginal gyrus, angular gyrus, middle temporal gyrus	L	−41	−71	48	1425	4.1
cMCI correct > cMCI misclassified						
R crus I of cerebellar hemisphere, L/R crus II of cerebellar hemisphere, R lobule VI of cerebellar hemisphere, R/L lobule VIIb, VIII, and IX of cerebellar hemisphere	L/R	3	−48	−59	4679	4.6
Lobule VI of cerebellar hemisphere	R	23	−65	−29	209	3.8
AD correct < AD misclassified						
Middle and inferior temporal gyri, inferior temporal gyrus	R	48	−11	−29	448	4.6
Lingual gyrus	R	12	−62	−6	334	4.4
AD correct > AD misclassified						
Gyrus rectus	R	9	42	−19	464	3.68
ncMCI correct < ncMCI misclassified						
Calcarine sulcus, lingual gyrus	L	−22	−74	3	210	3.83
ncMCI correct > ncMCI misclassified						
Parahippocampal gyrus, amygdala, fusiform gyrus, superior and middle temporal poles, middle and inferior temporal gyri	R	54	−12	−42	3278	5.49
Hippocampus, parahippocampal gyrus, amygdala, fusiform gyrus, superior temporal gyrus, superior middle and inferior temporal poles, middle temporal pole	L	−48	−2	−28	7796	5.34

Key: AD, Alzheimer's disease; cMCI, mild cognitive impairment (converters); L, left; MNI, Montreal Neurological Institute; ncMCI, mild cognitive impairment (nonconverters); R, right.

potential overfitting of selected features to a specific dataset. The main goal of our study was not the further development and optimization of current classification algorithms but rather to provide evidence that sMRI-based classification is able to detect AD substantially before conversion and to characterize the evolution of detection accuracy. We therefore avoided in our selection and classification procedures most potential sources of overfitting by using independent training and test datasets and by avoiding parameter optimization to improve the classification accuracy of test data. Similarly, previous studies indicated that training cohort is an important factor influencing achievable accuracy (Abdulkadir et al., 2011; Casanova and Hsu, 2012; Chu et al., 2012). We would therefore expect that the accuracy we obtained for classification of all clinical groups could be further increased by using larger training datasets.

Another related issue is the use of only AD patients in the training dataset. It is likely that higher accuracies could have been achieved for early AD detection if only cMCI subjects had been used as a training dataset. However, from our point of view, it is very important for clinical practice to establish a single classifier that will be capable to detect AD irrespective of stage. We therefore avoided training a separate classifier for application only on MCI subjects. Nonetheless, a classifier built on already manifested AD is also likely not to be the best for AD detection in all its stages. The issue of establishing an optimum classifier with maximum accuracy for early AD detection should be addressed in future research, for example, by integrating both cMCI and AD patients in the same training dataset to allow SVM or other techniques to detect a stage-independent AD pattern.

A potential source of limitation in our study that also needs consideration is that all the data used for training and testing, although treated independently, were extracted from the same cohort (ADNI). Any potential selection bias inherent to the ADNI cohort could therefore limit the generalizability of our results to other sources of patients.

In summary, our study provides first evidence that sMRI might contain information that can be used for detection of AD already at early disease stages—4 years before conversion from MCI. We further demonstrate that control subjects and ncMCI that are classified as AD using automated computer-based diagnostics show preclinical cognitive deficits align with an atrophy pattern that is characteristic for AD. Both findings are supportive for the capability of automated machine learning algorithms based on sMRI data to detect a possible incipient AD.

Disclosure statement

The authors have no conflicts of interest relevant to the subject of the manuscript, including no institutional contracts relating to this research or any other agreements of the authors or their institutions that could be seen as involving a financial interest in this work.

Acknowledgements

S.A. is supported by the Scientific Exchange Program Sciex-NMS-ch (Scientific Exchange Programme between the New Member States and Switzerland). J.D. is supported by the Swiss National Science Foundation (National Centres of Competence in Research [NCCR] Synapsy). B.D. is supported by the Swiss National Science Foundation (Project Grant Nr 320030_135679, NCCR Synapsy, and SPUM 33CM30_140332/1), Foundation Parkinson Switzerland, Foundation

Synapsis, Novartis Foundation for Medical-Biological Research, and Deutsche Forschungsgemeinschaft (Kfo 247). F.K. is supported by the Velux Stiftung. Data collection and sharing for this project were 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, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Amorfix Life Sciences Ltd, AstraZeneca, Bayer HealthCare, BioClinica, Inc, Biogen Idec, Inc, Bristol-Myers Squibb Company, Eisai, Inc, Elan Pharmaceuticals, Inc, Eli Lilly and Company, F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc, GE Healthcare, Innogenetics, N.V., IXICO Ltd, Janssen Alzheimer Immunotherapy Research and Development, LLC, Johnson & Johnson Pharmaceutical Research and Development LLC, Medpace, Inc, Merck & Co, Inc, Meso Scale Diagnostics, LLC, Novartis Pharmaceuticals Corporation, Pfizer, Inc, Servier, Synarc, Inc, and Takeda Pharmaceutical Company. 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, revised March 26, 2012, coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of California, Los Angeles. This research was also supported by National Institutes of Health grants P30 AG010129 and K01 AG030514.

References

Abdulkadir, A., Mortamet, B., Vemuri, P., Jack, C.R., Krueger, G., Klöppel, S., Initiative, A.s.D.N., 2011. Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. Neuroimage 58, 785–792. http://www.S1053-8119(11)00647-1 [pii]10.1016/j.neuroimage.2011.06.029.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Barnes, J., Godbolt, A.K., Frost, C., Boyes, R.G., Jones, B.F., Scahill, R.I., Rossor, M.N., Fox, N.C., 2007. Atrophy rates of the cingulate gyrus and hippocampus in AD and FTLD. Neurobiol. Aging 28, 20–28.

Barnes, J., Whitwell, J.L., Frost, C., Josephs, K.A., Rossor, M., Fox, N.C., 2006. Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. Arch. Neurol. 63, 1434–1439.

Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W., 2012. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J. Neuropathol. Exp. Neurol. 71, 266–273.

Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol. 82, 239–259.

Casanova, R., Hsu, F.-C., 2012. Classification of structural MRI images in Alzheimer's disease from the perspective of ill-posed problems. PloS One 7, e44877.

Chang, C., Lin, C. 2001. LIBSVM: a library for support vector machines. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm. Accessed January 8, 2012.

Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., 2012. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage 60, 59–70.

Chételat, G., Villemagne, V.L., Villain, N., Jones, G., Ellis, K.A., Ames, D., Martins, R.N., Masters, C.L., Rowe, C.C., Group, A.R., 2012. Accelerated cortical atrophy in cognitively normal elderly with high β-amyloid deposition. Neurology 78, 477–484.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging 32, 2322.e19–2322.e27.

Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M., 2008a. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol. Aging 29, 514–523.

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

Davatzikos, C., Xu, F., An, Y., Fan, Y., 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.

Devanand, D.P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G.H., Honig, L.S., Mayeux, R., Stern, Y., Tabert, M.H., de Leon, M.J., 2007. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. Neurology 68, 828–836.

Dickerson, B.C., Wolk, D.A., 2012. MRI cortical thickness biomarker predicts AD-like CSF and cognitive decline in normal adults. Neurology 78, 84–90.

Dougherty, E.R., Sima, C., Hanczar, B., Braga-Neto, U.M., 2010. Performance of error estimators for classification. Curr. Bioinform. 5, 53–67.

Dubois, B., Feldman, H.H., Jacova, C., Dekosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., Scheltens, P., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 6, 734–746.

Dukart, J., Mueller, K., Barthel, H., Villringer, A., Sabri, O., Schroeter, M.L., Initiative, A.D.N., 2012. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatry Res., 10.1016/j.pscychresns.2012.04.007. http://www.S0925-4927(12)00085-6.

Dukart, J., Mueller, K., Horstmann, A., Barthel, H., Möller, H.E., Villringer, A., Sabri, O., Schroeter, M.L., 2011a. Combined evaluation of FDG-PET and MRI improves detection and differentiation of dementia. PLoS One 6, e18111.

Dukart, J., Schroeter, M.L., Mueller, K., Initiative, A.D.N., 2011b. Age correction in dementia—matching to a healthy brain. PLoS One 6, e22193.

Ewers, M., Walsh, C., Trojanowski, J.Q., Shaw, L.M., Petersen, R.C., Jack, C.R., Feldman, H.H., Bokde, A.L., Alexander, G.E., Scheltens, P., Vellas, B., Dubois, B., Weiner, M., Hampel, H., Alzheimer's Disease Neuroimaging Initiative (ADNI), 2010. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol. Aging 33, 1203–1214.

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.

Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., Krampla, W., Tragl, K.H., 2007. Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. Neurology 68, 288–291.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.

Franke, K., Ziegler, G., Kloppel, S., Gaser, C., 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. Neuroimage 50, 883–892.

Frisoni, G.B., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., Laakso, M.P., 2002. Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. J. Neurol. Neurosurg. Psychiatry 73, 657–664.

Gerardin, E., Chetelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehericy, S., Garnero, L., Eustache, F., Colliot, O., 2009. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 47, 1476–1486.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., Alzheimer's Disease Neuroimaging Initiative (ADNI), 2009. Spatially augmented LP boosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48, 138–149.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., Initiative, A.D.N., 2011. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55, 574–589.

Jack, C.R., Knopman, D.S., 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.

Jain, A.K., Dubes, R.C., Chen, C.-C., 1987. Bootstrap techniques for error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 9, 628–633.

Karow, D.S., McEvoy, L.K., Fennema-Notestine, C., Hagler, D.J., Jennings, R.G., Brewer, J.B., Hoh, C.K., Dale, A.M., Alzheimer's Disease Neuroimaging Initiative (ADNI), 2010. Relative capability of MR imaging and FDG PET to depict changes associated with prodromal and early Alzheimer disease. Radiology 256, 932–942.

Kloppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., Mader, I., Mitchell, L.A., Patel, A.C., Roberts, C.C., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S., 2008a. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. Brain 131 (Pt 11), 2969–2974.

Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S., 2008b. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (Pt 3), 681–689.

Laakso, M.P., Soininen, H., Partanen, K., Helkala, E.L., Hartikainen, P., Vainio, P., Hallikainen, M., Hänninen, T., Riekkinen, P.J., 1995. Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer's disease: correlation with memory functions. J. Neural. Transm. Park. Dis. Dement. Sect. 9, 73–86.

Maldjian, J.A., Laurienti, P.J., Burdette, J.H., 2004. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. Neuroimage 21, 450–455.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19, 1233–1239.

McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M., Initiative, A.D.N., 2009. Alzheimer disease:

quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 251, 195–205.
McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–944.
Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage 44, 1415–1422.
Modrego, P.J., 2006. Predictors of conversion to dementia of probable Alzheimer type in patients with mild cognitive impairment. Curr. Alzheimer Res. 3, 161–170.
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., Schuff, N., Weiner, M.W., Thompson, P.M., Initiative, A.D.N., 2009a. Automated 3D mapping of hippocampal atrophy and its clinical correlates in 400 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Hum. Brain Mapp. 30, 2766–2788.
Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Toga, A.W., Jack, C.R., Schuff, N., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging Initiative (ADNI), 2009b. 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 1), S3–S15.
Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 50, 162–174.
Quiroz, Y.T., Stern, C.E., Reiman, E.M., Brickhouse, M., Ruiz, A., Sperling, R.A., Lopera, F., Dickerson, B.C., 2012. Cortical atrophy in presymptomatic Alzheimer's disease presenilin 1 mutation carriers. J. Neurol. Neurosurg. Psychiatry 84, 556–561.
Schroeter, M.L., Stein, T., Maslowski, N., Neumann, J., 2009. Neural correlates of Alzheimer's disease and mild cognitive impairment: a systematic and quantitative meta-analysis involving 1351 patients. Neuroimage 47, 1196–1206.
Smith, E.E., Egorova, S., Blacker, D., Killiany, R.J., Muzikansky, A., Dickerson, B.C., Tanzi, R.E., Albert, M.S., Greenberg, S.M., Guttmann, C.R., 2008. Magnetic resonance imaging white matter hyperintensities and brain volume in the prediction of mild cognitive impairment and dementia. Arch. Neurol. 65, 94.
Tondelli, M., Wilcock, G.K., Nichelli, P., De Jager, C.A., Jenkinson, M., Zamboni, G., 2012. Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. Neurobiol. Aging 33, 825.e25–825.e36.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.
van de Pol, L.A., Hensel, A., van der Flier, W.M., Visser, P.J., Pijnenburg, Y.A., Barkhof, F., Gertz, H.J., Scheltens, P., 2006. Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. J. Neurol. Neurosurg. Psychiatry 77, 439–442.
Varma, A.R., Snowden, J.S., Lloyd, J.J., Talbot, P.R., Mann, D.M., Neary, D., 1999. Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. J. Neurol. Neurosurg. Psychiatry 66, 184–188.
Visser, P.J., Kester, A., Jolles, J., Verhey, F., 2006. Ten-year risk of dementia in subjects with mild cognitive impairment. Neurology 67, 1201–1207.