NeuroImage 56 (2011) 766–781

Contents lists available at ScienceDirect

NeuroImage

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

Supplement

Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database

Rémi Cuingnet [a,b,c,d,*], Emilie Gerardin [a,b,c], Jérôme Tessieras [a,b,c], Guillaume Auzias [a,b,c], Stéphane Lehéricy [a,b,c,e], Marie-Odile Habert [d,f], Marie Chupin [a,b,c], Habib Benali [d], Olivier Colliot [a,b,c] and The Alzheimer's Disease Neuroimaging Initiative [1]

[a] UPMC Université Paris 6, UMR 7225, UMR_S 975, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), Paris, F-75013, France
[b] CNRS, UMR 7225, CRICM, Paris, F-75013, France
[c] Inserm, UMR_S 975, CRICM, Paris, F-75013, France
[d] Inserm, UMR_S 678, LIF, Paris, F-75013, France
[e] Centre for NeuroImaging Research, CENIR, Department of Neuroradiology, Groupe hospitalier Pitié-Salpêtrière, Paris, F-75013, France
[f] AP-HP, Department of Nuclear Medicine, Groupe hospitalier Pitié-Salpêtrière, Paris, F-75013, France

ARTICLE INFO

Article history:
Received 27 November 2009
Revised 31 May 2010
Accepted 5 June 2010
Available online 11 June 2010

Keywords:
Alzheimer's disease
AD
MCI
Converter
Prodromal
Classification
Magnetic resonance imaging
Support vector machines

ABSTRACT

Recently, several high dimensional classification methods have been proposed to automatically discriminate between patients with Alzheimer's disease (AD) or mild cognitive impairment (MCI) and elderly controls (CN) based on T1-weighted MRI. However, these methods were assessed on different populations, making it difficult to compare their performance. In this paper, we evaluated the performance of ten approaches (five voxel-based methods, three methods based on cortical thickness and two methods based on the hippocampus) using 509 subjects from the ADNI database. Three classification experiments were performed: CN *vs* AD, CN *vs* MCIc (MCI who had converted to AD within 18 months, MCI converters — MCIc) and MCIc *vs* MCInc (MCI who had not converted to AD within 18 months, MCI non-converters — MCInc). Data from 81 CN, 67 MCInc, 39 MCIc and 69 AD were used for training and hyperparameters optimization. The remaining independent samples of 81 CN, 67 MCInc, 37 MCIc and 68 AD were used to obtain an unbiased estimate of the performance of the methods. For AD *vs* CN, whole-brain methods (voxel-based or cortical thickness-based) achieved high accuracies (up to 81% sensitivity and 95% specificity). For the detection of prodromal AD (CN *vs* MCIc), the sensitivity was substantially lower. For the prediction of conversion, no classifier obtained significantly better results than chance. We also compared the results obtained using the DARTEL registration to that using SPM5 unified segmentation. DARTEL significantly improved six out of 20 classification experiments and led to lower results in only two cases. Overall, the use of feature selection did not improve the performance but substantially increased the computation times.

Introduction

Alzheimer's disease (AD) is the most frequent neurodegenerative dementia and a growing health problem. Definite diagnosis can only be made postmortem, and requires histopathological confirmation of amyloid plaques and neurofibrillary tangles. Early and accurate diagnosis of Alzheimer's Disease (AD) is not only challenging, but is crucial in the perspective of future treatments. Clinical diagnostic criteria are currently based on the clinical examination and neuropsychological assessment, with the identification of dementia and then of the Alzheimer's phenotype (Blennow et al., 2006). Patients suffering from AD at a prodromal stage are, mostly, clinically classified as amnestic mild cognitive impairment (MCI) (Petersen et al., 1999; Dubois and Albert, 2004), but not all patients with amnestic MCI will develop AD. Recently, more precise research criteria were proposed for the early diagnostic of AD at the prodromal stage of the disease (Dubois et al., 2007). These criteria are based on a clinical core of early episodic memory impairment and the presence of at least one additional supportive feature including abnormal MRI and PET neuroimaging or abnormal cerebrospinal fluid amyloid and tau biomarkers (Dubois et al., 2007). Neuroimaging therefore adds a positive predictive value to the diagnosis and includes measurements using structural MRI to assess medial temporal lobe atrophy and

* Corresponding author. CRICM, Equipe Cogimage (ex LENA), Hôpital de la Pitié-Salpêtrière, 47, boulevard de l'Hôpital, 75651 Paris Cedex 13, France.
E-mail address: remi.cuingnet@gmail.com (R. Cuingnet).

[1] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI Author ship list.pdf).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.06.013

positron emission tomography using fluorodeoxyglucose (FDG) or amyloid markers (Fox and Schott, 2004; Jagust, 2006).

Many group studies based on volumetric measurements of regions of interest (ROI) (Convit et al., 1997, 2000; Jack et al., 1997,1998; Juottonen et al., 1998; Laakso et al., 1998, 2000; Busatto et al., 2003; Xu et al., 2000; Good et al., 2002; Chételat and Baron, 2003; Rusinek et al., 2004; Tapiola et al., 2008), voxel-based morphometry (Good et al., 2002; Busatto et al., 2003; Karas et al., 2003, 2004; Chételat et al., 2005; Whitwell et al., 2007, 2008) or group comparison of cortical thickness (Thompson et al., 2001, 2003, 2004; Lerch et al., 2005, 2008; Bakkour et al., 2009, Dickerson et al., 2009; Hua et al., 2009; McDonald et al., 2009) have shown that brain atrophy in AD and prodromal AD is spatially distributed over many brain regions including the entorhinal cortex, the hippocampus, lateral and inferior temporal structures, anterior and posterior cingulate. However these analyses measure group differences and thus are of limited value for individual diagnosis.

Advances in statistical learning with the development of new machine learning algorithms capable of dealing with high dimensional data, such as the support vector machine (SVM) (Vapnik, 1995; Shawe-Taylor and Cristianini, 2000; Schölkopf and Smola, 2001), enable the development of new diagnostic tools based on T1-weighted MRI. Recently, several approaches have been proposed to automatically classify patients with AD and/or MCI from anatomical MRI (Fan et al., 2005, 2007, 2008a,b; Colliot et al., 2008; Davatzikos et al., 2008a,b; Klöppel et al., 2008; Vemuri et al., 2008; Chupin et al., 2009a,b; Desikan et al., 2009; Gerardin et al., 2009; Hinrichs et al., 2009; Magnin et al., 2009; Misra et al., 2009; Querbes et al., 2009). These approaches could have the potential to assist in the early diagnosis of AD. These approaches can roughly be grouped into three different categories, depending on the type of features extracted from the MRI (voxel-based, vertex-based or ROI-based). In the first category, the features are defined at the level of the MRI voxel. Specifically, the features are the probability of the different tissue classes (grey matter, white matter and cerebrospinal fluid) in a given voxel (Lao et al., 2004; Fan et al., 2007, 2008a,b; Davatzikos et al., 2008a,b; Klöppel et al., 2008; Vemuri et al., 2008; Hinrichs et al., 2009; Magnin et al., 2009; Misra et al., 2009). Klöppel et al. (2008) directly classified these features with an SVM. All other methods first reduce the dimensionality of the feature space relying on different types of features extraction, agglomeration and/or selection methods. Vemuri et al. (2008) used smoothing, voxel-downsampling, and then a feature selection step. Another solution is to group voxels into anatomical regions through the registration of a labeled atlas (Lao et al., 2004; Ye et al., 2008; Magnin et al., 2009). However, this anatomical parcellation may not be adapted to the pathology. In order to overcome this limitation, Fan et al. (2007) have proposed an adaptive parcellation approach in which the image space is divided into the most discriminative regions. This method has been used in several studies (Davatzikos et al., 2008a,b; Fan et al., 2008a,b; Misra et al., 2009). In the second category, the features are defined at the vertex-level on the cortical surface (Desikan et al., 2009; Querbes et al., 2009). The methods of the third category include only the hippocampus. Their approach is based on the analysis of the volume and/or shape of the hippocampus (Colliot et al., 2008, Chupin et al., 2009a,b; Gerardin et al., 2009).

These approaches achieve high accuracy (over 84%). However, they were evaluated on different study populations, making it difficult to compare their respective discriminative power. Indeed, many factors such as degree of impairment, age, gender, genotype, educational level and MR image quality perceptibly affect the evaluation of the prediction accuracy. This variability between evaluations is increased for statistical reasons when the number of subjects is small. Therefore a meta-analysis would be of limited value to compare the prediction accuracies of different methods.

The goal of this paper was to compare different methods for the classification of patients with AD based on anatomical MRI, using the same study population. To that purpose, we used the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Ten methods were evaluated. We tested five voxel-based approaches: a direct approach (Klöppel et al., 2008), an approach based on a volume of interest (Klöppel et al., 2008), an atlas-based approach (Magnin et al, 2009) and the approaches proposed by Vemuri et al. (2008) and Fan et al. (2008a,b) respectively. In order to assess the influence of the registration step and the features used on the classification accuracies, these latter methods were tested with two different registration steps: SPM5 (Ashburner and Friston, 2005) and DARTEL (Ashburner, 2007) and also with either only the grey matter (GM) probability maps or all the tissues probability maps including also white matter (WM) and cerebrospinal fluid (CSF). Three cortical approaches were evaluated as well: a direct one similar to (Klöppel et al., 2008), an atlas based one and an approach using only the regions found in (Desikan et al., 2009). Two methods respectively based on the volume (Colliot et al., 2008, Chupin et al., 2009a,b) and the shape (Gerardin et al., 2009) of the hippocampus were also tested.

Materials

Data

Data used in the preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

MRI acquisition

The MR scans are T1-weighted MR images. MRI acquisition had been done according to the ADNI acquisition protocol in (Jack et al., 2008). For each subject, we used the MRI scan from the baseline visit when available and from the screening visit otherwise. We only used images acquired at 1.5 T. To enhance standardization across sites and platforms of images acquired in the ADNI study, pre-processed images that have undergone some post-acquisition correction of certain image artifacts are available (Jack et al., 2008). We used those corrected in image geometry for gradient nonlinearity and corrected for intensity non-uniformity due to non-uniform receiver coil sensitivity. The image geometry correction was the 3D gradwarp correction (Hajnal et al., 2001; Jovicich et al. 2006). The B1 non-uniformity correction is detailed in Narayana et al. (1988). These two preprocessing steps can be performed directly on the MRI console and thus seem feasible in clinical routine. All subjects were scanned twice at each visit. As explained in Jack et al.(2008), MR scans were graded qualitatively by the ADNI investigators of the ADNI MRI quality control center at the Mayo Clinic for artifacts and general image quality. Each scan was graded on several separate criteria: blurring/ghosting, flow artifact, intensity and homogeneity, signal-to-noise ratio (SNR), susceptibility artifacts, and gray-white/cerebrospinal fluid contrast. For each subject, we used the MRI scan which was considered as the "best" quality scan by the ADNI investigators. In the description of the ADNI methods (http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml), the "best" quality image is the one which was used for the complete pre-processing

steps. We thus used the images which had been selected for the complete pre-processing pipeline. No other exclusion criteria based on image quality were applied. The identification numbers of the images used in this study are reported in Tables S2 to S9.

Participants

The criteria used for the inclusion of participants were those defined in the ADNI protocol (described in details at http://www.adni-info.org/Scientists/AboutADNI.aspx#). Enrolled subjects were between 55 and 90 (inclusive) years of age, had a study partner able to provide an independent evaluation of functioning, and spoke either English or Spanish. All subjects were willing and able to undergo all test procedures including neuroimaging and agreed to longitudinal follow up. Specific psychoactive medications were excluded. General inclusion/exclusion criteria were as follows: control subjects (CN) had MMSE scores between 24 and 30 (inclusive), a CDR (Clinical Dementia Rating) (Morris, 1993) of zero. They were non-depressed, non MCI, and non-demented. MCI subjects had MMSE scores between 24 and 30 (inclusive), a memory complaint, had objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II (Wechsler, 1987), a 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. AD patients had MMSE scores between 20 and 26 (inclusive), CDR of 0.5 or 1.0, and met NINCDS/ADRDA criteria for probable AD (McKhann et al., 1984).

We selected all the subjects for whom preprocessed images were available. The identification numbers of the subjects used in this study are reported in Tables S2 to S9. As a result, 509 subjects were selected: 162 cognitively normal elderly controls (CN) (76 males, 86 females, age ± SD = 76.3 ± 5.4 years, range = 60–90 years, and mini-mental score (MMS) = 29.2 ± 1.0, range = 25–30), 137 patients with AD (67 males, 70 females, age ± SD = 76.0 ± 7.3 years, range = 55–91 years, and MMS = 23.2 ± 2.0, range = 18–27), 76 patients with MCI who had converted to AD within 18 months (MCIc) (43 males, 33 females, age ± SD = 74.8 ± 7.4 years, range = 55–88 years, and MMS = 26.5 ± 1.9, range = 23–30) and 134 patients with MCI who had not converted to AD within 18 months (MCInc) (84 males, 50 females, age ± SD = 74.5 ± 7.2 years, range = 58–88 years, and MMS = 27.2 ± 1.7, range = 24–30). We did not consider MCI patients who had been followed less than 18 months and had not converted within this time frame. The 509 images came from 41 different centers.

To assess differences in demographic and clinical characteristics between groups, we used Student's t-test for age and MMS and Pearson's chi-square test for gender. Significance level was set at 0.05. No significant differences were found except for the MMS between controls and patients (AD or MCIc, $p<0.0001$).

In order to obtain unbiased estimates of the performances, the set of participants was then randomly split up into two groups of the same size: a training set and a testing set. The algorithms were trained on a training set and the measures of the diagnostic sensitivity and specificity were carried out with an independent test set. The division process preserved the age and sex distribution.

Demographic characteristics of the studied population selected from the ADNI database are presented in Table 1.

Methods

Classification experiments

Three classification experiments were performed to compare the different approaches. The first one is the classification between CN subjects and patients with probable AD and is referred to as "CN *vs* AD" in the following. The second one is the classification between CN subjects and MCI converters and is referred to as "CN *vs* MCIc". It corresponds to the detection of patients with prodromal AD as defined by Dubois and Albert (2004). Indeed, MCI patients who will convert to AD are, at baseline, patients with incipient AD but non-demented, i.e. patients with prodromal AD. The third one is the classification MCInc versus MCIc and is referred to as "MCInc *vs* MCIc". It corresponds to the prediction of conversion in MCI patients.

Classification methods

The different approaches we compared can be grouped into three categories with respect to the features used for the classification. In the first category, the features are defined at the level of the MRI voxel. Specifically, the features are the probability of the different tissue classes (GM, WM and CSF) in a given voxel. In the second category, the features are defined at the vertex-level on the cortical surface. Specifically, the features are the cortical thickness at each vertex of the cortex. The methods of the third category include only the hippocampus.

These methods are summarized in Table 2 and briefly presented in the following paragraphs.

First category: voxel-based segmented tissue probability maps

The features of the methods of the first category were computed as follows. All T1-weighted MR images were spatially normalized and segmented into GM, WM and CSF using the SPM5 (Statistical Parametric Mapping, London, UK) unified segmentation routine (Ashburner and Friston, 2005) with the default parameters. These maps constitute a first set of tissue probability maps and will be referred to respectively as SPM5_GM, SPM5_WM and SPM5_CSF.

To evaluate the impact of the registration step on the classification accuracy, the GM and WM probability maps in native space segmented by the SPM5 unified segmentation routine were also normalized to the population template generated from all the images, using the DARTEL diffeomorphic registration algorithm (Ashburner, 2007) with the default parameters. The obtained transformations were applied to the GM, WM and CSF tissue maps. These maps compose a second set of tissue probability maps and will be referred to respectively as

Table 1
Demographic characteristics of the studied population (from the ADNI database). Values are indicated as mean ± standard-deviation [range].

Group	Diagnostic	Number	Age	Gender	MMS	# Centers
Whole set	CN	162	76.3 ± 5.4 [60–90]	76 M/86 F	29.2 ± 1.0 [25–30]	40
	AD	137	76.0 ± 7.3 [55–91]	67 M/70 F	23.2 ± 2.0 [18–27]	39
	MCIc	76	74.8 ± 7.4 [55–88]	43 M/33 F	26.5 ± 1.9 [23–30]	30
	MCInc	134	74.5 ± 7.2 [58–88]	84 M/50 F	27.2 ± 1.7 [24–30]	36
Training set	CN	81	76.1 ± 5.6 [60–89]	38 M/43 F	29.2 ± 1.0 [25–30]	35
	AD	69	75.8 ± 7.5 [55–89]	34 M/35 F	23.3 ± 1.9 [18–26]	32
	MCIc	39	74.7 ± 7.8 [55–88]	22 M/17 F	26.0 ± 1.8 [23–30]	21
	MCInc	67	74.3 ± 7.3 [58–87]	42 M/25 F	27.1 ± 1.8 [24–30]	30
Testing set	CN	81	76.5 ± 5.2 [63–90]	38 M/43 F	29.2 ± 0.9 [26–30]	35
	AD	68	76.2 ± 7.2 [57–91]	33 M/35 F	23.2 ± 2.1 [20–27]	33
	MCIc	37	74.9 ± 7.0 [57–87]	21 M/16 F	26.9 ± 1.8 [24–30]	24
	MCInc	67	74.7 ± 7.3 [58–88]	42 M/25 F	27.3 ± 1.7 [24–30]	31

Table 2
Summary of the approaches tested in this study.

Features		Segmentation registration	Tissues probability maps	Classifier	Method #	Method's name
Voxel–segmented tissue probability maps	Direct	DARTEL	GM	Linear SVM	1.1.1 a	Voxel-Direct-D-gm
			GM + WM + CSF	Linear SVM	1.1.1 b	Voxel-Direct-D-all
		SPM5	GM	Linear SVM	1.1.2 a	Voxel-Direct-S-gm
			GM + WM + CSF	Linear SVM	1.1.2 b	Voxel-Direct-S-all
	Direct VOI	DARTEL	GM	Linear SVM	1.2.1 a	Voxel-Direct_VOI-D-gm
			GM + WM + CSF		1.2.1 b	Voxel-Direct_VOI-D-all
		SPM5	GM	Linear SVM	1.2.2 a	Voxel-Direct_VOI-S-gm
			GM + WM + CSF	Linear SVM	1.2.2 b	Voxel-Direct_VOI-S-all
	STAND-score	DARTEL	GM	Linear SVM	1.3.1 a	Voxel-STAND-D-gm
			GM + WM + CSF		1.3.1 b	Voxel-STAND-D-all
		SPM5	GM + WM + CSF	Linear SVM	1.3.2 a	Voxel-STAND-S-gm
				Linear SVM	1.3.2 b	Voxel-STAND-S-all
		SPM5 custom template	GM	Linear SVM	1.3.3 a	Voxel-STAND-Sc-gm
			GM + WM + CSF	Linear SVM	1.3.3 b	Voxel-STAND-Sc-all
	Atlas	DARTEL	GM	Linear SVM	1.4.1 a	Voxel-Atlas-D-gm
			GM + WM + CSF	Linear SVM	1.4.1 b	Voxel-Atlas-D-all
		SPM5	GM	Linear SVM	1.4.2 a	Voxel-Atlas-S-gm
			GM + WM + CSF	Linear SVM	1.4.2 b	Voxel-Atlas-S-all
	COMPARE	DARTEL	GM	Linear SVM	1.5.1 a	Voxel-COMPARE-D-gm
			GM + WM + CSF	Linear SVM	1.5.1 b	Voxel-COMPARE-D-all
		SPM5	GM	Gaussian SVM	1.5.2 a	Voxel-COMPARE-S-gm
			GM + WM + CSF	Gaussian SVM	1.5.2 b	Voxel-COMPARE-S-all
Cortical thickness	Direct	Freesurfer	–	Linear SVM	2.1	Thickness-Direct
	Atlas	Freesurfer	–	Linear SVM	2.2	Thickness-Atlas
	ROI	Freesurfer	–	Logistic Reg.	2.3	Thickness-ROI
Hippocampus	Volume	Freesurfer	–	Parzen	3.1.1	Hippo-Volume-F
	Volume	SACHA	–	Parzen	3.1.2	Hippo-Volume-S
	Shape	SACHA	–	Linear SVM	3.2	Hippo-Shape

DARTEL_GM, DARTEL_WM and DARTEL_CSF. Some papers used only GM maps while others included all three tissues. In our experiments, we systematically evaluated the added value of WM and CSF maps by comparing the classification obtained with only GM to that obtained with all three classes. All maps were then modulated to ensure that the overall tissue amount remains constant. No spatial smoothing was performed, unless when otherwise specified.

The different methods of this category differ by the way the features are extracted and/or selected from the voxel probability maps. This is detailed in the following paragraphs.

Direct

The simplest approach consists in considering the voxels of the tissue probability maps directly as features in the classification. This type of approach is referred to as "*Voxel-Direct*" in the following. Such an approach was proposed by Klöppel et al. (2008) with two different versions: one is based on whole brain datasets and the other includes only data from a volume of interest (VOI) located in the anterior medial temporal lobe, including part of the hippocampus. This volume of interest was defined as two rectangular cuboids centered on $x = -17$, $y = -8$, $z = -18$ and $x = 16$, $y = -9$, $z = -18$ in the MNI space. Their dimensions were 12 mm, 16 mm and 12 mm in the x, y and z directions respectively. The latter method will be referred to as "*Voxel-Direct_VOI*". In their paper, they used only DARTEL_GM maps. Here, we will test all approaches with the following sets of probability maps: SPM5_GM only, SPM5_GM and SPM5_WM and SPM5_CSF, DARTEL_GM only, DARTEL_GM, and DARTEL_WM and DARTEL_CSF.

STAND-score

Vemuri et al. (2008) proposed an approach called the STAND score, in which the dimensionality is reduced by a sequence of feature aggregation and selection steps. First, the tissue probability maps were smoothed and down-sampled by averaging. Then, voxels that contained less than 10% tissue density values and CSF in half or more of the images were not considered for further analysis. A feature selection step was then carried out. First, a linear SVM was applied for each tissue class, which attributes a weight to each feature. Only features of which weights are consistent with increased neurodegeneration in the pathological group were kept. Then a second feature selection step was performed on the remaining features. To ensure spatial consistency, neighboring voxels of the voxels selected so far were also selected. The features from the different tissue classes were concatenated and then used in the classification. This approach is referred to as "*Voxel-STAND*" in the following. In their paper, the features used for this approach were the GM, WM and CSF tissue probability maps segmented and registered with the SPM5 unified segmentation routine using a customized tissue probability maps. Thus we also tested the classification with customized tissue probability maps.

Atlas based

Another approach consists in grouping the voxels into anatomical regions using a labeled atlas. This type of approach is used in Lao et al. (2004); Magnin et al. (2009). Each tissue probability map in the stereotaxic space was parceled into 116 regions of interest (ROI) using the AAL (Automatic Anatomical Labeling) atlas (Tzourio-Mazoyer et al., 2002). In each ROI, we computed the mean tissue probability and used these values as features in the classification. Such an approach will be referred to as "*Voxel-Atlas*". Note that the AAL is a predefined anatomical atlas, which has not been specifically designed for studying patients with AD; its areas thus do not necessarily represent pathologically homogeneous regions.

COMPARE

Instead of using a predefined atlas, Fan et al. (2007, 2008a,b) proposed a parcellation that is adapted to the pathology. The thorough explanation of the method is in Fan et al. (2007). Very briefly, the concept of COMPARE is to create homogeneously discriminative regions. In these regions, the voxel values are aggregated to form the features of the classification. Feature selection steps are then performed to identify the most discriminative regions. In the following, we refer to this approach as "*Voxel-COMPARE*". We used the COMPARE software freely available on request for download online (https://www.rad.upenn.edu/sbia/software/index.html).

Second category: cortical thickness

In this second category, the features are the cortical thickness values at each vertex of the cortical surface. Cortical thickness represents a direct index of atrophy and thus is a potentially powerful candidate to assist in the diagnosis of AD (Thompson et al., 2001, 2003, 2004; Lerch et al., 2005, 2008; Bakkour et al., 2009; Dickerson et al., 2009; Hua et al., 2009; McDonald et al., 2009). Cortical thickness measures were performed with the FreeSurfer image analysis suite (Massachusetts General Hospital, Boston, MA), which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). The technical details of this procedure are described in Sled et al. (1998), Dale et al.(1999), Fischl et al. (1999a,b) and Fischl and Dale (2000). All the cortical thickness maps were registered onto the default FreeSurfer common template. Four subjects were not successfully processed by the FreeSurfer pipeline. Those subjects are marked by an asterisk in Tables S2 to S9. They could thus not be classified with the SVM and were excluded from the training set. For the testing set, the subjects were considered as 50% misclassified.

Direct

As in Klöppel et al. (2008) for voxel-based maps, the simplest way consists in considering cortical thickness values at every vertex directly as features in the classification with no other preprocessing step. This approach is referred to as "*Thickness-Direct*" in the following.

Atlas based

As in the voxel-based case, we also tested an approach where vertices are grouped into anatomical regions using an atlas. Such approach is used in (Querbes et al., 2009; Desikan et al., 2009). The cortical parcellation was carried out with the cortical atlas of (Desikan et al., 2006). The atlas is composed of 68 gyral based ROIs. In each ROI, we computed the mean cortical thickness and used these values as features in the classification. This approach is referred to as "*Thickness-Atlas*" in the following.

ROI

Desikan et al. (2009) parcellated the brain into neocortical and non-neocortical ROIs by wrapping an anatomical atlas (Desikan et al., 2006). They studied the discriminative power for CN *vs* MCIc of the mean thickness (neocortical regions) and the volume (both neocortical and non-neocortical regions). For their analysis, the mean thickness and the volumes of the right and the left hemispheres, for each ROI, were added together. The volumes were corrected using estimate of the total intracranial volume.

Their study was carried out on a cohort of 97 participants selected from the Open Access Series of Imaging Studies (OASIS) database (Marcus et al., 2007). They found out that, with a logistic regression analysis, the best set of discriminator was: the entorhinal cortex thickness, the supramarginal gyrus thickness and the hippocampal volume. They used these features with a logistic regression to classify CN and MCIc and to classify CN and AD. Therefore, in this approach, we used only these three features for the classification. This approach is referred to as "*Thickness-ROI*" in the following.

Third category: hippocampus

Finally, we tested the discriminative power of methods which consider only the hippocampus and not the whole brain or the whole cortex as in the two first categories. The hippocampus is affected at the earliest stages of the disease and has thus been used as a marker of early AD in a vast number of studies.

Here, the segmentation of the hippocampus was performed using SACHA, a fully automatic method we previously developed (Chupin et al., 2007, 2009a). This approach has been shown to be competitive with manual tracing for the discrimination of patients with AD and MCI (Colliot et al., 2008; Chupin et al., 2009b). This approach segments both the hippocampus and the amygdala simultaneously based on competitive region-growing between these two structures. It includes prior knowledge on the location of the hippocampus and the amygdala derived from a probabilistic atlas and on the relative positions of these structures with respect to anatomical landmarks which are automatically identified.

We also evaluated the hippocampal volume obtained with the FreeSurfer image analysis suite.

Volume

We first tested the classification accuracy obtained when the only feature is the hippocampal volume. For each subject, we computed the volume of the hippocampi. Volumes were normalized by the total intracranial volume (TIV) computed by summing SPM5 segmentation maps of grey matter, white matter, and cerebrospinal fluid (CSF), inside a bounding box defined in standard space to obtain a systematic inferior limit. For more robustness with respect to segmentation errors, left and right volumes were averaged. The thorough explanation of the method is in (Chupin et al., 2007, 2009a,b). This approach is referred to as "*Hippo-Volume-S*" in the following.

We also evaluated this approach with the hippocampal volume obtained with the FreeSurfer image analysis suite and corrected with the total intracranial volume also obtained with obtained with FreeSurfer. This approach will be referred to as "*Hippo-Volume-F*".

Shape

We then tested an approach in which the features describe the hippocampal shape (Gerardin et al., 2009). Each hippocampus was described by a series of spherical harmonics (SPHARM) to model the shape of the segmented hippocampi. The classification features were based on the SPHARM coefficients. Specifically, each subject was represented by two sets (one for each hippocampus) of three-dimensional SPHARM coefficients. The SPHARM coefficients were computed using the SPHARM-PDM (Spherical Harmonics-Point Distribution Model) software developed by the University of North Carolina and the National Alliance for Medical Imaging Computing (http://www.namic.org/Wiki/index.php/Algorithm:UNC:Shape_Analysis). In the original paper by our group describing this method (Gerardin et al., 2009), we used a feature selection step because the subjects groups were much smaller (less than 30 subjects in each group). When the number of subjects is small, the classifier can be more sensitive to uninformative features. In the present study, the number of subjects was larger and thus a feature selection step is less necessary and increases the risk of overfitting. We thus chose to avoid this selection step. We also tested the procedure with the selection step but it did not lead to further improvement in this study. Moreover, the degree of the SPHARM decomposition was set at four. Four subjects were not successfully processed by the SPHARM pipeline. Those subjects are marked by a dagger in Tables S2 to S9. They could thus not be classified with the SVM and were excluded from the training set. For the testing set, those subjects were considered as 50% misclassified. This approach is referred to as "*Hippo-Shape*" in the following.

Classification using SVM

Classifiers

A support vector machine is a supervised learning method. In brief: given a training set of size K: $(x_k, y_k)_{k=1\ldots K}$, where x_k in $\boldsymbol{R}^d$ are observations, and y_k in {-1,1} are corresponding labels, SVMs search for the optimal margin hyperplane (OMH) separating groups, i.e. the hyperplane for which the margin between groups is maximal. More details on SVM can be found in (Vapnik, 1995; Shawe-Taylor and Cristianini, 2000, 2004; Schölkopf and Smola, 2001). We used a linear C-SVM for all the approaches except COMPARE (Fan et al., 2007) for

which a non-linear C-SVM with a Gaussian kernel was used. The SVM implementation relied on the LIBSVM Library (Chang and Lin, 2001).

The dimension of the features of the approach *Hippo-Volume* is only one. Therefore a much simpler classifier can be used with no hyperparameter: each participant is assigned to the closest group. Specifically, if S_1 and S_2 are two groups of participants with respective centers of mass defined as m_1 and m_2, a new individual with hippocampus volume x is assigned to the closest group according to its Euclidean distance to the center of mass. This is a Parzen window classifier with the linear kernel and assuming a prevalence of 50% (Shawe-Taylor and Cristianini, 2004).

As in (Desikan et al., 2009) a logistic regression is used instead of a SVM, the classification step of *Thickness-ROI* was also based on a logistic regression.

Evaluation

In order to obtain unbiased estimates of the performances, the set of participants was randomly split into two groups of the same size: a training set and a testing set. The division process preserves the age and sex distribution. The training set was used to determine the optimal values of the hyperparameters of each method and to train the classifier. The testing set was then only used to evaluate the classification performances. The training and testing sets were identical for all methods, except for those four cases for which the cortical thickness pipeline failed and those other four for which the SPHARM pipeline failed. For the SPHARM and the cortical thickness methods, the subjects for whom the corresponding pipeline failed could not be classified with the SVM and were therefore excluded from the training set. As for the testing set, since those subjects were neither misclassified nor correctly classified, they were considered as 50% misclassified. This approach was chosen because a failure of the pipeline is a weakness of the methods.

On the training set, cross-validation (CV) was used to estimate the optimal values of hyperparameters. In general, there is only one hyperparameter which is the cost parameter C of the linear C-SVM. In *Voxel-STAND*, there is a second parameter which is the threshold t of feature selection. In *Voxel-COMPARE*, a second parameter is the size σ of the Gaussian kernel and the third parameter is the number n of selected features. In *Hippo-Volume*, there is no hyperparameter. The optimal parameter values were determined using a grid-search and leave-one-out cross validation (LOOCV) on the training set. The grid search was performed over the ranges $C = 10^{-5}$, $10^{-4.5}$, ..., $10^{2.5}$, 10^{3}, $t = 0.06, 0.08, ..., 0.98$, $\sigma = 100, 200, ..., 1000$ and $n = 1, 2, ..., 150$ (except for *Voxel-COMPARE* were $C = 10^{0}$, 10^{1},$10^{1.5}$, 10^{2}, $10^{2.5}$).

For each approach, the optimized set of hyperparameters was then used to train the classifier using the training group; the performance of the resulting classifier was then evaluated on the testing set. In this way, we achieved unbiased estimates of the performances of each method.

For each method, we computed the number of true positives TP (i.e. the number of diseased individuals which were correctly identified by the classifier), the number of true negatives TN (i.e. the number of healthy individuals which were correctly identified by the classifier), the number of false positives FP (i.e. the number of healthy individuals which were not correctly identified by the classifier), the number of false negatives FN (i.e. the number of diseased individuals which were not correctly identified by the classifier). We then computed the sensitivity defined as $TP/(TP+FN)$, the specificity defined as $TN/(TN+FP)$, the positive predictive value defined as $PPV = TP/(TP+FP)$, the negative predictive value defined as $NPV = TN/(TN+FN)$. Finally it should be noted that the number of subjects in each group is not the same. The classification accuracy does not enable to compare the performances between the different classification experiments. Thus we considered both the specificity and the sensitivity instead.

To assess whether each method performs significantly better than a random classifier, we used McNemar's chi square tests. Significance level was set at 0.05. We also used McNemar's chi square tests to assess differences between DARTEL and SPM5 registrations and between classification results obtained using only GM and using all three maps. The McNemar test investigates the difference between proportions in paired observations. We used it to assess the difference between proportions of correctly classified subjects, i.e. accuracy. The corresponding contingency table is presented in Table 3.

Results

Classification results

The results of the classification experiments are summarized in Tables 4, 5 and 6 respectively for CN *vs* AD, CN *vs* MCIc and CN *vs* MCInc. The classification results of CN *vs* AD and CN *vs* MCIc are also represented in Fig. 1. In each table, the different methods are referred to either by their abbreviation or by their number defined in Table 2.

CN vs AD

The classification results for CN *vs* AD are summarized in Table 4 and in Fig. 1. All methods performed significantly better than chance ($p<0.05$). The four *Voxel* methods (*Voxel-Direct*, *Voxel-STAND*, *Voxel-Atlas*, *Voxel-COMPARE*) classified AD from CN with very high specificity (over 89%) and high sensitivity: 75% for *Voxel*-STAND and over 81% for the other three methods. Methods based on the cortical thickness led to similar results with at least 90% specificity and 69%, 74% and 79% respectively for *Thickness-ROI*, *Thickness-Direct* and *Thickness-Atlas*. The hippocampus-based strategies were as sensitive but less specific: between 63% for *Hippo-Volume* and 84% for *Hippo-Shape*.

CN vs MCIc

Classification results for CN *vs* MCIc are summarized in Table 5 and in Fig. 1. Most methods were substantially less sensitive than for AD *vs* CN classification. All methods except *Voxel-COMPARE* and the *Hippo* methods obtained significantly better results than a random classifier ($p<0.05$). There was no substantial difference between the results obtained with *Voxel-Direct*, *Voxel-Atlas* and *Voxel-STAND*. All those methods reached a high specificity (over 85%) but a sensitivity ranging between 51% (*Voxel-COMPARE*) and 73% (*Voxel-STAND*). The methods based on cortical thickness behave as well as the previous ones. *Hippo-Volume* was slightly less specific but as sensitive as for the AD *vs* CN classification.

MCInc vs MCIc

The classification results for MCInc *vs* MCIc are summarized in Table 5. Only four methods managed to predict conversion slightly more accurately than a random classifier but none of them got significantly better results ($p>0.05$). *Thickness-Direct* reached 32% sensitivity and 91% specificity. *Voxel-STAND* reached 57% sensitivity and 78% specificity, *Voxel-COMPARE* reached 62% sensitivity and 67% specificity. *Hippo-Volume* distinguished MCIc from MCInc with 62% sensitivity and 69% specificity.

Table 3
Contingency table for the McNemar test. a: number of subjects correctly classified by both classifiers; b: number of subjects correctly classified by classifier 1 but misclassified by classifier 2; c: number of subjects misclassified by classifier 1 but correctly classified by classifier 2; and d: number of subjects misclassified by both classifiers.

	Classifier 2: correctly classified	Classifier 2: misclassified
Classifier 1: correctly classified	a	b
Classifier 1: misclassified	c	d

Table 4
Classification results CN *vs* AD.

		CN *vs* AD				
Method #	Method's name	SEN	SPE	PPV	NPV	McNemar test
1.1.1 a	Voxel-Direct-D-gm	81%	95%	93%	86%	$p<0.0001$
1.1.1 b	Voxel-Direct-D-all	68%	98%	96%	78%	$p<0.0001$
1.1.2 a	Voxel-Direct-S-gm	72%	89%	84%	79%	$p<0.0001$
1.1.2 b	Voxel-Direct-S-all	65%	88%	81%	75%	$p<0.0001$
1.2.1 a	Voxel-Direct_VOI-D-gm	71%	95%	92%	79%	$p<0.0001$
1.2.1 b	Voxel-Direct_VOI-D-all	65%	95%	92%	76%	$p<0.0001$
1.2.2 a	Voxel-Direct_VOI-S-gm	65%	91%	86%	76%	$p<0.0001$
1.2.2 b	Voxel-Direct_VOI-S-all	59%	81%	73%	70%	$p=0.0012$
1.3.1 a	Voxel-STAND-D-gm	69%	90%	85%	78%	$p<0.0001$
1.3.1 b	Voxel-STAND-D-all	71%	91%	87%	79%	$p<0.0001$
1.3.2 a	Voxel-STAND-S-gm	75%	91%	88%	81%	$p<0.0001$
1.3.2 b	Voxel-STAND-S-all	75%	86%	82%	80%	$p<0.0001$
1.3.3 a	Voxel-STAND-Sc-gm	72%	91%	88%	80%	$p<0.0001$
1.3.3 b	Voxel-STAND-Sc-all	71%	91%	87%	79%	$p<0.0001$
1.4.1 a	Voxel-Atlas-D-gm	78%	93%	90%	83%	$p<0.0001$
1.4.1 b	Voxel-Atlas-D-all	81%	90%	87%	85%	$p<0.0001$
1.4.2 a	Voxel-Atlas-S-gm	75%	93%	89%	82%	$p<0.0001$
1.4.2 b	Voxel-Atlas-S-all	74%	93%	89%	81%	$p<0.0001$
1.5.1 a	Voxel-COMPARE-D-gm	82%	89%	86%	86%	$p<0.0001$
1.5.1 b	Voxel-COMPARE-D-all	69%	81%	76%	76%	$p<0.0001$
1.5.2 a	Voxel-COMPARE-S-gm	66%	86%	80%	75%	$p<0.0001$
1.5.2 b	Voxel-COMPARE-S-all	72%	91%	88%	80%	$p<0.0001$
2.1	Thickness-Direct	74%	90%	86%	80%	$p<0.0001$
2.2	Thickness-Atlas	79%	90%	87%	84%	$p<0.0001$
2.3	Thickness-ROI	69%	94%	90%	78%	$p<0.0001$
3.1.1	Hippo-Volume-F	63%	80%	73%	72%	$p=0.0007$
3.1.2	Hippo-Volume-S	71%	77%	72%	76%	$p=0.0006$
3.2	Hippo-Shape	69%	84%	78%	76%	$p<0.0001$

SEN: sensitivity; SPE: specificity; PPV: positive predictive value; and NPV: negative predictive value.

Table 6
Classification results MCInc *vs* MCIc.

		MCInc *vs* MCIc				
Method #	Method's name	SEN	SPE	PPV	NPV	McNemar test
1.1.1 a	Voxel-Direct-D-gm	0%	100%	–	64%	$p=1.0$
1.1.1 b	Voxel-Direct-D-all	0%	100%	–	64%	$p=1.0$
1.1.2 a	Voxel-Direct-S-gm	0%	100%	–	64%	$p=1.0$
1.1.2 b	Voxel-Direct-S-all	0%	100%	–	64%	$p=1.0$
1.2.1 a	Voxel-Direct_VOI-D-gm	43%	70%	44%	69%	$p=0.62$
1.2.1 b	Voxel-Direct_VOI-D-all	0%	100%	–	64%	$p=1.0$
1.2.2 a	Voxel-Direct_VOI-S-gm	0%	100%	–	64%	$p=1.0$
1.2.2 b	Voxel-Direct_VOI-S-all	0%	100%	–	64%	$p=1.0$
1.3.1 a	Voxel-STAND-D-gm	57%	78%	58%	76%	$p=0.40$
1.3.1 b	Voxel-STAND-D-all	0%	100%	–	64%	$p=1.0$
1.3.2 a	Voxel-STAND-S-gm	22%	91%	57%	68%	$p=0.79$
1.3.2 b	Voxel-STAND-S-all	51%	79%	58%	75%	$p=0.49$
1.3.3 a	Voxel-STAND-Sc-gm	35%	70%	39%	66%	$p=0.30$
1.3.3 b	Voxel-STAND-Sc-all	41%	72%	44%	69%	$p=0.61$
1.4.1 a	Voxel-Atlas-D-gm	0%	100%		64%	$p=1.0$
1.4.1 b	Voxel-Atlas-D-all	0%	100%	–	64%	$p=1.0$
1.4.2 a	Voxel-Atlas-S-gm	0%	100%		64%	$p=1.0$
1.4.2 b	Voxel-Atlas-S-all	0%	100%	–	64%	$p=1.0$
1.5.1 a	Voxel-COMPARE-D-gm	62%	67%	51%	76%	$p=1.0$
1.5.1 b	Voxel-COMPARE-D-all	54%	78%	57%	75%	$p=0.50$
1.5.2 a	Voxel-COMPARE-S-gm	32%	82%	50%	69%	$p=0.84$
1.5.2 b	Voxel-COMPARE-S-all	51%	72%	50%	73%	$p=0.87$
2.1	Thickness-Direct	32%	91%	67%	71%	$p=0.24$
2.2	Thickness-Atlas	27%	85%	50%	68%	$p=0.82$
2.3	Thickness-ROI	24%	82%	43%	66%	$p=0.66$
3.1.1	Hippo-Volume-F	70%	61%	50%	79%	$p=0.89$
3.1.2	Hippo-Volume-S	62%	69%	52%	77%	$p=0.88$
3.2	Hippo-Shape	0%	100%	-	64%	$p=1.0$

SEN: sensitivity; SPE: specificity; PPV: positive predictive value; and NPV: negative predictive value.

Influence of the preprocessing

To evaluate the impact of the registration step, we tested both the registration using SPM5 unified segmentation and the registration DARTEL as described in the previous section. The influence of the registration step on the classification results is illustrated on Figs. 2 and 3. The performances obtained for the MCInc *vs* MCIc experiment were too low to be used to evaluate the impact of the registration step. Therefore we did not take them into account for this comparison. The use of the diffeomorphic registration algorithm DARTEL significantly improved the results of six out of 20 classification experiments ($p<0.05$). On the other hand, it led to significantly worse results in two cases. According to the results in Tables 4, 5, and 6, the use of customized tissue probability maps for the registration with SPM5 unified segmentation did not improve the results of *Voxel-STAND*.

We also compared the classification obtained with only the GM maps to those with GM, WM and CSF maps. Results are presented on Figs. 2 and 3. The use of all three maps led to significantly worse results ($p<0.05$) for two out of 20 classification experiments (*Voxel-Direct_VOI-S* and *Voxel-COMPARE-D*). It never led to significantly better results.

Table 5
Classification results CN *vs* MCIc.

		CN vs MCIc				
Method #	Method's name	SEN	SPE	PPV	NPV	McNemar test
1.1.1 a	Voxel-Direct-D-gm	57%	96%	88%	83%	$p=0.00052$
1.1.1 b	Voxel-Direct-D-all	49%	91%	72%	80%	$p=0.046$
1.1.2 a	Voxel-Direct-S-gm	32%	96%	80%	76%	$p=0.039$
1.1.2 b	Voxel-Direct-S-all	41%	94%	75%	78%	$p=0.044$
1.2.1 a	Voxel-Direct_VOI-D-gm	54%	95%	83%	82%	$p=0.0022$
1.2.1 b	Voxel-Direct_VOI-D-all	41%	96%	83%	78%	$p=0.0095$
1.2.2 a	Voxel-Direct_VOI-S-gm	32%	88%	55%	74%	$p=0.83$
1.2.2 b	Voxel-Direct_VOI-S-all	22%	99%	89%	73%	$p=0.046$
1.3.1 a	Voxel-STAND-D-gm	73%	85%	69%	87%	$p=0.025$
1.3.1 b	Voxel-STAND-D-all	65%	93%	80%	85%	$p=0.0019$
1.3.2 a	Voxel-STAND-S-gm	59%	86%	67%	82%	$p=0.082$
1.3.2 b	Voxel-STAND-S-all	49%	93%	75%	80%	$p=0.025$
1.3.3 a	Voxel-STAND-Sc-gm	62%	85%	66%	83%	$p=0.091$
1.3.3 b	Voxel-STAND-Sc-all	57%	90%	72%	82%	$p=0.026$
1.4.1 a	Voxel-Atlas-D-gm	65%	80%	60%	83%	$p=0.27$
1.4.1 b	Voxel-Atlas-D-all	54%	91%	74%	81%	$p=0.021$
1.4.2 a	Voxel-Atlas-S-gm	68%	95%	86%	87%	$p=0.00020$
1.4.2 b	Voxel-Atlas-S-all	59%	94%	81%	84%	$p=0.0021$
1.5.1 a	Voxel-COMPARE-D-gm	49%	81%	55%	78%	$p=0.73$
1.5.1 b	Voxel-COMPARE-D-all	51%	85%	61%	79%	$p=0.28$
1.5.2 a	Voxel-COMPARE-S-gm	49%	78%	50%	77%	$p=0.87$
1.5.2 b	Voxel-COMPARE-S-all	59%	78%	55%	81%	$p=0.64$
2.1	Thickness-Direct	54%	96%	87%	82%	$p=0.00084$
2.2	Thickness-Atlas	57%	93%	78%	82%	$p=0.0071$
2.3	Thickness-ROI	65%	94%	83%	85%	$p=0.00083$
3.1.1	Hippo-Volume-F	73%	74%	56%	86%	$p=0.47$
3.1.2	Hippo-Volume-S	70%	73%	54%	84%	$p=0.67$
3.2	Hippo-Shape	57%	88%	68%	82%	$p=0.072$

SEN: sensitivity; SPE: specificity; PPV: positive predictive value; and NPV: negative predictive value.

Complementariness of the methods

The different approaches tested tackle the classification problem with various angles and could thus be complementary. In order to quantify their similarity, we used the Jaccard similarity coefficient (Jaccard, 1901; Shattuck et al., 2001). In this case, the Jaccard index of two methods is the number of subjects correctly classified by both methods divided by the number of subjects correctly classified by at least one of the two methods. Results are presented on Figs. S1 and S2. All methods are in at least substantial agreement (Jaccard over 0.6) and most of them are in strong agreement. The most different results were obtained with the methods relying on the hippocampus.

We tested the combination of three approaches, one of each strategy: *Voxel-Direct-D-gm*, *Thickness-Atlas* and *Hippo-Volume-S*. A convenient approach to combine different SVM-based methods is to consider that the resulting classifier is a SVM which kernel is a linear

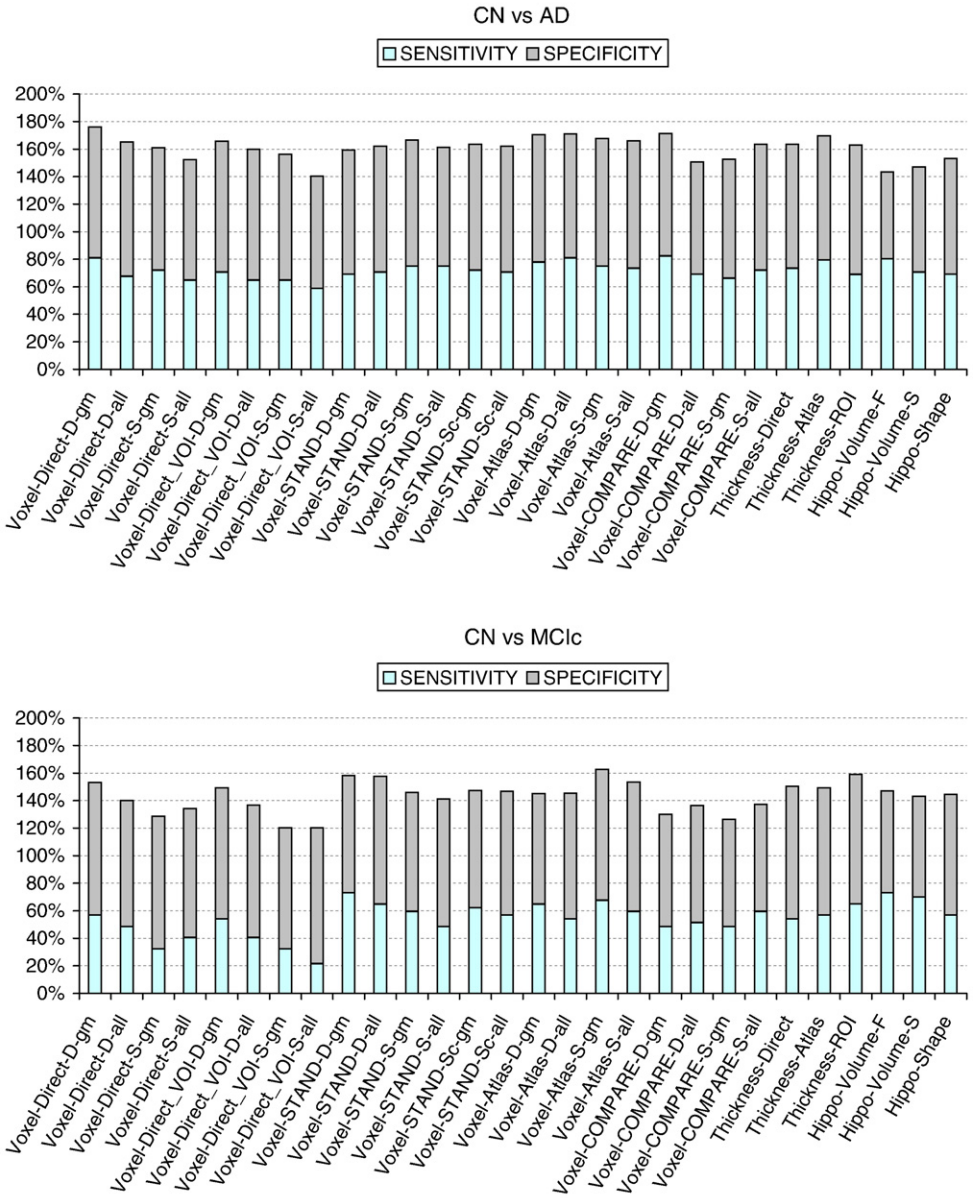

Fig. 1. Classification results for the different methods.

convex combination of the kernels of each method. The problem of learning both the coefficients of the best convex linear combination of kernels and the optimal margin hyperplane (OMH) is known as the multiple kernel learning (MKL) problem (Lanckriet et al., 2004; Bach et al., 2004; Sonnenburg et al., 2006). We used the SimpleMKL toolbox (Rakotomamonjy et al., 2008). All four possible combinations have been tested. The kernels are normalized with the trace of the Gram matrix of the training set. Note that for *Hippo-Volume-S*, the Parzen window classifier is replaced by a linear SVM.

None of these four combinations improved the accuracy in the CN *vs* AD experiment. Only the combination of *Hippo-Volume-S* and *Thickness-Atlas* improved only slightly the accuracy for the CN *vs* MCIc and the MCInc *vs* MCIc experiments. It distinguished MCIc from CN with 76% sensitivity and 85% specificity. The optimal coefficients of the linear combination were 0.057 and 0.943 for the kernels of *Hippo-Volume-S* and *Thickness-Atlas,* respectively. This combination classified MCIc and MCInc with 43% sensitivity and 83% specificity. The optimal coefficients of the linear combination were 0.030 and 0.970 respectively.

Influence of age and gender on classification results

We investigated whether the age of the subjects influences the classification results. We thus computed the average age of true positives, false positives, true negatives and false negatives. Overall, we found that the false positives were often older than the true negatives, meaning that the oldest controls were more often misclassified. Specifically, this was the case for 25 methods over 28 for CN *vs* AD and 24/28 for CN *vs* MCIc. Conversely, false negatives were often younger than the true positives, meaning that the youngest patients were more often misclassified. Specifically, this was the case for 26 methods over 28 for CN *vs* AD and 28/28 for CN *vs* MCIc. The number of misclassified subjects was too small to test for statistical significance of these differences. However, the fact that this difference was present for the vast majority of method suggests that it

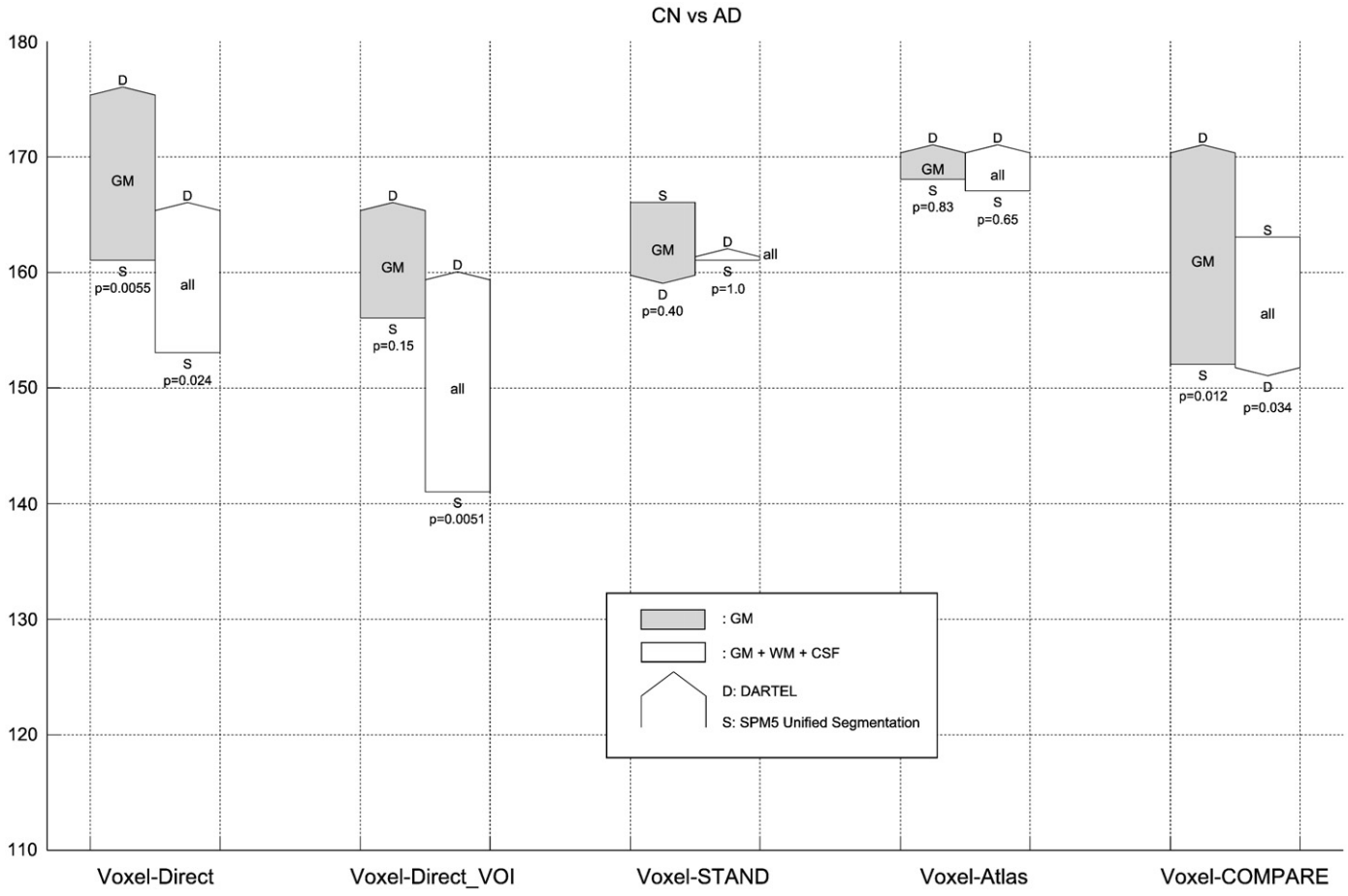

Fig. 2. Impact of the preprocessing on the accuracy for CN *vs* AD. The sum of the sensitivity and specificity is considered. The front tip of an arrow indicates the results obtained with DARTEL whereas the back tip indicates the results obtained with SPM5 unified segmentation. The color of the arrow indicates the features used. Grey arrows correspond to the use of GM probability maps only whereas white arrows correspond to the use of GM, WM and CSF probability maps. The p-values obtained with the McNemar's chi square test assessed the difference between the results obtained with DARTEL and SPM5.

may not be due to chance. We also investigated the influence of gender but did not find any difference.

Computation time

The computations were carried out with a processor running at 3.6 GHz with 2 GB of RAM. Table 7 presents, for each method, the order of magnitude of the computation time (i.e. minutes, hours, days, and weeks). For each method, we report the computation time of its three main phases: the feature computation step (segmentation and registration), the building of the classifier (including the grid search for the optimization of the hyperparameters and the learning of the classifier), and the classification of a new subject.

The order of magnitude of the computation time for the tissue segmentation and the registration step per subject is respectively about ten minutes and an hour with SPM5 and DARTEL. The cortical thickness computation and the registration of a single subject with FreeSurfer take roughly a day. The segmentation of the hippocampi of a subject lasts a few minutes and the shape analysis process with the SPHARM decomposition about one hour. The tuning of parameters and learning phase took from a few minutes to several weeks for the *Voxel-STAND* and *Voxel-COMPARE* methods. Once the hyperparameters are set and the learning is done, it takes at most minutes to classify a new subject.

Optimal margin hyperplanes

The classification function obtained with a linear SVM is the sign of the inner product of the features with w, a vector orthogonal to the optimal margin hyperplane (OMH) (Vapnik, 1995; Shawe-Taylor and Cristianini, 2000, 2004; Schölkopf and Smola, 2001). Therefore if the ith component w_i of the vector w is small, the ith feature will have a small influence on the classification. Conversely, if w_i is large, the ith feature will play an important role in the classifier. When the input features are the voxels of the image, each component of w also corresponds to a voxel. One can thus represent the values of w in the image space. Similarly, for the *Thickness* methods, the values of w can be represented on the cortical surface. The values of the optimal margin hyperplanes for the different methods are presented on Figs. from 4 to 7. This allows a qualitative comparison of the features used in the classifier. Our aim was not to perform a statistical analysis of differences between groups — for example using permutation tests on the coefficients (Mourao-Miranda et al., 2005).

Figs. 4 and 5 show the OMH for CN *vs* AD and CN *vs* MCIc respectively for the *Voxel* methods. Overall, the spatial patterns corresponding to CN *vs* AD and CN *vs* MCIc are similar. For *Voxel-Direct-D-gm*, the main regions were the medial temporal lobe (hippocampus, amygdala and the parahippocampal gyrus), the inferior and middle temporal gyri, the posterior cingulate gyrus and the posterior middle frontal gyrus. To a lesser extent, the OMH also included the inferior parietal lobule, the supramarginal gyrus, fusiform gyrus, the middle cingulate gyrus and in the thalamus. When all three tissue maps were used, the CSF maps mirrored the GM map (the enlargement of the ventricle mirroring GM reduction). This was also the case for part of the WM map, in particular in the hippocampal region. When using SPM5 unified segmentation instead of DARTEL, voxels were much more scattered and not grouped into anatomical regions except in the medial temporal lobe. For the AAL

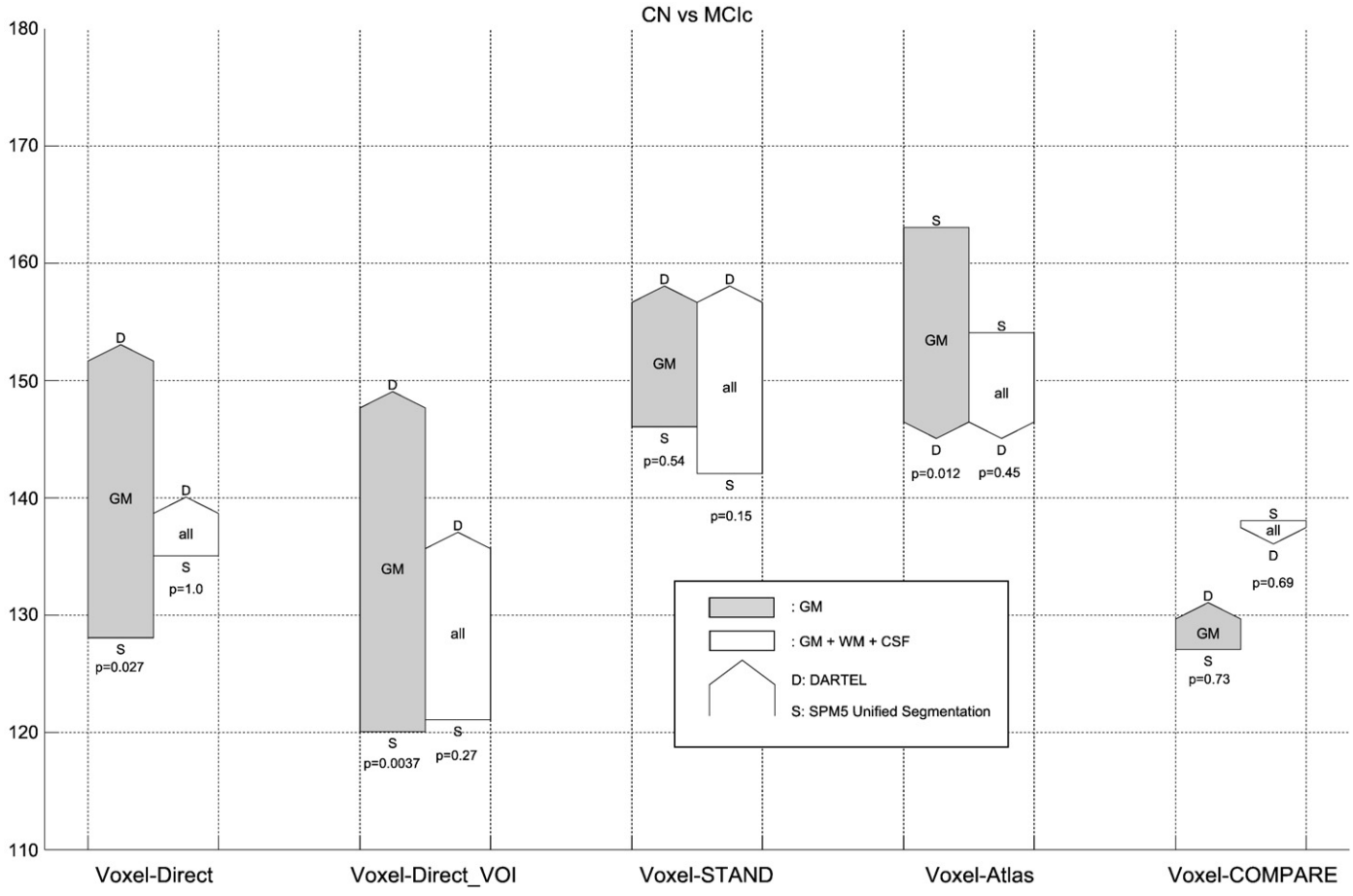

Fig. 3. Impact of the preprocessing on the accuracy for CN *vs* MCIc. The sum of the sensitivity and specificity is considered. The front tip of an arrow indicates the results obtained with DARTEL whereas the back tip indicates the results obtained with SPM5 unified segmentation. The color of the arrow indicates the features used. Grey arrows correspond to the used of GM probability maps only whereas white arrows correspond to the use of GM, WM and CSF probability maps. The p-values obtained with the McNemar's chi square test assessed the difference between the results obtained with DARTEL and SPM5.

atlas, regions included the hippocampus, the amygdala, the parahippocampal gyrus, the cingulum, the middle and inferior temporal gyri and the superior and inferior frontal gyri. The regions were very similar for the surface Atlas as shown on Fig. 6. Regions corresponding to *Thickness-Direct* (Fig. 7) were more restricted: the entorhinal cortex, the parahippocampal gyrus and to a lesser extent the lateral temporal lobe, the inferior parietal lobule and some prefrontal areas.

Optimal parameters of the classifiers

For each approach, the optimal values of the hyperparameters are summarized in Table S1. One should note that the *Hippo-Volume* method has no hyperparameter.

Discussion

In this paper, we compared different methods for the classification of patients with AD and MCI based on anatomical T1-weighted MRI. To evaluate and compare the performances of each method, three classification experiments were performed: CN *vs* AD, CN *vs* MCIc and CN *vs* MCInc. The set of participants was randomly split up into two groups of the same size: a training set and a testing set. For each approach, the optimal parameter values had been determined using a grid-search and LOOCV on the training set. Those values were then used to train the classifier using the training group; the performance of the resulting classifier was then evaluated on the testing set. In this way, we obtained unbiased estimates of the performances of each method.

Classification methods discriminate AD from normal aging

All the classification methods that we tested in this paper achieved accuracies significantly better than chance for the discrimination of patients with AD from normal aging. All methods except *Voxel-COMPARE* and *Hippo* methods performed significantly better than chance for the discrimination of patients with prodromal AD (MCIc) from normal aging. For AD *vs* CN, most methods achieved high sensitivity and specificity. However, at the prodromal stage, their sensitivity was substantially lower.

The classification results we obtained for AD *vs* CN with *Atlas* and *COMPARE* methods are lower than those reported in the respective papers: 94% accuracy for the *COMPARE* method in (Fan et al., 2008a) and 92% sensitivity and 97% specificity for the *Atlas* in Magnin et al. (2009). These differences can be explained by several factors. First, in the original papers, the hyperparameters were optimized on the testing set. This may lead to overfitting the testing set and thus to overestimate the sensitivity and specificity. On the contrary, in our evaluation, the learning step as well as the optimization of the hyperparameters had been carried out on a training set and the evaluation of the performance on a completely separated testing set. Thus our evaluation was unbiased. Another explanation may stem from differences between studied populations (sample size, stage of the disease). In particular, the ADNI population includes a large number of subjects with vascular lesions, which was not the case in Magnin et al. (2009). Finally, the image preprocessing step may also have an impact on the classification results. Davatzikos et al. (2008b) and Fan et al. (2008b) used the RAVENS maps (Goldszal et al., 1998),

Table 7
Order of magnitude of the computation time (i.e. minutes, hours, days, and weeks) for each method for its three main phases: feature computation step (segmentation and registration), building of the classifier (including the grid search for the optimization of the hyperparameters and the learning of the classifier), and classification of a new subject. The computations have been carried out with a processor running at 3.6 GHz with 2 GB of RAM.

Method #	Method's name	Segmentation registration	Grid search learning	Testing
1.1.1 a	Voxel-Direct-D-gm	Hour(s) per subject	Minute(s)	Minute(s)
1.1.1 b	Voxel-Direct-D-all	Hour(s) per subject	Minute(s)	Minute(s)
1.1.2 a	Voxel-Direct-S-gm	10 min per subject	Minute(s)	Minute(s)
1.1.2 b	Voxel-Direct-S-all	10 min per subject	Minute(s)	Minute(s)
1.2.1 a	Voxel-Direct_VOI-D-gm	Hour(s) per subject	Minute(s)	Minute(s)
1.2.1 b	Voxel-Direct_VOI-D-all	Hour(s) per subject	Minute(s)	Minute(s)
1.2.2 a	Voxel-Direct_VOI-S-gm	10 min per subject	Minute(s)	Minute(s)
1.2.2 b	Voxel-Direct_VOI-S-all	10 min per subject	Minute(s)	Minute(s)
1.3.1 a	Voxel-STAND-D-gm	Hour(s) per subject	Day(s)	Hour(s)
1.3.1 b	Voxel-STAND-D-all	Hour(s) per subject	Week(s)	Hour(s)
1.3.2 a	Voxel-STAND-S-gm	10 min per subject	Day(s)	Hour(s)
1.3.2 b	Voxel-STAND-S-all	10 min per subject	Week(s)	Hour(s)
1.3.3 a	Voxel-STAND-Sc-gm	20 min per subject	Day(s)	Hour(s)
1.3.3 b	Voxel-STAND-Sc-all	20 min per subject	Week(s)	Hour(s)
1.4.1 a	Voxel-Atlas-D-gm	Hour(s) per subject	Minute(s)	Minute(s)
1.4.1 b	Voxel-Atlas-D-all	Hour(s) per subject	Minute(s)	Minute(s)
1.4.2 a	Voxel-Atlas-S-gm	10 min per subject	Minute(s)	Minute(s)
1.4.2 b	Voxel-Atlas-S-all	10 min per subject	Minute(s)	Minute(s)
1.5.1 a	Voxel-COMPARE-D-gm	Hour(s) per subject	Week(s)	Hour(s)
1.5.1 b	Voxel-COMPARE-D-all	Hour(s) per subject	Week(s)	Hour(s)
1.5.2 a	Voxel-COMPARE-S-gm	10 min per subject	Week(s)	Hour(s)
1.5.2 b	Voxel-COMPARE-S-all	10 min per subject	Week(s)	Hour(s)
2.1	Thickness-Direct	Day(s) per subject	Minute(s)	Minute(s)
2.2	Thickness-Atlas	Day(s) per subject	Minute(s)	Minute(s)
2.3	Thickness-ROI	Day(s) per subject	Minute(s)	Seconds
3.1.1	Hippo-Volume-F	Day(s) per subject	Minute(s)	Seconds
3.1.2	Hippo-Volume-S	10 min per subject	Minute(s)	Seconds
3.2	Hippo-Shape	Hour(s) per subject	Minute(s)	Seconds

thus the registration and the segmentation step was different and might lead to different classification results. However, the aim of the present paper was to compare different classification strategies and it was thus necessary to use the same preprocessing for all methods. Since most of them relied on SPM, we chose to use this preprocessing for all methods. It is possible that using other registration approaches such as HAMMER would increase the classification performance but this is beyond the scope of this paper.

For the *Voxel-STAND* and *Voxel-Direct* methods, our results were similar to those reported in the original papers by Vemuri et al. (2008) and Klöppel et al. (2008). This can probably be explained by the fact that Vemuri et al.'s (2008) evaluation procedure is also based on independent testing group and that Klöppel et al. (2008) did not mention any optimization of the hyperparameters. As for the *Thickness-ROI*, the results (69% sensitivity and 94% specificity) were lower than those obtained by Desikan et al. (2009) (100% specificity and sensitivity). A possible explanation is that in their study the classifier was trained on a different population (patients with CDR = 0.5) selected from a different database (the OASIS database).

The results obtained with *Hippo-Volume* were similar to those that we previously reported for the ADNI database (Chupin et al., 2009b). The sensitivities and specificities were however lower than those found in our previous study on a different population (Colliot et al., 2008) (84% sensitivity and specificity for CN *vs* AD). This can be explained by several factors (Chupin et al., 2009b). First ADNI is a multi-site database whereas the data in the previous study came from a single scanner. Moreover the population included a large number of subjects with vascular lesions. The slight difference between the results obtained in Chupin et al. (2009b) and the present results mostly comes from the difference in the accuracy estimation: two separate groups instead of a LOOCV procedure. As for the *Hippo-Shape* method the results were substantially lower than our results reported in Gerardin et al. (2009) (86% for CN *vs* AD). This may result from the relatively small number of subjects used in our previous study. Besides, the estimation was carried out with a LOOCV. Moreover, this can also be due to that fact that all subjects were considered without

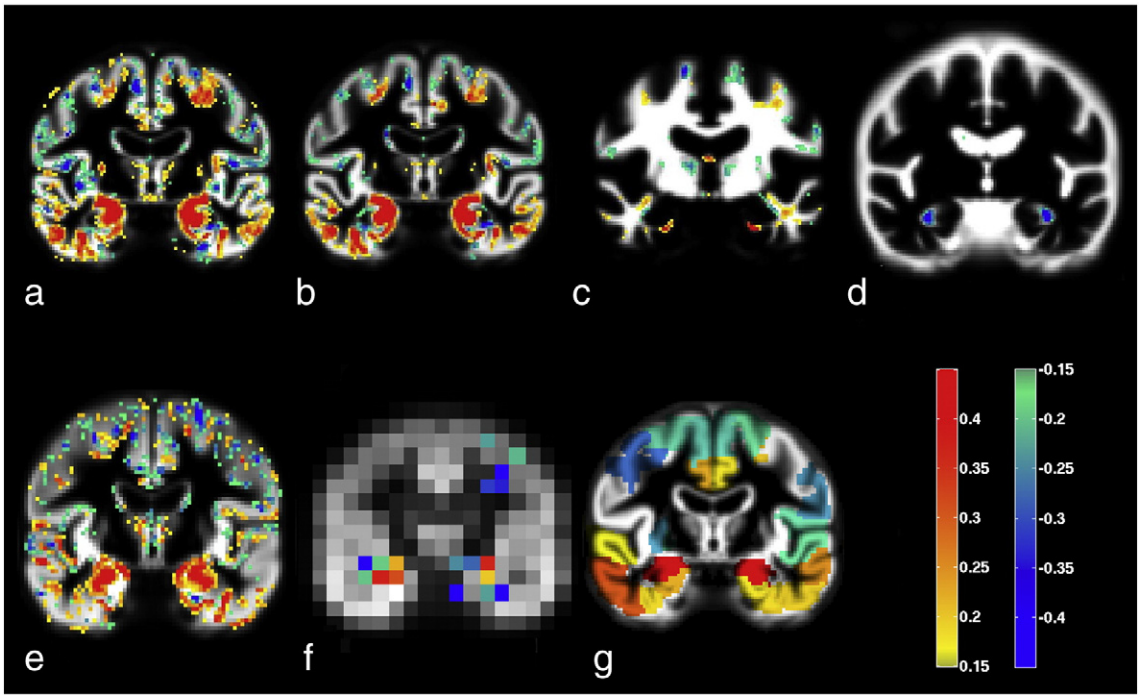

Fig. 4. Optimal margin hyperplane in the CN vs AD experiments for *Voxel-Direct-D-gm* (a), *Voxel-Direct-D-all* (b–d), *Voxel-Direct-S-gm* (e), *Voxel-STAND-D-gm* (f) and *Voxel-Atlas-D-gm* (g). The figure displays the normalized vector orthogonal to the hyperplane superimposed on the tissue average probability maps. The coronal slices are equivalent to $y = 9$ mm in the MNI-space. For visualization purposes, only coefficients w_i greater than 0.15 in absolute value are displayed. For regions in warm colors, tissue atrophy increases the likelihood of classification into AD or MCIc. For regions in cool colors, it is the opposite.

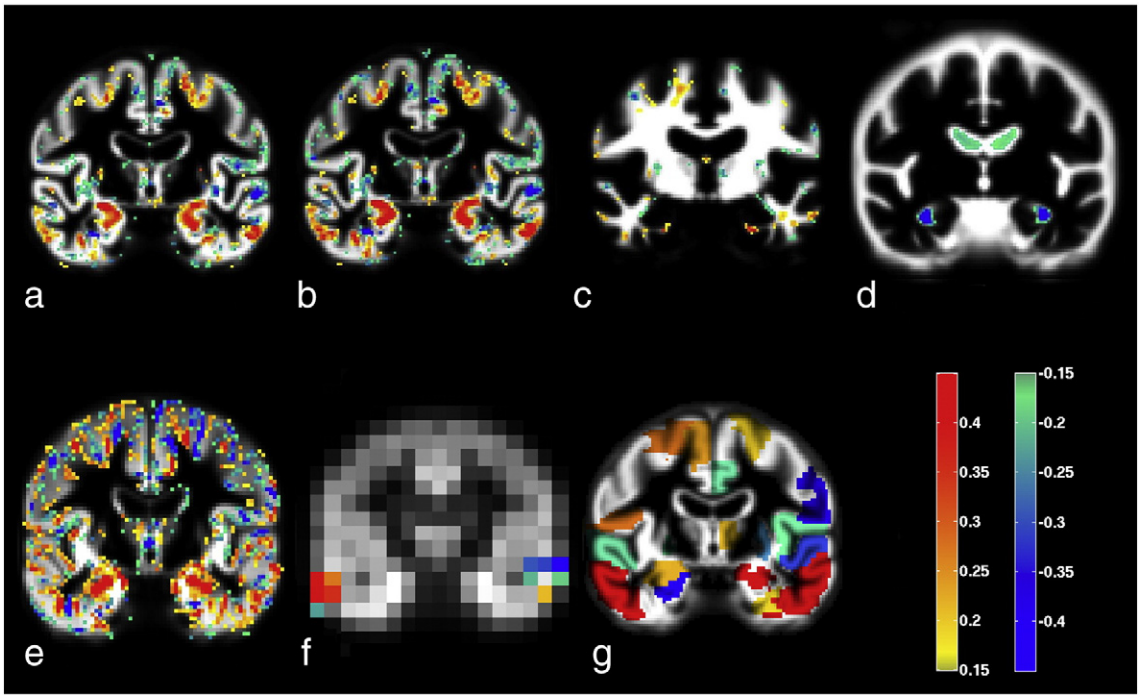

Fig. 5. Optimal margin hyperplane in the CN *vs* MCIc experiments for *Voxel-Direct-D-gm* (a), *Voxel-Direct-D-all* (b–d), *Voxel-Direct-S-gm* (e), *Voxel-STAND-D-gm* (f) and *Voxel-Atlas-D-gm* (g) (please refer to Fig. 4 for a complete description of the figure).

taking into consideration the quality control (Chupin et al., 2009b) of the hippocampus segmentation.

To our knowledge, the classification CN *vs* MCIc has only been addressed by Desikan et al. (2009). Davatzikos et al. (2008a) and Fan et al. (2008a,b) have performed the classification CN *vs* MCI with no distinction between converters and non-converters. The MCI group did not include only prodromal AD, hence the classification experiment cannot be compared to CN *vs* MCIc. Desikan et al. (2009) classified CN and MCI who converted within two years after baseline with 91% accuracy. This is substantially higher than the results obtained in our paper with the same method *Thickness-ROI* (65% sensitivity and 94% specificity).

Prediction of conversion in MCI patients

No method was able to predict conversion better than chance. The three most accurate methods were: *Voxel-STAND* (57% sensitivity and 78% specificity), *Voxel-COMPARE* (62% sensitivity and 67% specificity) and *Hippo-Volume* (62% sensitivity and 69% specificity). These three methods restricted their search to a portion of the brain. In *Voxel-STAND*

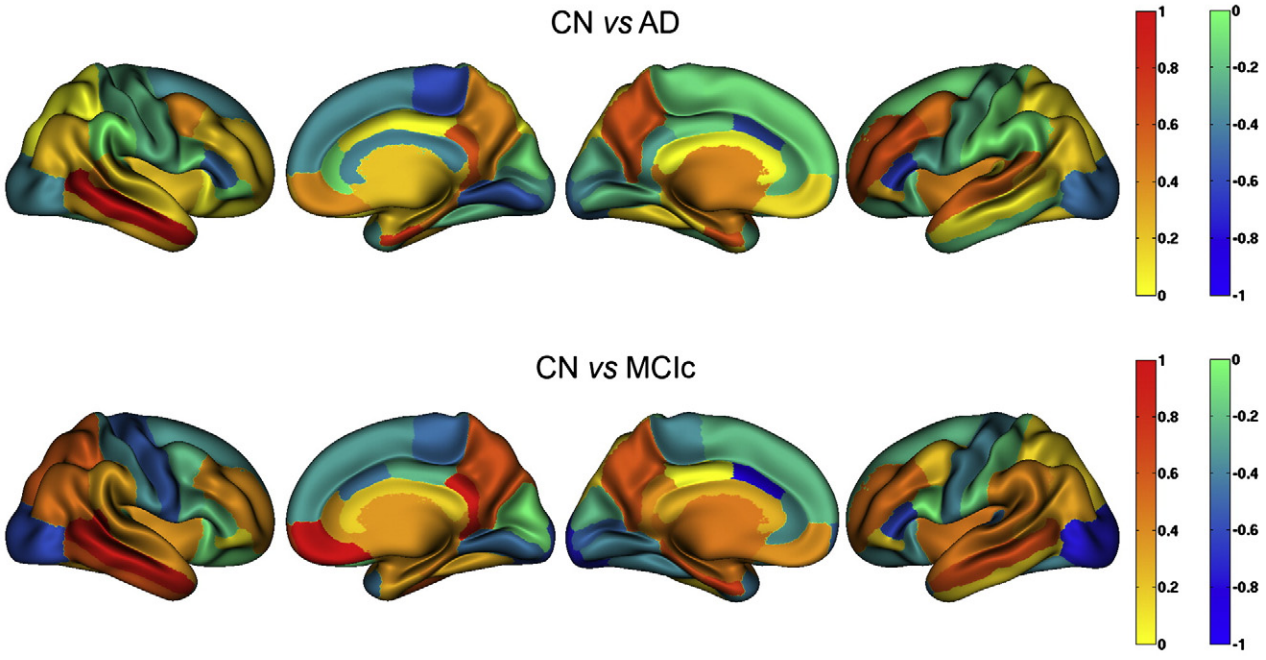

Fig. 6. Optimal margin hyperplane for *Thickness-Atlas*. Upper rows: CN *vs* AD experiment. Lower rows: CN *vs* MCIc experiment.

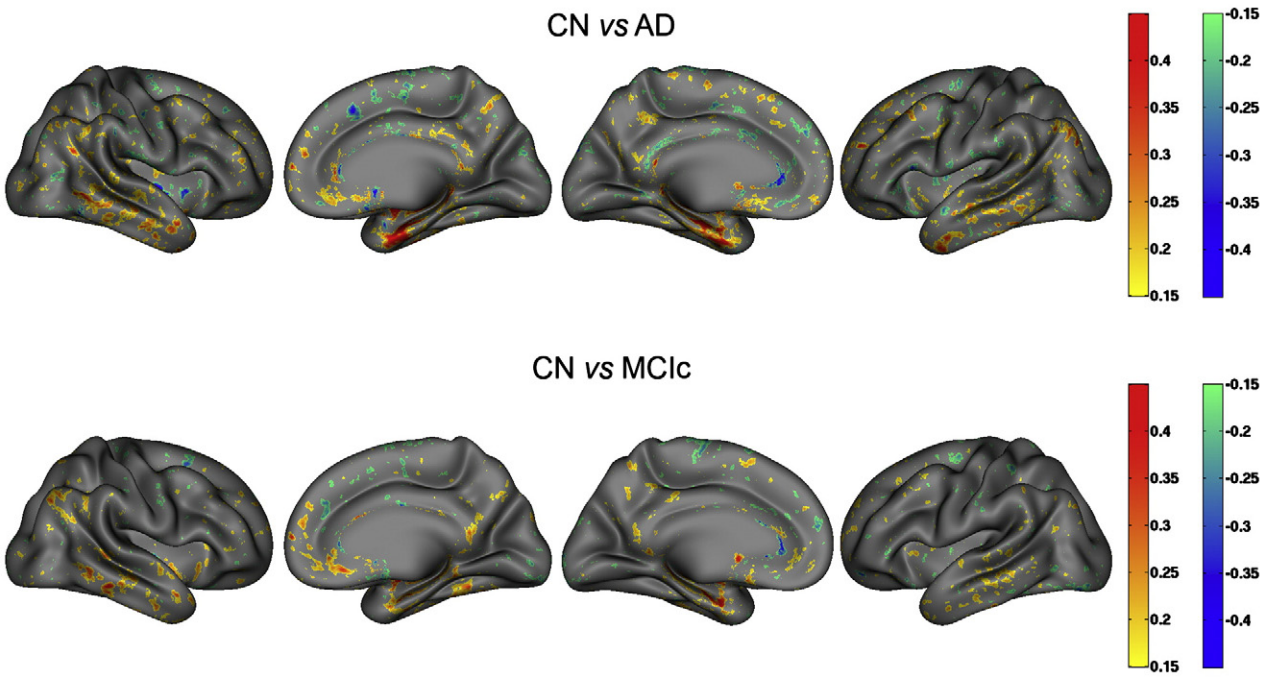

Fig. 7. Optimal margin hyperplane for *Thickness-Direct*. Upper rows: CN *vs* AD experiment. Lower rows: CN *vs* MCIc experiment.

and *Voxel-COMPARE*, this was done using feature selection: the selected regions are mainly in the medial temporal structures. In *Hippo-Volume*, this was done by considering only the hippocampus.

Even for these three methods, the performances remained particularly low. The main reason is certainly that MCI non converters are a very heterogeneous group: some patients would convert shortly after the end of the follow-up and are thus in fact prodromal AD patients while others would remain stable for a long period of time. We thus advocate that classification methods should be focused on the detection of prodromal AD (i.e. MCI converters) which is a much better defined entity.

To our knowledge, the classification MCInc *vs* MCIc has only been addressed by Misra et al. (2009) and Querbes et al. (2009). Misra et al. (2009) considered the conversion within 12 months and Querbes et al. (2009) within 24 months. They obtained substantially higher accuracy: respectively 81.5% and 76% accuracy. Misra et al. (2009) used the COMPARE (Fan et al., 2007) classification methods. The differences may result from the same reasons as explained in the previous paragraph: the use of separate training and testing sets and differing preprocessing steps. Querbes et al. (2009) used a feature selection step, which may explain the slightly higher accuracy.

Whole brain or hippocampus?

For CN *vs* AD, methods using the whole brain (or the whole cortex) reached substantially higher specificity (over 90%) than those based on the hippocampus (from 63% to 84%). For the detection of prodromal AD, hippocampal-based approaches remained competitive with whole-brain methods. It thus seems that considering the whole brain is advantageous mostly at the most advanced stages. Indeed, at these more advanced stages, the atrophy is much more widespread. Moreover, it should be noted that many subjects included in the ADNI have vascular lesions which may be, at least partially, captured by whole brain methods. For intermediate stages, an alternative would be to consider a set of selected regions instead of the whole brain or the hippocampus alone. For example, *Thickness-ROI* performs at least as well as whole brain approaches for the detection of prodromal AD. Even though they achieve lower accuracies, hippocampal-based methods may still be of interest to the clinician because they provide a direct and easily interpretable index to the clinician (the hippocampal volume) while the whole-brain approaches base their classification on a complex combination of different regions.

All methods presented substantial agreement (Jaccard index over 0.6). The most different results were obtained between hippocampal and whole brain methods. However, combining them through multiple kernel learning did not improve the classification results.

The registration step: is a fully deformable method advantageous?

The use of DARTEL significantly improved the classification results in six cases, while it led to lower results in only two cases. This is in line with other studies which reported that DARTEL led to higher overlap values (Klein et al., 2009; Yassa and Stark, 2009) and higher sensitivity for voxel-based morphometry (Bergouignan et al., 2009). In particular, the use of a fully deformable method was advantageous for the medial temporal lobe as shown in (Yassa and Stark, 2009; Bergouignan et al., 2009). Since the hippocampus is highly affected in AD, we expected that using a method which registers the hippocampus better, would result in higher classification accuracy.

Does adding WM and CSF maps increase the performance of the classifiers?

In their original description, some of the tested methods used the three tissue (GM, WM and CSF) maps (e.g. Vemuri et al., 2008, Fan et al., 2007, Magnin et al., 2009) while others used only the GM maps (e.g. Klöppel et al., 2008). In this paper, we systematically tested whether the compared methods performed better with the three maps or with only the GM maps. It should be noted that this does not aim at assessing the diagnostic value of WM or CSF in general but only to test if including all tissue maps is more effective for these particular classification approaches under study. On the whole, adding the WM and the CSF probability maps did not improve the classification performances. Adding WM and CSF maps increases the dimensionality of the feature space which can make the classifier unstable and lead to overfitting the data. This problem is well-known in machine learning as the curse of dimensionality. Besides, elder subjects are likely to have WM structural abnormalities caused by leucoaraiosis or other diseases. Therefore adding WM tissues may add noise in the features. Even if WM structural

abnormalities alter (Levy-Cooperman et al., 2008) the tissue segmentation step, GM probability maps are more robust features than WM tissues probability maps.

Adding the WM and the CSF in the features may improve the results in two instances. The first one is when the method encloses a feature selection step. Methods including feature selection steps are more able to keep only the added value and avoid considering the noise but, overall, the improvement is not substantial. Adding WM and CSF may also improve the results of methods grouping the voxels into ROIs via wrapping a labeled atlas. It may make up for the parcellation error due to the registration step but, again, the improvement is not substantial.

Is it worth performing feature selection?

The main objectives of the feature selection step are to keep only informative features and to reduce the dimensionality of the feature space. In our evaluation, two methods included a feature selection step: *Voxel-STAND* and *Voxel-COMPARE*. Overall, these methods did not perform substantially better than simpler ones. In particular, their results might be more sensitive to the training set. Indeed, feature selection can be regarded as a learning step. In such a case, the feature selection step increases the class of all possible classification functions, which could lead to overfitting the data. A more robust way to decrease the dimensionality of the features way would be to use more prior knowledge of the disease.

Besides features selection can be time consuming as it adds new hyperparameters and thus makes the grid search less tractable. Compared to *Voxel-Direct* and *Voxel-Atlas*, *Voxel-STAND* and *Voxel-COMPARE* are time consuming (up to weeks), mostly because of the number of hyperparameters to be tuned.

Nevertheless, feature selection proved useful in two specific cases. First, these methods proved less sensitive when increasing the dimensionality of the feature space by adding WM and CSF maps. They also tended to be more accurate for the MCIc *vs* MCInc experiment, where only a few brain regions are informative.

Does age influence the classification accuracy?

Overall, we found that the oldest controls and the youngest patients were more often misclassified. This may results from different causes. Normal aging is associated with atrophy of the grey and white matter and increase of the CSF (Good et al., 2001; Salat et al., 2004). Moreover, aging is also associated with alterations in tissue intensity and contrast, which can disrupt the segmentation step and thus artificially increase the measured atrophy (Salat et al., 2009). Besides, elderly subjects are more likely to have structural abnormalities of the white matter, which can also impede the tissue segmentation step (Levy-Cooperman et al., 2008) and increase the measured atrophy. In addition, elderly subjects have a propensity to suffer from mixed dementia (Zekry et al., 2002).

Optimal margin hyperplanes

In a linear SVM, the OMH can be easily represented. The OMH provides information about the regions of the brain which was used by the classifier. It should be noted that this only provides qualitative information on the hyperplanes, and that no statistical analysis of the OMH coefficients was performed.

With *Voxel-Direct-D*, *Voxel-Atlas* and *Thickness-Atlas*, the regions in which atrophy increased the likelihood of being classified as AD or MCIc were largely consistent with the pattern of atrophy demonstrated in previous morphometric studies. These regions included the medial temporal lobe, the inferior and middle temporal gyri (Chételat and Baron, 2003; Good et al., 2002; Busatto et al., 2003; Rusinek et al., 2004; Tapiola et al., 2008), the posterior cingulate gyrus (Karas et al., 2004; Chételat et al., 2005; Laakso et al., 1998) and the posterior middle frontal gyrus (Whitwell et al., 2007), the fusiform gyrus, the thalamus (Karas et al., 2003, 2004; Chételat et al., 2005). As for the cortical methods, the main regions in the medial temporal, middle and inferior lateral temporal, inferior parietal, and posterior cingulated cortices and with a lesser extent parietal, frontal, and lateral occipital cortices, which is consistent with the previous group studies based on cortical thickness (Thompson et al., 2004; Lerch et al., 2005, 2008; McDonald et al., 2009).

In conclusion, we compared different automatic classification methods to assist in the early diagnosis of Alzheimer's disease using the ADNI database. Most of them classify AD and CN with high accuracy. However, at the prodromal stage, their sensitivity was substantially lower. Combinations with other markers and/or more sophisticated prior knowledge seem necessary to be able to detect prodromal AD with high accuracy.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; Principal Investigator: Michael Weiner; NIH grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and through generous contributions from the following: Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for the Study of Aging, with participation from the U.S. Food and Drug Administration. Industry partnerships are coordinated through the Foundation for the National Institutes of Health. 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 of Neuro Imaging at the University of California, Los Angeles.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2010.06.013.

References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26 (3), 839–851.
Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.
Bach, F., Lanckriet, G., Jordan, M.I., 2004. Multiple kernel learning, conic duality, and the SMO algorithm. Proceedings of the 21st International Conference on Machine Learning, pp. 41–48.
Bakkour, A., Morris, J.C., Dickerson, B.C., 2009. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. Neurology 72 (12), 1048–1055.
Bergouignan, L., Chupin, M., Czechowska, Y., Kinkingnéhun, S., Lemogne, C., Le Bastard, G., Lepage, M., Garnero, L., Colliot, O., Fossati, P., 2009. Can voxel based morphometry, manual segmentation and automated segmentation equally detect hippocampal volume differences in acute depression? Neuroimage 45 (1), 29–37.
Blennow, K., de Leon, M.J., Zetterberg, H., 2006. Alzheimer's disease. Lancet 368 (9533), 387–403.
Busatto, G.F., Garrido, G.E., Almeida, O.P., Castro, C.C., Camargo, C.H., Cid, C.G., Buchpiguel, C.A., Furuie, S., Bottino, C.M., 2003. A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease. Neurobiol. Aging 24 (2), 221–231.
Chang, C.C., Lin, C.J., 2001. LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
Chételat, G., Baron, J.C., 2003. Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. Neuroimage 18 (2), 525–541.
Chételat, G., Landeau, B., Eustache, F., Mézenge, F., Viader, F., de la Sayette, V., Desgranges, B., Baron, J.C., 2005. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. Neuroimage 27 (4), 934–946.
Chupin, M., Mukuna-Bantumbakulu, A.R., Hasboun, D., Bardinet, E., Baillet, S., Kinkingnéhun, S., Lemieux, L., Dubois, B., Garnero, L., 2007. Automated segmentation of the hippocampus and the amygdala driven by competition and anatomical priors: Method and validation on healthy subjects and patients with Alzheimer's disease. Neuroimage 34, 996–1019.

Chupin, M., Hammers, A., Liu, R.S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J.S., Garnero, L., Lemieux, L., 2009a. Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. Neuroimage 46 (3), 749–761.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O., Alzheimer's Disease Neuroimaging Initiative, 2009b. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19 (6), 579–587.

Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., Dubois, B., Garnero, L., Eustache, F., Lehéricy, S., 2008. Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. Radiology 248 (1), 194–201.

Convit, A., De Leon, M.J., 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 (2), 131–138.

Convit, A., de Asis, J., de Leon, M.J., Tarshish, C.Y., De Santi, S., Rusinek, H., 2000. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiol. Aging 21 (1), 19–26.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9 (2), 179–194.

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 (4), 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 (4), 1220–1227.

Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., Fischl, B., Alzheimer's Disease Neuroimaging Initiative, 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 132 (8), 2048–2057.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980.

Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B. T., Morris, J.C., Fischl, B., Buckner, R.L., 2009. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cereb. Cortex 19 (3), 497–510.

Dubois, B., Albert, M.L., 2004. Amnestic MCI or prodromal Alzheimer's disease? Lancet Neurol. 3 (4), 246–248.

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 (8), 734–746.

Fan, Y., Shen, D., Davatzikos, C., 2005. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention 8 (Pt 1), pp. 1–8.

Fan, Y., Shen, D., Gur, R.C., Davatzikosa, C., 2007. COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Trans. Med. Imaging 26 (1), 93–105.

Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., Alzheimer's Disease Neuroimaging Initiative, 2008b. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39 (4), 1731–1743.

Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C., 2008a. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage 41 (2), 277–285.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis. II: Inflation, flattening, and a surface based coordinate system. Neuroimage 9 (2), 195–207.

Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8 (4), 272–284.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl Acad. Sci. USA 97, 11050–11055.

Fox, N.C., Schott, J.M., 2004. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet 363 (9406), 392–394.

Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.-S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., Francis, E., Colliot, O., 2009. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage 47 (4), 1476–1486.

Goldszal, A.F., Davatzikos, C., Pham, D.L., Yan, M.X., Bryan, R.N., Resnick, S.M., 1998. An image-processing system for qualitative and quantitative volumetric analysis of brain images. J. Comput. Assist. Tomogr. 22 (5), 827–837.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 14 (1), 21–36.

Good, C.D., Scahill, R.I., Fox, N.C., Ashburner, J., Friston, K.J., Chan, D., Crum, W.R., Rossor, M.N., Frackowiak, R.S., 2002. Automatic differentiation of anatomical patterns in the human brain: validation with studies of degenerative dementias. Neuroimage 17 (1), 29–46.

Hajnal, J.V., Hill, D.L.G., Hawkes, D.J., 2001. Medical Image Registration. CRC Press, New York.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., Alzheimer's Disease Neuroimaging Initiative, 2009. Spatially augmented LP boosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48 (1), 138–149.

Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Chou, Y.Y., Ho, A.J., Gutman, B., Toga, A.W., Jack Jr., C.R., Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging Initiative, 2009. Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. Neuroimage 48 (4), 668–681.

Jaccard P., 1901. Etude Comparative de la Distribution Florale dans une Portion des Alpes et du Jura, Bulletin de la Société vaudoise des Sciences Naturelles 37, 547–79

Jack Jr., C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E., 1997. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 49 (3), 786–794.

Jack Jr., C.R., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1998. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. Neurology 51 (4), 993–999.

Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, L., Ward, J., Dale, C., Felmlee, A.M., Gunter, J.P., Hill, J.L., Killiany, D.L., Schuff, R., Fox-Bosetti, N., Lin, S., Studholme, C., DeCarli, C., Krueger, C.S., Ward, G., Metzger, H.A., Scott, G.J., Mallozzi, K.T., Blezek, R., Levy, D., Debbins, J., Fleisher, J.P., Albert M., A.S., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W., 2008. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27 (4), 685–691.

Jagust, W., 2006. Positron emission tomography and magnetic resonance imaging in the diagnosis and prediction of dementia. Alzheimers Dement. 2, 36–42.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., Fischl, B., Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. Neuroimage 30, 436–443.

Juottonen, K., Laakso, M.P., 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 (1), 15–22.

Karas, G.B., Burton, E.J., Rombouts, S.A., van Schijndel, R.A., O'Brien, J.T., Scheltens, P., McKeith, I.G., Williams, D., Ballard, C., Barkhof, F., 2003. A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. Neuroimage 18 (4), 895–907.

Karas, G.B., Scheltens, P., Rombouts, S.A., Visser, P.J., van Schijndel, R.A., Fox, N.C., Barkhof, F., 2004. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. Neuroimage 23 (2), 708–716.

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46 (3), 786–802.

Klöppel, 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.J., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681–689.

Laakso, M.P., Soininen, H., Partanen, K., Lehtovirta, M., Hallikainen, M., Hänninen, T., Helkala, E.L., Vainio, P., Riekkinen Sr., P.J., 1998. MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects. Neurobiol. Aging 19 (1), 23–31.

Laakso, M.P., Frisoni, G.B., Könönen, M., Mikkonen, M., Beltramello, A., Geroldi, C., Bianchetti, A., Trabucchi, M., Soininen, H., Aronen, H.J., 2000. Hippocampus and entorhinal cortex in frontotemporal dementia and Alzheimer's disease: a morphometric MRI study. Biol. Psychiatry 47 (12), 1056–1063.

Lanckriet, G.R.G., De Bie, T., Cristianini, N., Jordan, M.I., Noble, W.S., 2004. A statistical framework for genomic data fusion. Bioinformatics 20 (16), 2626–2635.

Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21 (1), 46–57.

Lerch, J.P., Pruessner, J.C., Zijdenbos, A., Hampel, H., Teipel, S.J., Evans, A.C., 2005. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. Cereb. Cortex 15 (7), 995–1001.

Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A.C., 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. Neurobiol. Aging 29 (1), 23–30.

Levy-Cooperman, N., Ramirez, J., Lobaugh, N.J., Black, S.E., 2008. Misclassified tissue volumes in Alzheimer disease patients with white matter hyperintensities: importance of lesion segmentation procedures for volumetric analysis. Stroke 39 (4), 1134–1141.

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S., Benali, H., 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51 (2), 73–83.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. 19 (9), 1498–1507.

McDonald, C.R., McEvoy, L.K., Gharapetian, L., Fennema-Notestine, C., Hagler Jr., D.J., Holland, D., Koyama, A., Brewer, J.B., Dale, A.M., Alzheimer's Disease Neuroimaging Initiative, 2009. Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. Neurology 73 (6), 457–465.

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 (7), 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 (4), 1415–1422.
Morris, J.C., 1993. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43, 2412–2414.
Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. Neuroimage 28 (4), 980–995.
Narayana, P.A., Brey, W.W., Kulkarni, M.V., Sievenpiper, C.L., 1988. Compensation for surface coil sensitivity variation in magnetic resonance imaging. Magn. Reson. Imaging 6, 271–274.
Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303–308.
Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Démonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P., Alzheimer's Disease Neuroimaging Initiative, 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain 32 (8), 2036–2047.
Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y., 2008. SimpleMKL. J. Mach. Learn. Res. 9, 2491–2521.
Rusinek, H., Endo, Y., De Santi, S., Frid, D., Tsui, W.H., Segal, S., Convit, A., de Leon, M.J., 2004. Atrophy rate in medial temporal lobe during progression of Alzheimer disease. Neurology 63 (12), 2354–2359.
Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S., Busa, E., Morris, J.C., Dale, A.M., Fischl, B., 2004. Thinning of the cerebral cortex in aging. Cereb. Cortex 14 (7), 721–730.
Salat, D.H., Lee, S.Y., van der Kouwe, A.J., Greve, D.N., Fischl, B., Rosas, H.D., 2009. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. Neuroimage 48 (1), 21–28.
Schölkopf, B., Smola, A.J., 2001. Learning with Kernels. MIT Press.
Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. M. Resonance Image Tissue Classification Using a Partial Volume Model. Neuroimage 13 (5), 856–876.
Shawe-Taylor, J., Cristianini, N., 2000. Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.
Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
Sonnenburg, S., Rätsch, G., Schäfer, C., Schölkopf, B., 2006. Large scale multiple kernel learning. J. Mach. Learn. Res. Arch. 7, 1531–1565.
Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., Pihlajamäki, M., Laakso, M.P., Hallikainen, M., Hämäläinen, A., Vanhanen, M., Helkala, E.L., Vanninen, R., Nissinen, A., Rossi, R., Frisoni, G.B., Soininen, H., 2008. MRI of hippocampus and entorhinal cortex in mild cognitive impairment: a follow-up study. Neurobiol. Aging 29 (1), 31–38.
Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield, C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W., 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. Cereb. Cortex 11 (1), 1–16.
Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W., 2003. Dynamics of gray matter loss in Alzheimer's disease. J. Neurosci. 23 (3), 994–1005.
Thompson, P.M., Hayashi, K.M., Sowell, E.R., Gogtay, N., Giedd, J.N., Rapoport, J.L., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Doddrell, D.M., Wang, Y., van Erp, T. G., Cannon, T.D., Toga, A.W., 2004. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. Neuroimage 23 (Suppl 1), S2–S18.
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.
Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. New York Inc, Springer-Verlag.
Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B. F., Petersen, R.C., Jack Jr., C.R., 2008. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. Neuroimage 39 (3), 1186–1197.
Wechsler, D., 1987. Manual for the Wechsler Memory Scale-Revised. The Psychological Corporation, San Antonio.
Whitwell, J.L., Przybelski, S.A., Weigand, S.D., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr., C.R., 2007. Brain 130 (7), 1777–1786.
Whitwell, J.L., Shiung, M.M., Przybelski, S.A., Weigand, S.D., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack Jr., C.R., 2008. MRI patterns of atrophy associated with progression to AD in amnestic mild cognitive impairment. Neurology 70 (7), 512–520.
Xu, Y., Jack Jr., C.R., O'Brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., Petersen, R.C., 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. Neurology 54 (9), 1760–1767.
Yassa, M.A., Stark, C.E., 2009. A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. Neuroimage 44 (2), 319–327.
Ye, J., Chen, K., Wu, T., Li, J., Zhao, Z., Patel, R., Bae, M., Janardan, R., Liu, H., Alexander, G., Reiman, E., 2008. Heterogeneous data fusion for Alzheimer's disease study. Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '08, pp. 1025–1033.
Zekry, D., Hauw, J.J., Gold, G., 2002. Mixed dementia: epidemiology, diagnosis, and treatment. J. Am. Geriatr. Soc. 8, 1431–1438.