Alzheimer's & Dementia 10 (2014) 456–467

Automatic temporal lobe atrophy assessment in prodromal AD: Data from the DESCRIPA study

Andrea Chincarini[a,*], Paolo Bosco[a,b], Gianluca Gemme[a], Mario Esposito[a,b], Luca Rei[a,b], Sandro Squarcia[a,b], Roberto Bellotti[c,d], Lennart Minthon[e], Giovanni Frisoni[f], Philip Scheltens[i], Lutz Frölich[g], Hilkka Soininen[h], Pieter-Jelle Visser[i,j], Flavio Nobili[k], for the Alzheimer's Disease Neuroimaging Initiative

[a]*Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Genova, Italy*
[b]*Dipartimento di Fisica, Università degli Studi di Genova, Genova, Italy*
[c]*Dipartimento Interateneo di Fisica "M. Merlin" and TIRES, Università degli Studi di Bari, Bari, Italy*
[d]*Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Bari, Italy*
[e]*Clinical Memory Research Unit, Department of Clinical Sciences, Malmö Lund University, Malmo, Sweden*
[f]*IRCCS Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy*
[g]*Department of Geriatric Psychiatry, Zentralinstitut für Seelische Gesundheit, University of Heidelberg, Mannheim, Germany*
[h]*Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland*
[i]*Alzheimer Centre, Department of Neurology, VU University Medical Centre, Amsterdam, The Netherlands*
[j]*Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands*
[k]*Neurofisiologia Clinica, Dipartimento di Neuroscienze, Oftalmologia e Genetica, Azienda Ospedale-Università S. Martino, Genova, Italy*

Abstract

Background: In the framework of the clinical validation of research tools, this investigation presents a validation study of an automatic medial temporal lobe atrophy measure that is applied to a naturalistic population sampled from memory clinic patients across Europe.

Methods: The procedure was developed on 1.5-T magnetic resonance images from the Alzheimer's Disease Neuroimaging Initiative database, and it was validated on an independent data set coming from the DESCRIPA study. All images underwent an automatic processing procedure to assess tissue atrophy that was targeted at the hippocampal region. For each subject, the procedure returns a classification index. Once provided with the clinical assessment at baseline and follow-up, subjects were grouped into cohorts to assess classification performance. Each cohort was divided into converters (*co*) and nonconverters (*nc*) depending on the clinical outcome at follow-up visit.

Results: We found the area under the receiver operating characteristic curve (AUC) was 0.81 for all *co* versus *nc* subjects, and AUC was 0.90 for subjective memory complaint (SMC_{nc}) versus all *co* subjects. Furthermore, when training on mild cognitive impairment (MCI-*nc*/MCI-*co*), the classification performance generally exceeds that found when training on controls versus Alzheimer's disease (CTRL/AD).

Conclusions: Automatic magnetic resonance imaging analysis may assist clinical classification of subjects in a memory clinic setting even when images are not specifically acquired for automatic analysis.

Keywords: MRI; Image analysis; Memory clinics; Naturalistic population; Alzheimer's disease; Medial temporal lobe; Hippocampus

1. Introduction

Between 2007 and 2011, new research and operational criteria for Alzheimer's disease (AD) were published by an expert dementia panel [1,2] and the U.S. National Institute of

*Corresponding author. Tel.: +39-010-353-6496; Fax: +39-010-313358.
E-mail address: andrea.chincarini@ge.infn.it

1552-5260/$ - see front matter © 2014 The Alzheimer's Association. All rights reserved.
http://dx.doi.org/10.1016/j.jalz.2013.05.1774

Table 1
Demographics and clinical findings for the training data set from ADNI

Cohort	Sample size	Age (years)	M/F	MMSE
Training set A				
CTRL	189	76.6 (5.1)	95/94	29.1 (0.9)
AD	144	75.5 (7.5)	78/66	22.3 (3.3)
Training set B				
MCI–*nc*	166	75.7 (7.3)	106/60	27.2 (2.4)
MCI–*co*	136	75.1 (7.1)	80/56	25.2 (2.7)

Abbreviations: CTRL, controls; AD, Alzheimer's disease; MCI–*nc*, MCI nonconverters; MCI–*co*, MCI converters; M, male; F, female; MMSE, Mini-Mental State Examination score.

Aging (NIA)-Alzheimer Association (AA) [3–5]. To be transferred to clinical practice, the criteria need to be validated in large patient groups, although scientific evidence is rather convincing and promising. Tools showing amyloidosis and neurodegeneration (i.e., the biomarkers) must be applied in naturalistic series of subjects presenting with cognitive complaints but without dementia. In this framework, they should be able to identify patients in the predementia stage of AD from all of the other patients who may manifest mild cognitive impairment (MCI) due to other neurological or systemic diseases or to drug abuse.

A typical paradigm involves the computation of biomarkers from several data sources (such as structural and functional imaging, proteomics, genetic profile), sometimes combining more than one to boost the classification power. The main goal is to find a tool (or a combination of tools) that is useful for diagnostic purposes by means of quick, low-cost, and widely available procedures. Magnetic resonance imaging (MRI) is probably the most available tool in this framework.

By disclosing atrophy as the last step of neurodegeneration, MRI is, in principle, less sensitive than tools showing signs of neurodegeneration at the synaptic level before massive cell death occurs, such as [^{18}F]-fluorodeoxyglucose positron emission tomography (FDG-PET) and phospho-tau protein in cerebrospinal fluid (CSF) [6]. A notable counterexample is provided by Bateman and colleagues [7], who suggest that atrophy may be an even earlier indicator of neurodegeneration than hypometabolism in familial AD. Although a discussion on this study applicability to sporadic AD is outside of the scope of this work, it is worth noting that neurodegeneration markers were taken on different brain areas (i.e., hippocampal volume on MRI and glucose metabolism in the precuneus on FDG-PET), which may well exhibit a nontrivial time mismatch in showing a measurable neurodegeneration signature. On regional FDG-PET studies on the hippocampus, the reader is referred to Mosconi and colleagues [8] and Clerici and colleagues [9]. With that said improved acquisition sequences, higher magnetic fields, and quicker and more reliable analysis tools make structural MRI still competitive at an earlier stage.

Automatic analysis tools often rely on a homogeneous database for their training and for their validation. Database homogeneity typically involves subject inclusion criteria, image acquisition protocols, and neuropsychological tests. In addition, automatic analysis tools are usually developed and trained on "research-grade" images (i.e., images for which the characteristics are uniform across the study and artifact presence is limited or filtered out at the source).

In the last years, there has been a strong development of automatic tools to detect atrophy; for instance, see Klöppel et al. [10], Klauschen et al. [11], Calvini et al. [12], Heckemann et al. [13], Aksu et al. [14], Matsuda et al. [15], Cui et al. [16], Plant et al. [17], and the review paper of Cuingnet et al. [18]. Literature works on analysis tools developed and tested on the Alzheimer's Disease Neuroimaging Initiative (ADNI; http://www.adni.loni.ucla.edu) data set are ample, but there are, to our knowledge, very few works on the validation of these tools and their related biomarkers on an independent data set of memory clinics origin.

In the present investigation, we tested the performance of a recent analysis method from Chincarini and colleagues [19] that was trained on 1.5-T MRI data from the ADNI database. The method was applied to a set of images coming from the DESCRIPA study (www.descripa.eu), a European effort sponsored by the European Alzheimer's Disease Consortium (EADC) and funded by the European Community.

2. Materials and methods

At the time of analysis, we were blind to the clinical assessment and to the subjects' metadata, except for their age. The first step was testing the procedure on a blind, "clinical-grade" image set and comparing the results to those achieved on the ADNI data set. We then tested the hypothesis that the MCI cohort could include a subgroup of subjects with neuroimaging characteristics of early AD mixed with subjects without morphological signs of neurodegeneration. To this end, we trained our procedure onto two different data sets: one consisting of control (CTRL) and AD subjects and one consisting of MCI subjects.

2.1. Subjects used for training

Data used in the preparation of this article were obtained from the ADNI database. The ADNI was launched in 2003 by the NIA, the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the U.S. Food and Drug Administration (FDA), private pharmaceutical companies, and nonprofit organizations. For up-to-date information, see www.adni-info.org.

Statistical data of the included subjects are summarized in Table 1. All subjects were required to have baseline and at least 2 years of clinical information available. Healthy CTRL and AD subjects were selected only if they had the same diagnosis at follow-up.

Images for the MCI cohorts were taken at baseline, and they were divided into "MCI converters" (MCI-*co*) and "MCI nonconverters" (MCI-*nc*) according to the 2 years of clinical follow-up.

Training subjects were divided into two sets: a training set (A) composed of 333 age- and sex-matched subjects (namely 189 CTRL and 144 AD) and a training set (B) consisting of 302 MCI subjects, 136 of which converted to AD within 2 years. Sample sizes and demographics used in the training are the same as in Chincarini et al. [19].

2.2. Subjects used for validation

Subjects for validation were selected from the DESCRIPA study. The DESCRIPA study aims at developing clinical criteria and screening guidelines for AD in the predementia stage. Recruiting centers were selected from EADC members in 11 European countries and included 20 memory clinics specialized in the diagnosis and treatment of memory disorders.

The inclusion criteria basically consisted of outpatients aged 55 years or older that were newly referred for cognitive complaints to a European center dedicated to the evaluation of cognitive disorders. All referrals were considered, including self- or relative-referral, referral from a general practitioner, and referral from first-level neurological or geriatric clinics. Cognitive complaints mainly included memory complaints, but they could also include difficulties in other cognitive domains, such as attention and orientation.

Exclusion criteria were dementia or any somatic, metabolic (e.g., vitamin deficiency; endocrine untreated disorders; and kidney, liver, or heart failure), psychiatric, or neurological disorder that may cause cognitive impairment (e.g., cerebrovascular accidents, neurodegenerative diseases such as Parkinson's disease, severe head trauma, brain tumor, history of alcohol abuse, severe depression). In more detail, dementia was excluded by the clinical interviews with patients and caregivers and by means of formal questionnaires assessing the basic and instrumental activities of daily living as outlined in the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS)-Alzheimer's Disease and Related Disorders Association (ADRDA) and Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) criteria [20,21].

The DESCRIPA study was designed to reflect routine clinical practice such that findings should allow for easy implementation into a clinical routine. Centers agreed on the collection of a minimal data set whereas other tests were optional depending on local clinical practice and local funding possibilities for data collection. For that reason, certain variability among centers in cognitive tests, clinical rating scales and neuroimaging tools was allowed. Further information on clinical tests in the DESCRIPA study can be found in Visser et al. [22].

DESCRIPA centers recruited subjects between January 2003 and June 2005. Subjects were classified as having subjective memory complaints (SMC), nonamnestic MCI (naMCI), and amnestic MCI (aMCI), with the MCI subjects also comprising the deficit profile (single- and multidomain). Magnetic resonance images [23,24], CSF [25], single-photon emission computed tomography (SPECT) [26,27], and electroencephalography [28] data in the DESCRIPA population have already been published.

The MRI reader was blind to the diagnosis so that, within the framework of a nondemented naturalistic population, we were blind to the baseline and the follow-up clinical assessment.

2.3. Neuropsychology and subgroup definition

Digital MRI was available at baseline in 245 subjects with cognitive complaints (including aMCI, naMCI, and SMC). All subjects underwent a standard battery of examinations, including clinical history, medical and neurological examinations, laboratory tests, functional evaluation using the Clinical Dementia Rating (CDR) scale, rating scales for depression and neuropsychiatric symptoms, a neuropsychological test battery, and structural neuroimaging.

General cognition was assessed using the Mini-Mental State Examination (MMSE). Depression was assessed by the 15-item Geriatric Depression Scale or the Center for Epidemiologic Symptoms of Depression (CES-D) scale [29]. A depressive trait was defined according to the standard cutoff of each scale.

In each center, a battery of neuropsychological tests was performed to assess cognitive performance in the domains of memory, language, executive function and attention, and visuoconstruction. Raw scores were converted to age, education, and gender corrected z scores according to locally collected or published normative data. At baseline, patients were classified into three groups on the basis of test performances in these cognitive domains. Impairment was defined as a z score of -1.5 or lower.

Subjects without impairment in any domain were classified as SMC. Subjects with impairment in the memory domain only or with impairment in the memory domain plus impairment in nonmemory domains were defined as aMCI. Subjects with impairment in one or more nonmemory domains were defined as naMCI.

The aMCI and naMCI subgroups could include subjects with either single or multiple domain deficit. Because of variability among the neuropsychological protocols, the tests used to define MCI subtypes varied between centers. The tests for memory were the learning measure and delayed recall measure of the Rey Auditory Verbal Learning test, the Selective Reminding test [30], or the Grober-Buschke test [31].

The tests for language were 1-minute verbal fluency for animals; 2-minute verbal fluency for animals; or 1-minute verbal fluency for fruits, animals, and car trades. Executive function and attention were assessed with the Trail Making Test part A and B (TMT A and B) in all centers. The tests for visuoconstruction were the copy subtest of the Rey-Osterrieth complex figure and the copy of figures from the Mental Deterioration Battery [32].

These subjects were followed up with the same clinical and neuropsychological assessments for 1–3 years. At

Table 2
Demographics and clinical findings for the validation data set from the DESCRIPA study

Cohort description	Sample size	Age (years)	M/F	MMSE	NC/CO	Time to AD* (years)	Deficit profile sd/md†	*APOE* +		*APOE* −	n.t.
								ε4	ε2		
SMC	53	70.2 (7.6)	21/32	27.1 (2.3)	51/2	1.7 (0.5)	–	15	7	22	9
naMCI	56	69.9 (7.7)	16/40	26.9 (2.4)	48/8	1.6 (0.9)	36/20	14	8	31	3
aMCI	85	70.2 (8.3)	44/41	27.3 (2.2)	55/30	1.6 (1.1)	37/48	29	3	30	23
Total	194	70.1 (7.9)	81/113	27.1 (2.3)	154/40	1.6 (1.0)	73/68	58	18	83	35

Abbreviations: M, male; F, female, MMSE, Mini-Mental State Examination score; AD, Alzheimer's disease; SMC, subjective memory complaints; naMCI, nonamnestic MCI; aMCI = amnestic MCI; sd, single domain; md, multidomain; APOE+, APOE positive on allele ε2/4; APOE−, APOE without ε2 or ε4 alleles; n.t., APOE not tested; MRI, magnetic resonance imaging; NC, nonconverters; CO, converters.

NOTE. Subjects described here are those for which their MRI did pass inclusion criteria for automatic analysis. Column NC/CO shows the number of converter/nonconverter subjects for each cohort.

*Average time to conversion after baseline examination refers to converted subjects only.

†Deficit profile class is based on neuropsychological tests and is divided into sd/ md, where sd refers to memory deficit only. Deficit profile numbers in the last row (total) refer to naMCI and aMCI only.

follow-up visit, the onset of dementia of the Alzheimer type was diagnosed according to the NINCDS-ADRDA and DSM-IV criteria. Those patients fulfilling the diagnostic criteria for MCI were labeled as MCI-*nc* whereas subjects with no objective cognitive deficit were termed SMC, thus following the same rules as for baseline classification.

2.4. Image inclusion criteria

ADNI selected a magnetization-prepared rapid acquisition with gradient echo (MPRAGE) sequence, which was defined across selected systems from GE Healthcare, Philips Medical Systems, and Siemens Medical Solutions with the objective of minimizing cross-platform differences. Platform-specific protocols were distributed through the MRI vendors to minimize inconsistencies expected to arise from building the protocol manually on individual scanners. In addition, the ADNI protocol involved a specific scanner calibration procedure, subject positioning, and intensity postprocessing to correct for various image nonuniformities.

Unlike ADNI, the DESCRIPA study did not include a specific protocol for structural MRI scans. Structural images in the study are rather heterogeneous and range from separate slices (two-dimensional scans) to fully volumetric MPRAGE-like scans. The MPRAGE-like scan type was explicitly stated in only 47% of the files, and images were acquired with four scanners: Siemens Magnetom Expert (36%), Siemens Sonata Vision (13%), Phylips Gyroscan NT (43%), and Phylips Intera (8%); these are scanner models not included in the ADNI-1 MRI protocols. Furthermore, voxel sizes are more heterogeneous than those found in ADNI images. Voxel volume for the ADNI images used in the training phase of this work ranges from 1.0 to 1.8 mm^3 with a mean of 1.4 mm^3, whereas in the DESCRIPA images we find 0.9–2.9 mm^3 with a mean of 1.3 mm^3.

Table 3
Exclusion criteria used to filter raw scans in input

Property	Value (sample size)	Number of discarded images
MRI scan	Available (221), not available (24)	24
Scan type	Volumetric (215), 2D (6)	6
Voxel size	<1.6 mm (197), >1.6 mm (18)	18
Image	Subject misplacement (1)	1
Image	Sizable artifacts (2)	2

Abbreviation: MRI, magnetic resonance imaging.

NOTE. Of the 215 available volumetric images, we had to discard 21 because of various nonidealities that could have severely compromised the analysis result of the subject. "Voxel size" refers to the maximum value of the MRI voxel size in either direction. "Subject misplacement" refers to images that may contain anatomical parts other than the subject head (subject shoulders) or omit part of the subject head. "Sizeable artifacts" refers to artifacts for which the extent and severity may compromise intensity and texture measures.

By design, our method works on volumetric scans, from which it selects small, salient regions where intensity and texture features are computed. For this reason, we had to define image inclusion criteria.

Of the $N_{tot} = 245$ subjects with complete clinical follow-up, only $N_{baseline} = 215$ baseline scans were found to be volumetric MRI potentially suitable for automatic analysis. In addition, $N_{disc} = 21$ had to be discarded because they exhibited abnormal characteristics, such as sizeable artifacts or excessive voxel-size anisotropy, which rendered them unsuitable for automatic processing and measurement.

We established the criteria for image inclusion as follows: MRI was volumetric, the maximum voxel dimension was $\leq$1.6 mm, and there was an absence of sizeable artifacts.

With these criteria, we were left with $N_a = 194$ images to analyze. Table 2 details the demographics and clinical findings for the N_a subjects and Table 3 details the number of discarded images and the grounds on which they were excluded. We checked that the statistical properties of the subjects' ensemble were not significantly altered by the image selection; therefore, the image inclusion criteria cannot be considered as a population enrichment filter.

Except for the scan type (volumetric or two dimensional) and the selection on the maximum voxel size, analysis progressed without human intervention.

Table 4
Main gray matter structures captured in the VOIs and the size of the box used to encapsulate them

	VOI	Main anatomical structure	Size (mm)
	1	Hippocampus, entorhinal cortex (right)	30 × 70 × 30
	2	Hippocampus, entorhinal cortex (left)	30 × 70 × 30
	3	Amygdala (right)	34 × 34 × 34
*	4	Amygdala (left)	34 × 34 × 34
	5	Middle and inferior temporal gyrus (right)	30 × 50 × 30
	6	Middle and inferior temporal gyrus (left)	30 × 50 × 30
	7	Insula, superior temporal gyrus (left)	30 × 60 × 30
†	8	Rolandic (right)	36 × 36 × 36
	9	Rolandic (left)	36 × 36 × 36

Abbreviation: VOI, volume of interest.
*Potentially significant regions.
†Control regions.

2.5. Image processing

Image processing closely follows the method detailed in Chincarini et al. [19]. We summarize here the procedure applied to each magnetic resonance (MR) image up to the extraction of its feature data set, which was used for classification purposes.

MR images underwent a series of filters designed for noise reduction, volume normalization, anatomical structure registration, and gray-level intensity equalization. No cohort-specific template was used for the Montreal Neurological Institute (MNI) spatial normalization. As a result of the preprocessing steps, images were aligned to the MNI reference, were volume corrected, and the mean gray-level intensities of the three major brain constituents (CSF, gray matter [GM] and white matter [WM]) were matched to reference values.

Once preprocessed, each image was sampled with nine volumes of interest (VOIs; see Table 4), seven of which were placed in brain areas relevant to the AD and two of which were placed in the Rolandic area and served as control regions. Intensity and texture information from these nine VOIs constitutes the feature set used in the classification step.

To preserve accurate anatomical alignment, the VOIs were extracted from each subject by means of a further rigid registration using templates (i.e., a registration of several reference VOIs onto the subject MNI-normalized brain). There are typically five to nine templates per VOI. They are design to capture the structural differences among subjects with varying degrees of neurodegeneration, ranging from healthy elderly to severe AD. Details on the generation of VOI templates can be found in reference Calvini et al. [12].

Although overall volume normalization is calculated using a single MNI reference, with VOI templates we can compensate for anatomical differences among individuals and, once the WM/GM/CSF intensity mapping is performed, we can directly compare the VOI content.

During the training phase, the random forest (RF) classifier computed "variable importance measures" [33,34], which are a byproduct of the RF classification procedure. This quantity was used for selecting the most probable predictors; therefore, it was useful to prune irrelevant or confounding variables [35].

Variable importance measures were then combined to give the important features map (IFM), which assigned a weight to each feature, measuring how relevant it was to the cohort separation. The application of a thresholded IFM on the raw feature set gave a reduced feature set, which amounted to approximately to 10^4 for each MRI. It is this reduced data set that was finally fed to a set of support vector machine (SVM) classifiers, the output of which was the classification index (CI). The CI is a number ranging from −1 (AD) to 1 (CTRL), and it was assigned to each subject.

2.6. Experiment

Two experiments were performed, differing only in the training set choice. Classifiers were trained either on data set A (i.e., a training set consisting of confirmed CTRL and AD subjects) or B (training set consisting of baseline MCI subjects, some of who were found to be converted to AD after a follow-up of 2 years; see Table 1). All training subjects came from the ADNI population.

The purpose was to test the discrimination ability of the information coming from the MCI population, which may contain late converter subjects within the *nc* cohort, and compare it with that coming from a clinically robust set. For each experiment, we calculated the related IFM and hence the feature subset used to train the SVM classifier. On the basis of the SVM classifier algorithm from training sets A and B, we calculated a CI for each subject in the DESCRIPA study.

DESCRIPA subjects' data were never included in the training phase. At the time of analysis, we were blind to all clinical data except for the subjects' ages, which were used only to check that the average training group age was not significantly different from that of the validation group.

Therefore, each subject from the DESCRIPA study was labeled with two CI indexes: the first coming from the classifiers trained on data set A and the second coming from the classifiers trained on data set B.

We tested the performance of our approach on five different groupings of the DESCRIPA subjects. We tested converters versus nonconverters within a single cohort, and we merged cohorts together. Groupings have different sample sizes and may contain mixed cohorts. They are detailed in Table 5.

Table 5
Classification performance for the different cohort grouping and training set

Test	Cohorts	Sample size	AUC(σ)		*Sp*/*Sn*		$AUC^B \neq AUC^A$
			Training set A	Training set B	Training set A	Training set B	(*P* value)
i	$naMCI_{nc}$/$naMCI_{co}$	48/8	0.60 (0.10)	0.76 (0.08)	0.62/0.62	0.75/0.69	.03
ii	$aMCI_{nc}$/$aMCI_{co}$	55/30	0.73 (0.05)	0.72 (0.06)	0.80/0.64	0.70/0.71	.69
iii	MCI_{nc}/MCI_{co}	103/38	0.74 (0.04)	0.76 (0.04)	0.68/0.76	0.66/0.75	.34
iv	SMC_{nc}/All_{co}	51/40	0.84 (0.04)	0.90 (0.03)	0.82/0.78	0.93/0.74	.01
v	All_{nc}/All_{co}	154/40	0.77 (0.04)	0.81 (0.03)	0.72/0.75	0.68/0.81	.08

Abbreviations: AUC, area under the receiver operating characteristic curve; *Sp*, specificity; *Sn*, sensitivity; naMCI, nonamnestic MCI; aMCI, amnestic MCI; MCI, mild cognitive impairment; SMC, subjective memory complaints; All, all cohorts (SMC + aMCI + naMCI).

NOTE. *Sn* and *Sp* values are given for both curves with a balanced cutoff. *nc* and *co* subscripts indicate the nonconverter and converter fraction of the cohort. The last column indicates whether it is statistically significant that the AUC values obtained when training on set B differ from those obtained when training on set A.

3. Results

3.1. Classification

The descriptive statistics of the CI according to MCI subgroup at baseline is shown in Figure 1. Subjects with more severe neuropsychological impairment show more MTL atrophy.

Using the CI statistics, we find SMC and naMCI single-domain populations indistinguishable at the 95% confidence level for both experiments. The same test performed on SMC and naMCI multidomain populations rejects the null hypothesis only in experiment B ($P < .006$).

Without subdivision in single/multiple domains, CI data coming from experiment B show that SMC, aMCI, and naMCI cohorts are all significantly different ($P < .02$ for SMC and naMCI, $P < 10^{-4}$ for aMCI vs.SMC and naMCI). The same test performed on experiment A only rejects the null hypothesis for aMCI against SMC and naMCI.

The overall accuracy of the CI for predicting AD-type dementia after 2 years in each subgroup and in the total sample is shown in Table 5.

Classifiers seem to perform better when obtained from training set B, although this difference is only significant in the comparison i and iv (i.e., when SMC and

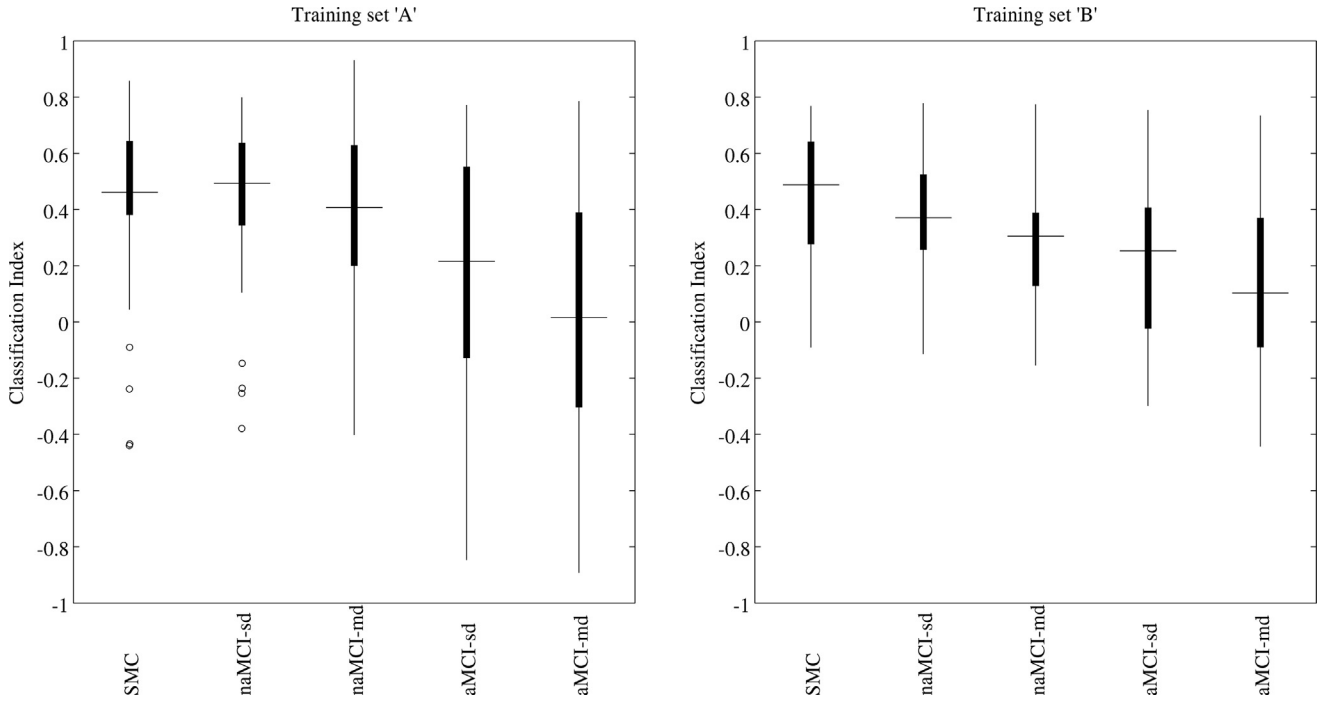

Fig. 1. Descriptive statistics of the CI for the DESCRIPA subjects detailed in Table 2. The left and right plots refer to data coming from classifiers trained on CTRL/AD cohorts and on MCI–*nc*/MCI–*co* cohorts, respectively. A CI value close to 1 represents a normalcy condition whereas a value close to −1 represents an AD-like condition. The central mark is the median, and the edges of the thicker line are the 25th and 75th percentiles. Whiskers extend up to approximately ±2.7σ if the data were normally distributed. Most extreme data points are plotted individually. CNTL, control; AD, Alzheimer's disease; CI, classification index; SMC, subjective memory complaints; naMCI-sd, nonamnestic MCI, single domain; naMCI-md, nonamnestic MCI, multidomain; aMCI-sd, amnestic MCI, single domain; aMCI-md, amnestic MCI, multidomain; MCI-*nc*, mild cognitive impairment nonconverter; MCI-*co*, mild cognitive impairment converter.

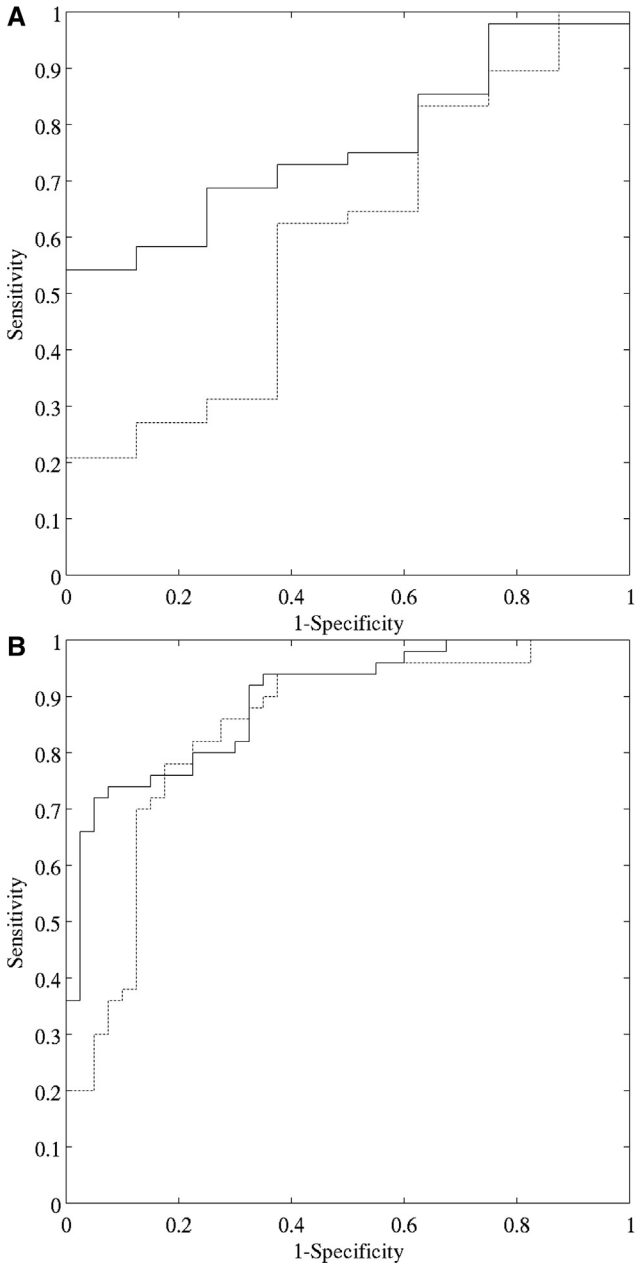

Fig. 2. ROC curves for test i and iv with training set A and B (see Table 5): (a) test i with training set A (solid line) and B (dashed line); (b) test iv with training set A (solid line) and B (dashed line). ROC, receiver operating characteristic.

naMCI are included in the test). For these tests, receiver operating characteristics (ROC) curves are shown in Figure 2.

3.2. Relevant regions

A significant difference when the procedure is trained with set A or B comes from the IFM, which weights those image features that are passed to the SVM classifier. Figure 3 shows the thresholded IFM superimposed on the MNI reference image.

Figure 3a shows that the IFM trained on set A is localized in the head of the hippocampus and amygdala (axial section) and in the tail of the hippocampus (parahippocampal gyrus), close to the fornix (sagittal section). On the other hand, Figure 3b shows that the IFM trained on set B is much more sparse within the temporal lobe, including some small clusters in the middle and inferior temporal gyri (Brodmann areas 21 and 22), in the body of the hippocampus and in the fusiform gyrus in both hemispheres, beside the head of hippocampus, and in a small part of the amygdala mainly in the left hemisphere.

Differences between IFM(A) and IFM(B) can be more readily appreciated when looking at a single VOI. Figure 4 shows a series of sagittal, axial, and coronal sections of a sample right hippocampus (VOI 1) together with the normalized, unthresholded IFM(A) and IFM(B) intensities. IFM(B) is clearly much sparser and involves many more voxels than its counterpart.

3.3. Statistics

On the CI distribution plotted in Figure 1, we tested group discrimination by means of a two-sample Kolmogorov-Smirnov test [36], which has the advantage of making no assumption about the distribution of data. In fact, CI distribution is bounded by definition on the interval $[-1, 1]$, which makes the distributions unsuitable to be tested with the more common t test. Tests on SMC and naMCI single-domain populations found them indistinguishable at the 95% confidence level for both experiments. The same test performed on the SMC and naMCI multidomain only rejects the null hypothesis in experiment B ($P < .006$).

As far as converter/nonconverter discrimination is concerned, as summarized in Table 5, the cutoff point used to compute specificity and sensitivity values was chosen to minimize the distance

$$d = \sqrt{(1-Sn)^2 + (1-Sp)^2}$$

thus achieving a balance between sensitivity (Sn) and specificity (Sp). For easier comparison to literature works, we report the results using another widely accepted cutoff rule (Youden index) in Table 7.

We also checked whether there was any performance difference when considering images coming from a single scanner, and we did not find any. Given the relatively few images involved, the check was only performed on test iii and v, in which we singled out images coming from the Phylips Gyroscan NT scanner, which accounts for approximately 43% of all scans.

Image processing was performed on a dedicated computational farm running the LONI pipeline software (www.loni.ucla.edu) using MATLAB (www.mathworks.com) and ITK (www.itk.org) as algorithm libraries. All

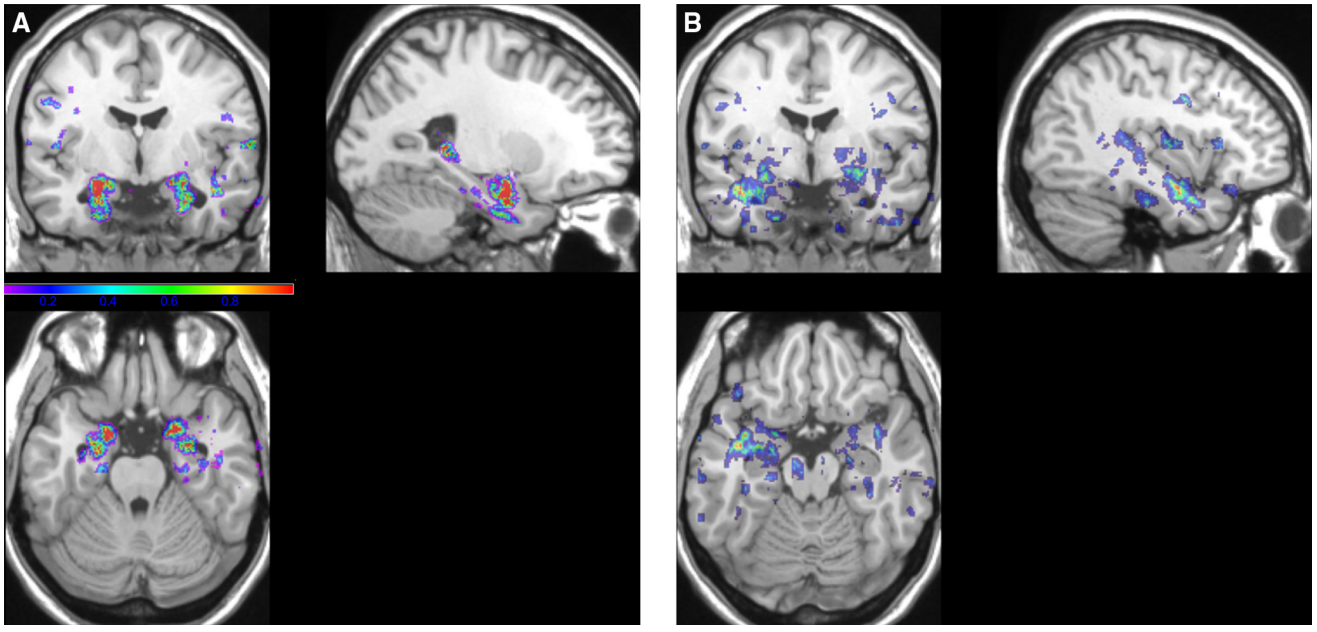

Fig. 3. Thresholded IFM when classifiers are trained on (a) set A and (b) set B. Axial (bottom left), coronal (top left), and sagittal (top right) sections show the IFM superimposed on the MNI reference image. Color scale is proportional to the normalized IFM value after thresholding. For displaying purposes, normalization was performed separately for IFM (A) and (B). IFM, important features map; MNI, Montreal Neurological Institute.

statistical analyses were performed within the MATLAB environment. Error estimation and *P* value statistics for the area under the ROC curve (AUC) values in Table 5 are computed according to Hanley and McNeil [37,38].

4. Discussion

The present study shows that automatic analysis of atrophy on structural MRI can help identifying those with underlying AD pathology with satisfactory accuracy in

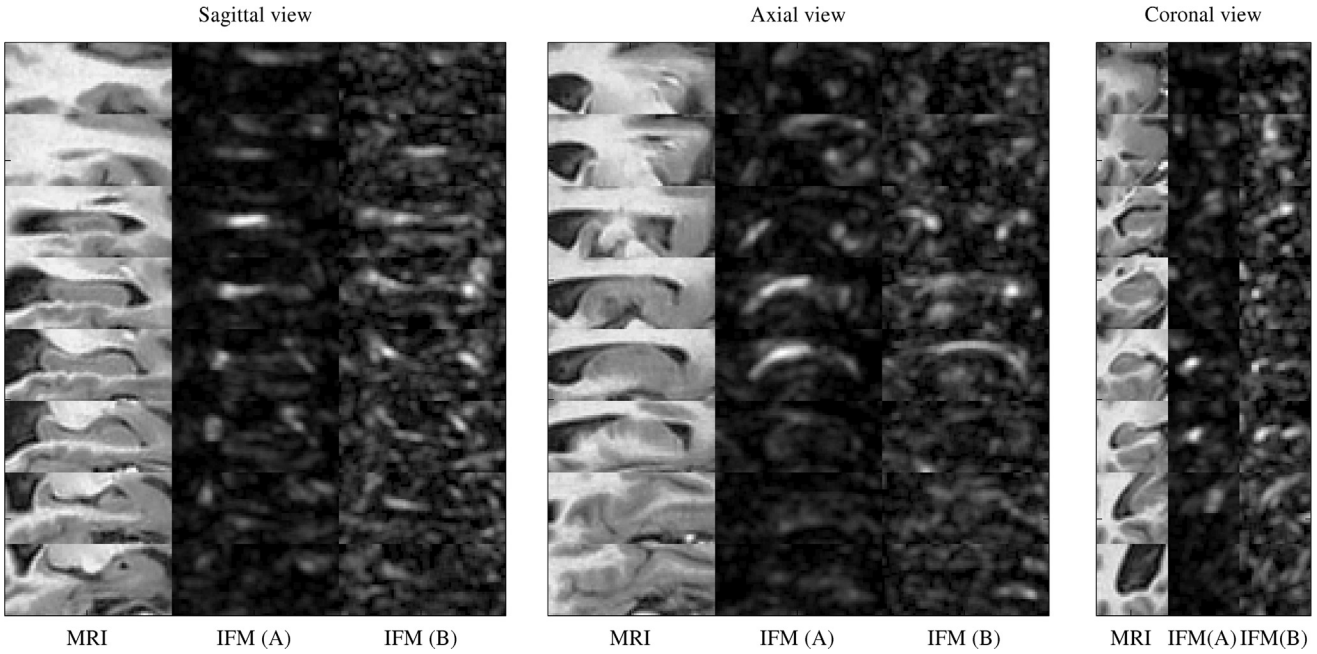

Fig. 4. From left to right: sagittal, axial, and coronal sections of a sample hippocampal area (VOI 1) together with the unthresholded IFM. Sagittal view top-to-bottom = right-to-left; axial view top-to-bottom = top-to-bottom; coronal view top-to-bottom = anterior-to-posterior. The left, central, and right columns of each image are, respectively, MRI sections, IFM based on CTRL/AD subjects (training set A), and IFM based on MCI-*nc*/MCI-*co* subjects (training set B). Intensity scale in IFM images is proportional to the feature relevance and was separately normalized on a scale (0–1) for easier reading. IFM, important features map; CNTL, control; AD, Alzheimer's disease; MCI-*nc*, mild cognitive impairment nonconverter; MCI-*co*, mild cognitive impairment converter; VOI, volume of interest.

Table 6
Performance comparison among available MRI measures and our procedure (experiment B) computed on a common DESCRIPA data set

Test	Cohorts	Common sample size†	$AUC^B(\sigma)$	Best matching measure*	AUC* (σ)	$AUC^B \neq AUC^*$ (P value)
i	$naMCI_{nc}/naMCI_{co}$	42/6	0.73 (0.09)	Lat. Ventr.	0.80 (0.08)	.29
ii	$aMCI_{nc}/aMCI_{co}$	54/30	0.72 (0.06)	LEAP (ivc)	0.71 (0.06)	.88
iii	MCI_{nc}/MCI_{co}	100/37	0.76 (0.04)	LEAP (ivc)	0.72 (0.05)	.26
iv	SMC_{nc}/All_{co}	49/40	0.90 (0.03)	MTA	0.81 (0.05)	$<10^{-3}$
v	All_{nc}/All_{co}	140/39	0.80 (0.04)	Hip. man.	0.71 (0.04)	.005

Abbreviations: LEAP (ivc), learning embeddings for atlas propagation with intracranial volume correction; Hip. man. (ivc), hippocampal volume, manual tracing with intracranial volume correction; Lat. Ventr., lateral ventricle volume; MTA, global atrophy visual assessment (Sheltens' visual rating scale); AUC, area under the receiver operating characteristic curve; naMCI, nonamnestic MCI; aMCI, amnestic MCI; MCI, mild cognitive impairment; SMC, subjective memory complaints; All, all cohorts (SMC + aMCI + naMCI).

NOTE. *nc* and *co* subscripts indicate the nonconverter and converter fraction of the cohort. For each test, i–v, we evaluated the performance of all available measures and chose that with the highest AUC value to compare it to our results. Because each measure was available on a subset of all subjects, the † shows the common number of subjects on which both statistics were available: these are the shared subjects between the best matching measure and our procedure. Compared with Table 5, we get a smaller number of subjects as the result of the set intersection. The last column indicates whether the two AUC values are significantly different.

*Measure and AUC refers to the best matching outcome among the available analyses.

a naturalistic population of subjects presenting with subtle or mild cognitive complaints to a memory clinic.

The first remark is that these data were collected by means of different equipment in several European countries and that MRI scans were performed on clinical demand and not for research purposes. As such, the results are even more important. Moreover, the studied population received no other screening procedure than having subjective or objective MCI (i.e., it is not the overselected population of clinical trials). In fact, the DESCRIPA design did not include the acquisition of MRI images; therefore, no selection based on image quality or availability was applied to the population.

The lack of a common protocol and of dedicated image quality assessment can be used to test the performance of an automatic analysis method when input images are "clinical quality" rather than "research quality". The paid price is that not all images can be accepted, depending on the tolerances of the analysis method. Nevertheless, this study shows that consistent results can be achieved when the analysis procedure is based on a sufficiently large number of images coming from different centers and scanners (as is the case with ADNI images).

Confounding factors, such as mild-to-moderate cerebrovascular disease, depression, and drug use, may have influenced brain atrophy to some extent. Nevertheless, the automatic procedure is able to capture those atrophy peculiarities in the temporal lobe, allowing group identification with satisfying accuracy.

An issue deserving particular attention is that training the classifiers on AD patients and healthy CNTL of the ADNI population has a similar effect on accuracy as classifiers trained on MCI–*co*/*nc* when aMCI converter patients are compared with aMCI nonconverters (test ii). This finding stresses that AD pathology features can be well identified in aMCI if a memory deficit is identified first.

The same assumption is not confirmed when an objective memory deficit cannot be demonstrated, as in naMCI patients and in SMC subjects (tests i and iv), in whom training the classifier on a MCI population rather than on AD and healthy CNTL led to a very significant increase in diagnostic accuracy. Thus, the atrophy pattern may be different in patients converted to AD dementia if their clinical presentation at baseline does not fit the "classical" episodic memory deficit paradigm.

4.1. Comparison to other measures

DESCRIPA data include several data fields and imaging types. Following the data usage policy, the DESCRIPA steering committee approved other analyses on the same MRI data set, which were performed blind to the subjects' clinical assessment at baseline and follow-up.

Analyses included an application of the LEAP algorithm to the hippocampus [39], hippocampal volume measured with manual segmentation, global atrophy visual assessment (medial temporal lobe atrophy [MTA]) [40], and volumetric assessment in the lateral ventricle [41,42]. Some measures were given with and without intracranial volume correction (IVC). A thorough comparison of these methods can be found in Clerx et al. [43].

We compared the performance of all analyses to the results of experiment B by means of the AUC. For each of the tests in Table 5, we computed AUC values on the common data set between our procedure and any of the aforementioned measures. The best matching measure (i.e., with the higher AUC) was taken for comparison. Results are detailed in Table 6.

We see that for each test, there is an analysis method competitive with ours (AUC^B) but none of them is consistently competitive in all tests. In addition, our procedure performs significantly better on tests iv and v, suggesting that the inclusion of SMC subjects is best taken into consideration in our training set.

Further qualitative comparisons on an equivalent DESCRIPA data set but with radically different approaches can be found in Babiloni et al. [28] and Nobili et al. [27].

Table 7
Sensitivity and specificity values with different cutoffs

	Balanced cutoff			Youden index		
Test	CI value	*Sp*/*Sn*	PPV/NPV	CI value	*Sp*/*Sn*	PPV/NPV
i	0.30	0.75/0.69	0.94/0.29	0.37	1.00/0.54	1.00/0.27
ii	0.11	0.70/0.71	0.81/0.57	0.11	0.70/0.71	0.81/0.57
iii	0.14	0.66/0.75	0.86/0.49	0.11	0.63/0.78	0.85/0.51
iv	0.36	0.93/0.74	0.93/0.74	0.37	0.95/0.72	0.95/0.73
v	0.14	0.68/0.81	0.91/0.47	0.36	0.93/0.56	0.97/0.36

Abbreviations: CI, classification index; *Sp*, specificity; *Sn*, sensitivity; PPV, positive predictive value; NPV, negative predictive value.

NOTE. Cutoff value (CI threshold), *Sp*, *Sn*, PPV, and NPV for the tests are detailed in Table 5 using two different cutoff rules on the receiver operating characteristic curves. Statistics evaluated on values from experiment B.

Babiloni et al. [28] show that electroencephalographic (EEG) analysis can convey group discrimination among SMC, naMCI, and aMCI. Their results suggest that along the transition line between normal and pathological aging, aMCI subjects present peculiar alterations of global neural synchronization and that the classification of subjects into MCI subtypes and SMC has a neurophysiological basis.

Our findings are in accordance with their conclusions because we readily see that aMCI is the most diverse group among the three cohorts. In addition, the fact that classifiers trained on MCI generally deliver better results could be explained by the differences in the IFM obtained on MCI subjects when compared with the one computed on CTRL/AD.

IFM differences translate into feature selection (i.e., information captured in different regions and by different intensity and texture filters). Therefore, it is reasonable to assume that training on MCI subjects is better suited for capturing transitional features that are either flattened out in the more advanced stage of the pathology or are not significant enough during the normal aging process.

4.2. Study limitation

From a methodological point of view, this work does not present a novel analysis but follows the procedure published in Chincarini et al. [19], except for the use of MCI subjects for training and their comparison to CTRL/AD on the clinically relevant regions (IFM).

The choice of the procedure was driven by its simplicity and robustness while providing reasonable accuracy so that it could be used to assist in the clinical classification of subjects in a memory clinic setting. However, it could be argued that more sophisticated approaches (i.e., using nonlinear registration algorithms) could have improved the performance. This issue was already addressed in Chincarini et al. [19], although it is likely that complex procedures are more susceptible to image quality issues.

In this study we had to limit the number of images to those complying with certain quality requirements. With these constraints, the analysis procedure did prove satisfactory but the presence of sizeable artifacts or the threshold on voxel anisotropy was decided arbitrarily. However, these constraints depend on the particular algorithm chosen to analyze the data; for instance, a procedure looking for ventricular enlargement could be more robust with respect to the image voxel size and presence of artifacts than a procedure that looks for subtle intensity and texture variances in a small set of voxels.

The present study could be extended in two directions. From a clinical point of view, we would welcome a longer follow-up timeframe, which could improve the prediction ability on AD converters. Data from the ADNI-II phase could well comply with such a requirement.

From a methodological point of view, different approaches could be used to improve on the simple MRI data on the algorithm side and on the selection criteria. Bearing in mind the applicability of the method in everyday clinical practice, we envisage a prognostic improvement by means of a multidomain approach, possibly with CSF analysis.

5. Conclusion

We validated a fully automatic analysis and classification method using only structural MR images on a set of clinical-grade images coming from the DESCRIPA study. The DESCRIPA enrolled SMC, aMCI, and naMCI subjects and acquired a series of neuroimaging and clinical data with a follow-up in the range of 1–3 years.

Each of the 215 baseline digital MRI from the DESCRIPA study was checked against a minimal set of basic quality parameters to determine its eligibility for automatic analysis. A total of 194 images survived the inclusion criteria and were analyzed.

For each subject, the analysis assigns a number (CI) that correlates with the clinical assessment and probability of converting to the AD state in a timeframe of 1–2 years. The method was developed on 1.5-T MR images coming from the ADNI database, and classifiers were trained on two separate data sets: one consisting of CTRL and AD subjects and one consisting of MCI subjects. It was shown that the classification performance on DESCRIPA subjects benefits from a dedicated training on MCI subjects.

Further analysis of the relevant regions (IFM) when the procedure was trained on MCI versus CTRL/AD subjects revealed subtle differences in spatial and intensity distribution that reflect on the classifiers ability to better distinguish between converters and nonconverters, particularly when SMC subjects were included in the test.

Our findings suggest that, even from the structural point of view, the cognitive impairment stage (be it SMC, naMCI, or aMCI) cannot simply be regarded as a transitional phase between normalcy and dementia. A dedicated training on specific and well-selected populations significantly improves the odds of correct classification and prediction.

Comparison to other analysis methods on the very same subjects shows that our approach is competitive and has the benefit of being completely automatic.

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

This research was supported by Istituto Nazionale di Fisica Nucleare (INFN), Italy. This research was also supported by grants to L.R., P.B., and M.E. from Università degli Studi di Genova, Italy.

Validation data collection and sharing for this project was funded by EADC and by the European Commission as part of the 5th Framework Programme (DESCRIPA, QLK-6-CT-2002-02455).

EADC is a nonprofit organization networking over 50 European centers of clinical and biomedical research excellence working in the field of AD and related dementias. It aims to provide a setting in which to increase the basic scientific understanding of AD and to develop ways to prevent, slow, or ameliorate the primary and secondary symptoms of AD. Funding for the original realization of this network was received from the European Commission, which supported work toward standardization of diagnostic criteria, assessment tools, and data collection methods with a view to this being followed by a trial period involving the testing and practical application of the tools agreed.

Data collection and sharing for this project was funded by ADNI (National Institutes of Health Grant U01 AG024904). ADNI is funded by NIA, NIBIB, and through generous contributions from the following: Abbott; AA; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences, Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec, Inc.; Bristol-Myers Squibb Company; Eisai, Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche, Ltd. and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO, Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development, LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer, Inc.; Servier; Synarc, Inc.; and Takeda Pharmaceutical Company.

The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

All authors disclose any actual or potential conflicts of interest, including any financial, personal, or other relationships with other people or organizations that could inappropriately influence their work. All experiments were performed with the informed consent of each participant or caregiver in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Local institutional ethics committees approved the study.

RESEARCH IN CONTEXT

1. Systematic review: A workgroup commissioned by the NIA-AA recently published research criteria for preclinical AD. Literature works show that there is an increasing need to validate these criteria and the biomarkers outside of the research environment. Most works on image-based biomarkers derive their results on research-driven image databases, in which data quality and consistency are guaranteed by shared protocols and study criteria.
2. Interpretation: Validation on a clinical-grade data set would provide information on the marker clinical usefulness and on the procedure robustness. Our test shows that the marker performance is satisfactory, provided some minimal criteria are met. These findings support a more widespread use of automatic procedures in everyday clinical practice.
3. Future directions: We should address the potential problems arising from less restrictive inclusion procedures on the general population and the relevancy of biomarker outcome on clinical diagnosis and prognosis.

References

[1] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007; 6:734–46.

[2] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol 2010;9:1118–27.

[3] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–9.

[4] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's

Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.
[5] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270–9.
[6] Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.
[7] Bateman RJ, Xiong C, Benzinger TLS, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med 2012;367:795–804.
[8] Mosconi L, De Santi S, Li J, Tsui WH, Li Y, Boppana M, et al. Hippocampal hypometabolism predicts cognitive decline from normal aging. Neurobiol Aging 2008;29:676–92.
[9] Clerici F, Del Sole A, Chiti A, Maggiore L, Lecchi M, Pomati S, et al. Differences in hippocampal metabolism between amnestic and non-amnestic MCI subjects: automated FDG-PET image analysis. Q J Nucl Med Mol Imaging 2009;53:646–57.
[10] Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. Brain 2008;131:681–9.
[11] Klauschen F, Goldman A, Barra V, Meyer-Lindenberg A, Lundervold A. Evaluation of automated brain MR image segmentation and volumetry methods. Hum Brain Mapp 2009;30:1310–27.
[12] Calvini P, Chincarini A, Gemme G, Penco MA, Squarcia S, Nobili F, et al. Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer disease. Med Phys 2009;36:3737–47.
[13] Heckemann RA, Keihaninejad S, Aljabar P, Gray KR, Nielsen C, Rueckert D, et al. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. Neuroimage 2011;56:2024–37.
[14] Aksu Y, Miller DJ, Kesidis G, Bigler DC, Yang QX. An MRI-derived definition of MCI-to-AD conversion for long-term, automatic prognosis of MCI patients. PLoS ONE 2011;6:e25074.
[15] Matsuda H, Mizumura S, Nemoto K, Yamashita F, Imabayashi E, Sato N, et al. Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer Disease. AJNR Am J Neuroradiol 2012;33:1109–14.
[16] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, et al. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLoS ONE 2011;6:e21896.
[17] Plant C, Teipel SJ, Oswald A, Böhm C, Meindl T, Mourao-Miranda J, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 2010;50:162–74.
[18] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert MO, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage 2011;56:766–81.
[19] Chincarini A, Bosco P, Calvini P, Gemme G, Esposito M, Olivieri C, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage 2011;58:469–80.
[20] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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 1984;34:939–44.
[21] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: American Psychiatric Association; 2000, p.886.
[22] Visser PJ, Verhey FRJ, Boada M, Bullock R, De Deyn PP, Frisoni GB, et al. Development of screening guidelines and clinical criteria for predementia Alzheimer's disease. The DESCRIPA Study. Neuroepidemiology 2008;30:254–65.
[23] Jacobs HIL, Visser PJ, Van Boxtel MPJ, Frisoni GB, Tsolaki M, Papapostolou P, et al. Association between white matter hyperintensities and executive decline in mild cognitive impairment is network dependent. Neurobiol Aging 2012;33:201.e1–8.
[24] van de Pol LA, Verhey F, Frisoni GB, Tsolaki M, Papapostolou P, Nobili F, et al. White matter hyperintensities and medial temporal lobe atrophy in clinical subtypes of mild cognitive impairment: the DESCRIPA study. J Neurol Neurosurg Psychiatr 2009; 80:1069–74.
[25] Visser PJ, Verhey F, Knol DL, Scheltens P, Wahlund LO, Freund-Levi Y, et al. Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: a prospective cohort study. Lancet Neurol 2009;8:619–27.
[26] Nobili F, Frisoni GB, Portet F, Verhey F, Rodriguez G, Caroli A, et al. Brain SPECT in subtypes of mild cognitive impairment. Findings from the DESCRIPA multicenter study. J Neurol 2008;255:1344–53.
[27] Nobili F, De Carli F, Frisoni GB, Portet F, Verhey F, Rodriguez G, et al. SPECT predictors of cognitive decline and Alzheimer's disease in mild cognitive impairment. J Alzheimers Dis 2009;17:761–72.
[28] Babiloni C, Visser PJ, Frisoni G, De Deyn PP, Bresciani L, Jelic V, et al. Cortical sources of resting EEG rhythms in mild cognitive impairment and subjective memory complaint. Neurobiol Aging 2010;31:1787–98.
[29] Radloff L. The CES-D scale: A self-report depression scale for research in the general population. Appl Psychol Measurem 1977;1:385–401.
[30] Masur DM, Fuld PA, Blau AD, Thal LJ, Levin HS, Aronson MK. Distinguishing normal and demented elderly with the selective reminding test. J Clin Exp Neuropsychol 1989;11:615–30.
[31] Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. Neurology 1988;38:900–3.
[32] Carlesimo GA, Caltagirone C, Gainotti G. The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. The Group for the Standardization of the Mental Deterioration Battery. Eur Neurol 1996;36:378–84.
[33] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[34] Bylander T. Estimating generalization error on two-class datasets using out-of-bag estimates. Mach Learn 2002;48:287–97.
[35] Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. Comput Stat Data Anal 2008; 52:2249–60.
[36] Eadie W, James F. Statistical methods in experimental physics. 2nd ed. Singapore: World Scientific Publishing Company; 2006, pp. 313–7.
[37] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.
[38] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–43.
[39] Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D. LEAP: learning embeddings for atlas propagation. Neuroimage 2010;49:1316–25.
[40] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatr 1992;55:967–72.
[41] Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2002;17:143–55.
[42] Smith SM, De Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. J Comput Assist Tomogr 2001;25:466–75.
[43] Clerx L, van der Pol L, Rueckert D, de Jong R, van Schijndel R, Verhey FR, et al. Comparison of measurements of medial temporal lobe atrophy in the prediction of Alzheimer's Disease in subjects with MCI. Alzheimers Dement 2011;7:S133.