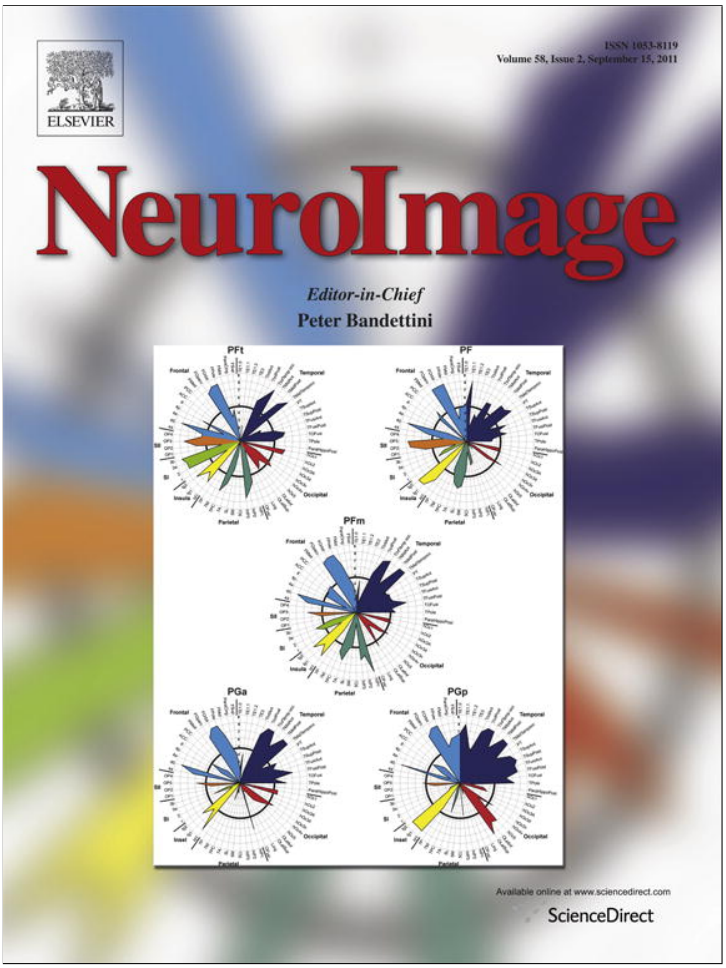

NeuroImage 58 (2011) 469–480

Contents lists available at ScienceDirect

NeuroImage

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

Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease☆

Andrea Chincarini [a,*], Paolo Bosco [a,b], Piero Calvini [a,b], Gianluca Gemme [a], Mario Esposito [a,b], Chiara Olivieri [c], Luca Rei [a,b], Sandro Squarcia [a,b], Guido Rodriguez [d], Roberto Bellotti [e,f], Piergiorgio Cerello [g], Ivan De Mitri [i,h], Alessandra Retico [j], Flavio Nobili [d] and The Alzheimer's Disease Neuroimaging Initiative

[a] Istituto Nazionale di Fisica Nucleare, Sezione di Genova, I-16146, Genova, Italy
[b] Dipartimento di Fisica, Università degli Studi di Genova, I-16146, Genova, Italy
[c] Dipartimento di Informatica e Scienze dell'Informazione, Università degli Studi di Genova, I-16146, Genova, Italy
[d] Neurofisiologia Clinica, Dipartimento di Neuroscienze, Oftalmologia e Genetica, Azienda Ospedale-Università S. Martino, Genova, I-16132, Genova, Italy
[e] Dipartimento Interateneo di Fisica "M. Merlin" and TIRES, Università degli Studi di Bari, Italy
[f] Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy
[g] Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Italy
[h] Dipartimento di Fisica, Università del Salento, Italy
[i] Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, Italy
[j] Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, I-56127 Pisa, Italy

ARTICLE INFO

Article history:
Received 20 January 2011
Revised 24 May 2011
Accepted 27 May 2011
Available online 16 June 2011

Keywords:
MRI
Image analysis
Alzheimer's disease
Medial temporal lobe
Hippocampus

ABSTRACT

Background: Medial temporal lobe (MTL) atrophy is one of the key biomarkers to detect early neurodegenerative changes in the course of Alzheimer's disease (AD). There is active research aimed at identifying automated methodologies able to extract accurate classification indexes from T1-weighted magnetic resonance images (MRI). Such indexes should be fit for identifying AD patients as early as possible.
Subjects: A reference group composed of 144 AD patients and 189 age-matched controls was used to train and test the procedure. It was then applied on a study group composed of 302 MCI subjects, 136 having progressed to clinically probable AD (MCI-converters) and 166 having remained stable or recovered to normal condition after a 24 month follow-up (MCI-non converters). All subjects came from the ADNI database.
Methods: We sampled the brain with 7 relatively small volumes, mainly centered on the MTL, and 2 control regions. These volumes were filtered to give intensity and textural MRI-based features. Each filtered region was analyzed with a Random Forest (RF) classifier to extract relevant features, which were subsequently processed with a Support Vector Machine (SVM) classifier. Once a prediction model was trained and tested on the reference group, it was used to compute a classification index (CI) on the MCI cohort and to assess its accuracy in predicting AD conversion in MCI patients. The performance of the classification based on the features extracted by the whole 9 volumes is compared with that derived from each single volume. All experiments were performed using a bootstrap sampling estimation, and classifier performance was cross-validated with a 20-fold paradigm.
Results: We identified a restricted set of image features correlated with the conversion to AD. It is shown that most information originate from a small subset of the total available features, and that it is enough to give a reliable assessment. We found multiple, highly localized image-based features which alone are responsible for the overall clinical diagnosis and prognosis. The classification index is able to discriminate Controls from AD with an Area Under Curve (AUC) = 0.97 (sensitivity ≃89% at specificity ≃94%) and Controls from MCI-converters with an AUC = 0.92 (sensitivity ≃89% at specificity ≃80%). MCI-converters are separated from MCI-non converters with AUC = 0.74(sensitivity ≃72% at specificity ≃65%).
Findings: The present automated MRI-based technique revealed a strong relationship between highly localized baseline-MRI features and the baseline clinical assessment. In addition, the classification index was

Abbreviations: AD, Alzheimer's disease; CTRL, Control subjects; MCI, Mild cognitive impairment; MRI, Magnetic Resonance Imaging; RF, Random Forest; SVM, Support Vector Machine; AUC, Area Under Curve; VOI, Volume of Interest.

☆ 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. A complete listing of ADNI investigators is available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf.

* Corresponding author at: INFN, via Dodecaneso 33, I-16146 Genova, Italy. Fax: +39 010 313358.
E-mail address: andrea.chincarini@ge.infn.it (A. Chincarini).

1053-8119/$ – see front matter
doi:10.1016/j.neuroimage.2011.05.083

also used to predict the probability of AD conversion within a time frame of two years. The definition of a single index combining local analysis of several regions can be useful to detect AD neurodegeneration in a typical MCI population.

Introduction

In recent years, the early clinical signs of Alzheimer's disease (AD) have been extensively investigated, leading to the concept of amnestic Mild Cognitive Impairment (aMCI), an intermediate cognitive state between normal aging and dementia (Winblad et al., 2004). The aMCI condition is currently identified by both a reported and an objective memory impairment, either associated with a slight impairment in other cognitive areas (multi-domain aMCI) or not (single-domain aMCI).

In longitudinal studies the aMCI subjects are experienced either to convert to AD (converters) or not (non-converters). In the latter case they may remain stable in the aMCI state or they may even revert to normalcy. Therefore, aMCI is a clinically and pathologically heterogeneous state in need of effective and reliable strategies to predict the clinical evolution. When these are available, hopefully upcoming disease-modifying drugs will be administered only to the aMCI subjects with prodromal AD as diagnosed with the help of specific biomarkers (Dubois et al., 2007, 2010).

Neuropsychology is the clinical cornerstone in the effort of performing a good and reliable predictive classification of the aMCI subjects. However such tests alone prove to be not completely satisfactory because of relatively low specificity between aMCI converters and non-converters and because of floor effects that sometimes make recall measures relatively insensitive to longitudinal changes.

The new proposed criteria for early diagnosis of AD suggest that one or more biomarkers should show typical findings ("supportive features") for an aMCI subject to be diagnosed as affected by early AD (Dubois et al., 2007).

The evidence of medial temporal lobe (MTL) atrophy in magnetic resonance imaging (MRI) is probably the most easily accessed worldwide.

Studies carried out in the last decade indicate that MRI can be used to quantify regional atrophy in MCI population, distinguishing early and later preclinical stages of AD (Pruessner et al., 2000; Shen et al., 2002).

A key contribution to biomarker findings came from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a large study 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. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

The noticeable growth in the number of candidate biomarkers poses the question on which one can add more value to the routinely performed episodic memory tests.

In principle, an ideal biomarker should detect a specific pathophysiological feature of AD, not present in the healthy condition, in other primary dementias, or in confounding conditions. Besides being reliable, a biomarker should be detectable by means of procedures which must be relatively non-invasive, simple to perform, widely available and not too expensive.

Among others, various morphological brain measures performed by means of MRI, ranging from brain-wide, voxel-wise multivariate measures to the selective volume estimate of restricted regions of interest, such as the hippocampal formation, have been proposed as candidate biomarkers (Frisoni et al., 2006; Karow et al., 2010; Desikan et al., 2009). In particular, regional neuro-anatomical changes have been investigated as biomarkers for Alzheimer's disease (Holland et al., 2009; Vemuri et al., 2010), and aMCI conversion (Risacher et al., 2010).

The increased interest in such "supportive features" derived from neuroimaging is due to the improved image quality and to the development of novel computer-assisted image processing tools giving the possibility of an automated volume and shape quantification of brain structures. Moreover, the interest is also raised by the availability of large sets of imaging data collected by large multicenter studies such as the ADNI.

In the context of neuroimaging applied to AD, computer-aided analysis techniques have been proposed to discover and study biomarkers via texture changes in signal intensity (Freeborough and Fox, 1998), gray matter concentrations differences, atrophy of subcortical limbic structures (Thompson et al., 2004; Frisoni et al., 2006) and general cortical atrophy (Thompson et al., 2003; Lerch et al., 2005). The underlying assumption being, in the general case, that changes in neuropsychological or neurological functions under consideration have a morphological counterpart, detectable via structural MRI.

A growing body of literature has used machine learning methods to extract high-dimensional features of interest from MRI, on which classification functions are built to assist in clinical diagnosis of probable AD or predict future clinical status for individuals with MCI (Klöppel et al., 2008; Fan et al., 2008; Lao, 2004; Davatzikos et al., 2008).

Hippocampal atrophy analysis performed by means of three-dimensional (3D) MRI seems to obtain a rather good discriminative value (Ferreira et al., in press). However, according to the traditional approach such volumetric measurements typically rely on manual/semi-automated outlining of the hippocampal structures on serial MR images, which is time consuming and prone to inter-rater and intra-rater variability.

In order to overcome these difficulties a novel approach was proposed in a previous paper (Calvini et al., 2009). It is based on a simple, quick, and operator independent method for the automatic extraction of two regions around the hippocampus (one for each side of the brain) from an MR image. From such subimages, denoted there as hippocampal boxes (HBs) and containing both the hippocampus and the perihippocampal region, a statistical indicator was able to separate the AD, aMCI and controls cohorts with good accuracy.

We propose, as an evolution of the previous technique, a procedure which is able to find other pathology-specific volumes of interest (VOIs) where a high discrepancy exists between healthy controls and either MCI converters or AD patients. Within a given VOI, we shall see that only a small volume contributes to the cohort discrimination, that is, the relevant information for the clinical assessment is highly localized. A Classification Index (hereafter designated as CI) could then be computed and found potentially more accurate than the discrimination based on the previous HBs.

Materials and methods

Our procedure is meant to be fully automated and consists of five steps summarized as follows: (i) image preprocessing (noise removal, affine registration, gray-level intensity normalization), (ii) multiple VOI extraction by means of template-matching and rigid registration, (iii) feature computation and Random Forest-based (RF) feature classification, (iv) Support Vector Machine (SVM) analysis and CI computation, (v) CI validation and conversion probability estimation based on follow-up clinical assessment.

Subjects

The raw data used in the preparation of this article were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database (http://www.loni.ucla.edu/ADNI/Data). Up to the date of this writing, the ADNI has recruited229 healthy elderly (CTRL), 398 MCI and 192 AD patients to participate and be followed for 2–3 years.

Statistical data of the subjects included in our analysis are summarized in Table 1. Subjects have been divided into 2 categories: a training/testing set and a trial set.

The training/testing set consisted of 333 age and sex-matched subjects, namely 189 CTRL and 144 AD. The trial set consisted of 302 MCI subjects, 136 of whom converted to AD in a time-frame of 2 years from the baseline scans. These subjects were selected from the larger ADNI data on the basis of having both baseline and at least 2 years information fully available. Moreover, training subjects were chosen if confirmed to be healthy controls/AD at follow-up assessment.

MRI acquisition and preprocessing

All images used here were acquired with 1.5 T scanners. Data were collected across a variety of scanners with protocols and processed. For up-to-date information on ADNI eligibility criteria and protocols see http://www.adni-info.org.

Raw NIFTI-converted MRI scans were downloaded from the ADNI site, automatically reviewed by signal-to-noise statistics for quality, and processed with a wavelet based noise-filtering algorithm to improve signal-to-noise ratio and image uniformity across different sites.

Wavelet based denoising methods for MRI are quite common in literature, as wavelet domain processing has many beneficial properties compared to classical image domain processing. Through the wavelet transform of the image we gain access to a sparse hierarchical representation where most of the information is contained in few high magnitude coefficients and the remaining low magnitude coefficients can be related to the noise present in the image.

Among the firsts to use wavelet techniques we can cite the work of Healy et al. (1992) and Hilton et al. (1996) for the case of functional MRI. Nowak (1999) was the first to address the problem of the bias introduced by the Rician noise model that affects MR images, and proposed a wavelet scheme that operates on the squared magnitude of the images. Pizurica et al. (2003) performed a preliminary coefficients classifications in order to empirically estimate the statistical distributions of the coefficients that represent useful image features and employed these probabilities in a Bayesian de-noising scheme.

For our work we propose a soft-threshold de-noising scheme in the Full Steerable Pyramid. The Steerable Pyramid is a particular wavelet transform first proposed by Castleman et al. (1998) and then extended by Portilla et al. (2003). The SteerablePyramid is a redundant transform; contrary to the Discrete Wavelet Transform (DWT) it is translation and rotation-invariant. This means that for each scale we can select as many orientations as we want, gaining a precise response to every structure contained in the image. It is thus particularly suited for image analysis and de-noising.

One drawback of the Steerable Pyramid is that the transform is inherently 2D (it is not straightforward nor computationally feasible to obtain a 3D extension). On the other hand using a 3D wavelet transform such as the 3D-DWT does not seem the best approach either, as we lose the fine orientation selectivity of the Steerable Pyramid. Three-dimensional wavelets are used in literature for compression (Schiavi et al., 2004) but it is the 2D approach which is the one most suitable for processing and analysis.

In our work we solve the problem by sequentially applying the de-noising technique for each slice and each direction, therefore a total of three noise-filtered MRIs are produced. The three filtered images are merged into a mean MRI to further improve the signal-to-noise ratio (see Fig. 1 for a schematic representation of the filtering procedure).

A novel method for filtering the noisy wavelet coefficients was devised for this work. The idea for the selection of an appropriate size for the soft-thresholding function is based on preserving the structures of the image. The noise filtering procedure uses an automatic calibration of the noise threshold N_t and three thresholds N_t^* are required to process a 3D image (one threshold for each direction).

To look for the appropriate N_t^*, we computed the Structural Similarity Index (SSI) between the raw image and the filtered one as function of N_t. The structure similarity index was first proposed by Wang et al. (2004) and it is commonly used as an image quality index. The computation is carried out separately on the three principal slices (axial, sagittal, coronal) passing through the image center of mass.

The SSI is well suited to account for visual detail preservation versus image information content, surpassing the classical image quality indexes, such as Mean Square Error based indexes. The SSI is defined as:

$$\mathrm{SSI}\left(g^k, f^k\right) = \frac{\left(2\mu_{g^k}\mu_{f^k} + c_1\right)\left(2\mathrm{cov}_{g^k f^k} + c_2\right)}{\left(\mu_{g^k}^2 + \mu_{f^k}^2 + c_1\right)\left(\sigma_{g^k}^2 + \sigma_{f^k}^2 + c_2\right)}$$

where g^k and f^k are local image patches taken from the same location of two images g and f. The local SSI index measures the similarities of three elements of the image patches: the local patch luminances, the local patch contrasts, and the local patch structures. μ_{g^k} and μ_{f^k} are the means of g^k and f^k, σ_{g^k} and σ_{f^k} are the standard deviations of g^k and f^k, cov_{g^k, f^k} is the covariance between g^k and f^k and c_1 and c_2 are user-defined constants. The resulting measure of similarity is finally given by the average of the local SSI (Wang and Bovik, 2009).

Since the SSI is a symmetric measure, it can be thought of as a similarity measure for comparing any two signals, and it is thus not limited to a usage for image quality assessment. In addition, with the proposed approach we are not making any assumption on the underlying statistics of the image noise, we are only looking at the preservation of the relevant structures, thus avoiding any approximate definition of the distribution of the noise and obtaining a fully automatic de-noising of our data.

The SSI is a monotonically decreasing function of the threshold value N_t. In order to get a good compromise between structure preservation and noise removal, the N_t^* value is chosen to be right

Table 1
Demographics and clinical findings at baseline. (⋆) training/testing set; (†) trial set. "APoE x/y" stands for Apolipoprotein E $-\epsilon(y)$ on the allele x; subject number is within parenthesis.

Cohort	Sample size	Age [y]	M/F	MMSE	APoE				
					2/2	2/3	3/3	3/4	4/4
CTRL ⋆	189	76.6 ± 5.1	95/94	29.1 ± 0.9	0.5% (1)	11.1% (21)	60.0% (114)	24.8% (47)	1.5% (3)
MCI-NC †	166	75.7 ± 7.3	106/60	27.2 ± 2.4	0.0% (0)	6.0% (10)	46.0% (77)	35.5% (59)	10.8% (18)
MCI-C †	136	75.1 ± 7.1	80/56	25.2 ± 2.7	0.0% (0)	2.2% (3)	31.6% (43)	45.6% (62)	17.6% (24)
AD ⋆	144	75.5 ± 7.5	78/66	22.3 ± 3.3	0.0% (0)	1.4% (2)	28.4% (41)	45.1% (65)	22.2% (32)

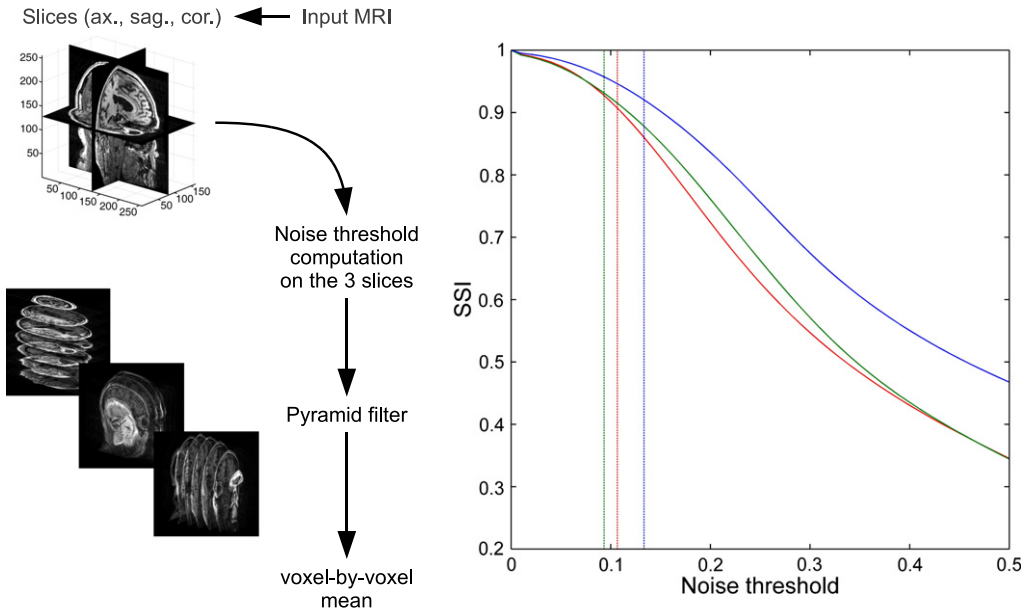

Fig. 1. Schematic representation of the noise-filtering process. The graph on the upper right corner shows the typical trend of the SSI vs the noise threshold N_t, expressed as percent of the maximum image intensity. The dash-dotted lines show the positions of the three thresholds N_t^* computed on the axial, sagittal and coronal slice.

after the "knee" of the function, defined as the intermediate value between the zero of the first derivative and the inflection point.

Automatic noise thresholding is targeted at improving MRI scan uniformity throughout the subjects, yielding better analysis stability with respect to images coming from different scanners. Fig. 2 shows a sample sagittal cross-sectioncontaining the right hippocampus as the original image (left) and after the application of the de-noise filter (right).

Registration and intensity normalization

De-noised scans were then registered onto the Montreal Neurological Institute (MNI) reference with a 12 parameter affine registration and resampled onto a 1 mm^3 isotropic grid (Mazziotta, 1995).

In order to ensure an accurate registration, each image underwent a 2-fold affine registration process, one with the normalized correlation and one with the normalized mutual information metric. The two-fold registration is a split process, where each image undergoes two registration processes in parallel, each one having its own metric and optimization parameters. The reason for using a split registration process is that of circumventing possible local minima in the parameter space. Ultimately, the double registration proves more reliable, as each single method failed to provide accurate registration on some images. To decide which image to keep after the split registration process, they were compared to the MNI reference using the normalized correlation metric and the best one was kept as final registered image.

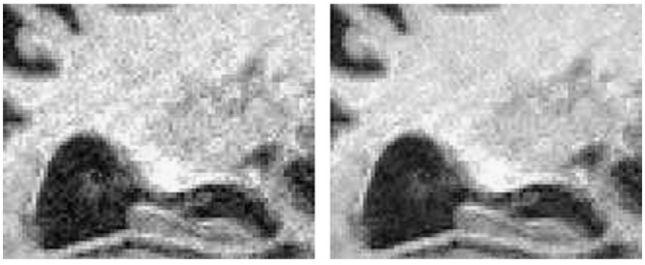

Fig. 2. Sample of a sagittal slice including the right hippocampus: original image (left), noise-filtered image (right).

The registered images were then intensity normalized by means of a segmen-tation-based method. Our method relies on a region-of-interest-based intensity segmentation. The chosen ROI, identified on the reference image, measures $n = 50 \times 120 \times 50 = 300{,}000$ voxels and is oriented along the long axis of the corpus callosum (see Fig. 3). This region proves to be relatively easy to classify into Cerebro-Spinal Fluid/Gray Matter/White Matter (CSF/GM/WM) and—thanks to the morphological characteristics of the included anatomical structures—it can be easily registered onto any target image.

The initial ROI, set as reference, is manually outlined on the MNI image and is subsequently registered onto each target image by means of a rigid transformation. The registration outputs the corresponding

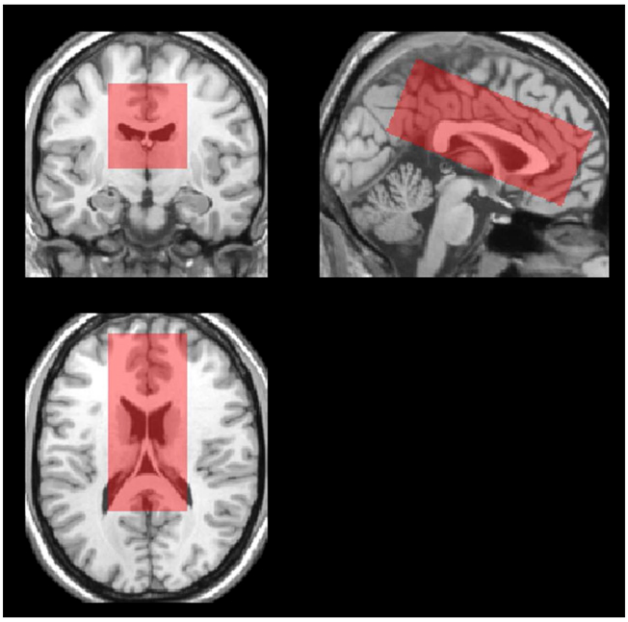

Fig. 3. Intensity normalization ROI size and positioning displayed on the MNI reference image (ventricula and corpus callosum). Similar ROIs are extracted on each image to serve as basis for the histogram matching procedure, following the segmentation into CSF/GM/WM.

ROI in the target image, which is thereafter segmented into CSF/GM/WM with k-means cluster analysis (Seber, 1984).

The three cluster means are matched to the corresponding CSF/GM/WM mean levels on the reference image. These are the fixed points through which a cubic spline is fitted. This non-linear intensity normalization pairs the three mean cluster intensities in the ROI (CSF/GM/WM) between each subject and the MNI reference image, and extends the mapping to the other gray levels by a smooth piecewise polynomial curve. Fig. 4 shows a schematic representation of the intensity mapping.

With this procedure we achieve good histogram normalization among images coming from different scanners. We also ensure that the mean gray levels of the three main cerebral matter contributions are mapped onto those of the MNI reference.

VOI extraction

The objective of this step is to extract selected VOIs on which salient image-based features can be computed. This step is accomplished by template matching techniques and rigid registration.

Each VOI is a parallelepiped-shaped volume with specific dimensions and it is associated with a set of atlas-like templates. Briefly, templates represent classes of similar structures, and they are computed once and for all from a set of matched aged Controls, MCI and AD patients. We have taken the VOI templates as a given atlas, as they were extracted in previous works (Calvini et al., 2009). Their use came into play only as a means to ensure correct registration of the various anatomical regions among subjects.

Once a set of templates is available for a given VOI, we used rigid registration to map them onto the target MRI. A correlation coefficient is calculated between each template and the tentative VOI extracted by registering that template onto the target image. The tentative VOI with the best normalized correlation value to its template is returned.

Nine VOIs are extracted from any given MRI (see Fig. 5). They include or capture part of anatomical structures, as listed in Table 2. Due to a minor computational setback, the right counterpart of VOI n. 7 (insula-sup. temporal gyrus) could not be consistently extracted from all MRIs. For this reason it has not been included in this work.

Although there is virtually no limitation to how finely the brain parcelization with VOIs can be made, we selected some candidate volumes relevant to AD-related pathology, based on previous MRI and FDG-PET findings (Mosconi et al., 2007; Fennema-Notestine et al., 2009; Walhovd et al., 2010b; Liu et al., 2010).

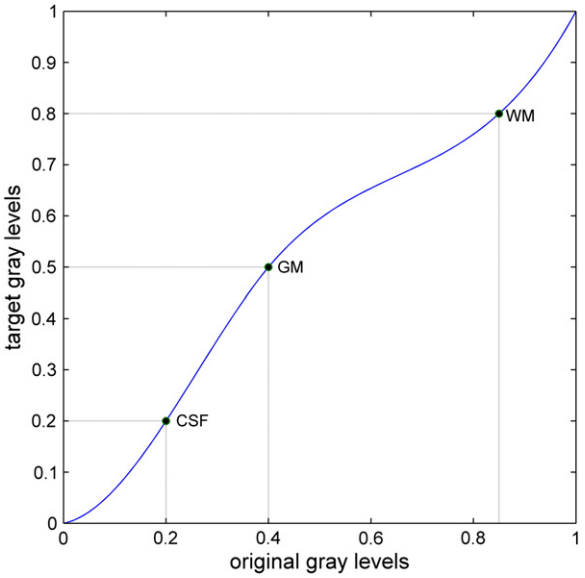

Fig. 4. Sample cubic piecewise polynomial curve mapping the ROI original gray level intensities. The three points are the mean levels of segmented CSF/GM/WM from the ROI. Target CSF/GM/WM levels are fixed while original ones are computed on the ROI extracted from the image.

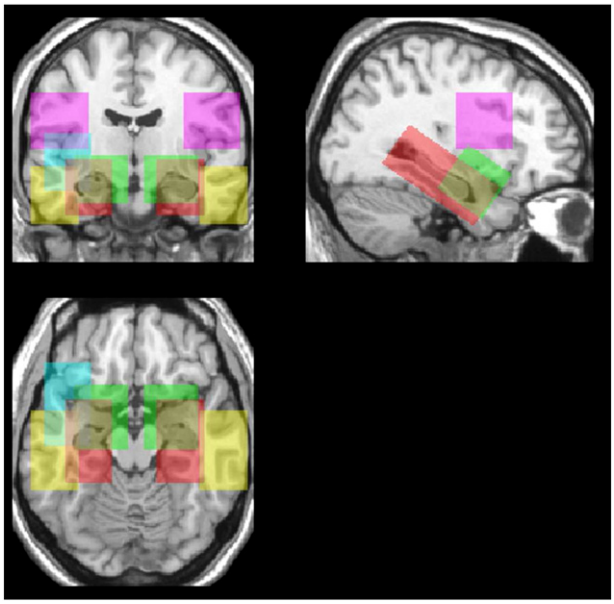

Fig. 5. VOI size and positioning displayed on the MNI reference image. The main structures captured by each VOI are listed in Table 2. VOI n. 1,2: red; VOI n. 3,4: green; VOI n. 5,6: yellow; VOI n. 7: cyan; VOI n. 8,9: magenta.

Seven VOIs (Table 2[a]) were chosen to include those temporal lobe structures that are known to be affected in early AD, such as the entorhinal, perirhinal cortex, hippocampus and parahippocampal gyri, irrespective of normal inter- and intra-individual variability. Two additional VOIs (Table 2[b]) were chosen as control volumes in regions known to be relatively spared in early AD.

However, VOI positions other than those considered in this work could be explored by means of the same approach, to include regions showing early atrophic changes such as, for instance, the thalami and the parietal cortex.

Global and local alignment optimizations between reference and subject are achieved via affine and rigid transformation, on the whole MRI and on the VOI respectively. The two-step global registration process has been shown in other studies to reduce positional variability of the VOI (Duchesne, 2006), which in turn lowers the chances of propagating as unwanted noise in the morphometric modeling. We are avoiding nonlinear registrations here on purpose, as we shall show they are not needed to achieve high accuracy in cohorts discrimination.

Table 2
Main gray matter structures captured in the VOIs (see Fig. 5). [a] potentially significant regions; [b] control regions.

	VOI n.	Main anatomical structures
[a]	1	Hippocampus, enthorinal cortex (right)
	2	Hippocampus, enthorinal cortex (left)
	3	Amigdala (right)
	4	Amigdala (left)
	5	Middle and inf. temp. gyrus (right)
	6	Middle and inf. temp. gyrus (left)
	7	Insula-Sup. temporal gyrus (left)
[b]	8	Rolandic (right)
	9	Rolandic (left)

Features and classification

For each subject, a total of $Z=9$ VOIs are extracted and each is filtered with $f=18$ different filters. Choices in image filtering are unlimited in principle but practical computational considerations suggested some restrictions. We therefore opted for a small set chosen to enhance both intensity and texture-based characteristics: Gaussian mean, standard deviation, range, entropy and Mexican-hat filters calculated on different voxel neighborhoods (see Table 3). The result of the application of the chosen filters to a sample image (a sagittal cross-section of the right hippocampus) is shown in Fig. 6.

Each VOI z is filtered according to Table 3 leading to a total of

$$F = f \times \sum_{z=1}^{Z} V_z$$

features, where V_z is the number of voxels in VOI z. The feature set consists therefore of the ensemble of all voxels of the filtered VOIs extracted from the MRI under analysis. The number of features for each MRI easily reaches the order of 10^6 and it is reasonable that not all of them be relevant in the CTRL/AD classification.

Random Forest algorithms were demonstrated to easily deal with high numbers of input features and it was also shown they improve on accuracy in comparison to other supervised learning methods (Breiman, 2001; Svetnik et al., 2003). Aside from improved accuracy, there are two extremely useful byproducts: out-of-bag estimates of generalization error and variable importance measures (Breiman, 2001; Bylander, 2002). It has been shown that the importance information can be used to select the most probable predictors (Archer and Kimes, 2008).

During training and for each feature, the RF classifier returns a number representing the relative importance of the feature for the classification. This "important features map" (IFM) weights those voxels within a VOI (and for each filter f), which are most relevant to the CTRL/AD cohort separation.

According to Archer and Kimes (2008), the RF algorithm together with the important variable information is most useful when there is a large number of highly collinear covariates, such as the case where voxels represent morphological structures.

The IFM can be used to prune the less relevant features, cutting the original number by approximately two orders of magnitude. Pruning was performed by thresholding the IFM, cutting out the lowermost 95%. However, this value is not a critical parameter. CTRL/AD classification performance was plotted against the threshold to find a reasonable trade-off. We checked that lower pruning thresholds did not improve on the classifier cross-validation results, whereas higher values (e.g. 98% and up) impoverished the feature set of too much information and therefore they did impact on the result quality. We opted for a high enough threshold to reduce the features to a more manageable number without impacting on the classification performance.

This feature pruning significantly reduced the computational time and resources necessary to train and cross-validate our classifiers. Once our input set was pruned from all but the most relevant features we fed it to a set of SVM classifiers (Cortes and Vapnik, 1995) and their outcome was averaged out to give the Classification Index.

Table 3
List of filters used to enhance both intensity and texture characteristics of the VOIs.

Filter type	Neighborhood (voxels)				
	3×3×3	5×5×5	7×7×7	9×9×9	11×11×11
Null	–	–	–	–	–
Gaussian mean	x	x	x		
Average	x	x	x		
Entropy	x	x	x		
Range	x	x	x		
Std. deviation	x	x	x		
Mexican hat				x	x

Statistical and computational considerations

In order to assess the statistical significance for the VOI extraction, feature selection and classification, we resorted to the bootstrap procedure and the k-fold cross-validation (Zhu et al., 2006). Fig. 7 displays a schematic overview of the training and testing procedure.

Bootstrapping involves the generation of multiple versions of the training group, serving to ensure maximum learning efficiency from a limited dataset and involves the generation of several random samples with replacement. In our procedure it was used to train many RF classifiers in order to get the IFM.

The chained RF and SVM classifiers were validated on the CTRL/AD set with a 20-fold, cross-validation paradigm. As shown in Fig. 7, $m \simeq N/20$ subjects are left out from the N CTRL + AD subjects and the training process is carried on the remaining N-m. MCI population was used neither in the training nor in the testing process.

Cohort classification was performed with an SVM algorithm, whose input is a subset of the feature space computed on all the VOIs (pruned features). As stated in Section 2.5, this subset is obtained by masking the whole feature set F with a thresholded IFM to reduce the total number of features per subject. The resulting number of features $\tilde{F}$ is the same for each subject and it is determined by the threshold applied to the IFM.

Still, the number of the pruned features $\tilde{F}$ is generally much greater than the number of training subjects. In order to minimize over-training effects, several SVM classifiers (n_{svm}) are trained on a random sample taken from the pruned feature space $\tilde{F}$.

$$n_{\mathrm{svm}} = 1.2 \times \frac{2\tilde{F}}{N-m}$$

Being N-m the number of subjects in the training set, each SVM classifier takes $\frac{N-m}{2}$ features in input. In order to take most of the features into account, n_{svm} was chosen to cover 1.2 times the number of the pruned feature set $\tilde{F}$. The CI value is computed by averaging the dichotomic outcome of the single SVM classifier on the n_{svm} set.

Although the SVM algorithm is capable of dealing with a large number of features, it is known to perform more reliably when the most irrelevant inputs are purged out (Guyon et al., 2002). In this way, the preliminary step with the RF can improve the classification stability.

Volumes Of Interest have been centered onto specific anatomical regions and their sizes are chosen to assure accurate registration, therefore some VOIs may share common volume. Volume redundancy may seem relevant among some VOIs (like, for instance, between the hippocampus and the amigdala region) but the actual analysis is carried out on the IFM pruned features and not on the whole VOI volume. On the IFM map, we measured a volume overlap of about 10%.

Even though some redundancy is still present in the IFM from neighboring VOIs, SVM classification is not impaired because it only means that a small percentage of features are more likely to be picked by the random sampling.

Clinical data on all cohorts include follow-up classification. We are interested in the MCI cohort follow-up which is usually taken every 6 months up to 2 years after the baseline acquisition. Regarding MRI data, we analyzed baseline scans only, as we aim to determine the conversion-to-AD probability based on the initial scan analysis.

Image processing from raw data to VOI extraction was completed using a combination of ITK tools (Yoo et al., 2002), FLIRT (Smith et al., 2004) and MATLAB (http://www.mathworks.com) in the framework of the LONI pipeline environment (Rex et al., 2003). RF and SVM classifiers, model prediction and the whole statistical analysis were performed using MATLAB.

The computational framework used for processing, storing and analyzing all images is based on a 48 cores Sun Grid Engine cluster running a LONI pipeline server. The whole process, from raw data to the CI, takes about 15 minutes per image and is fully automated.

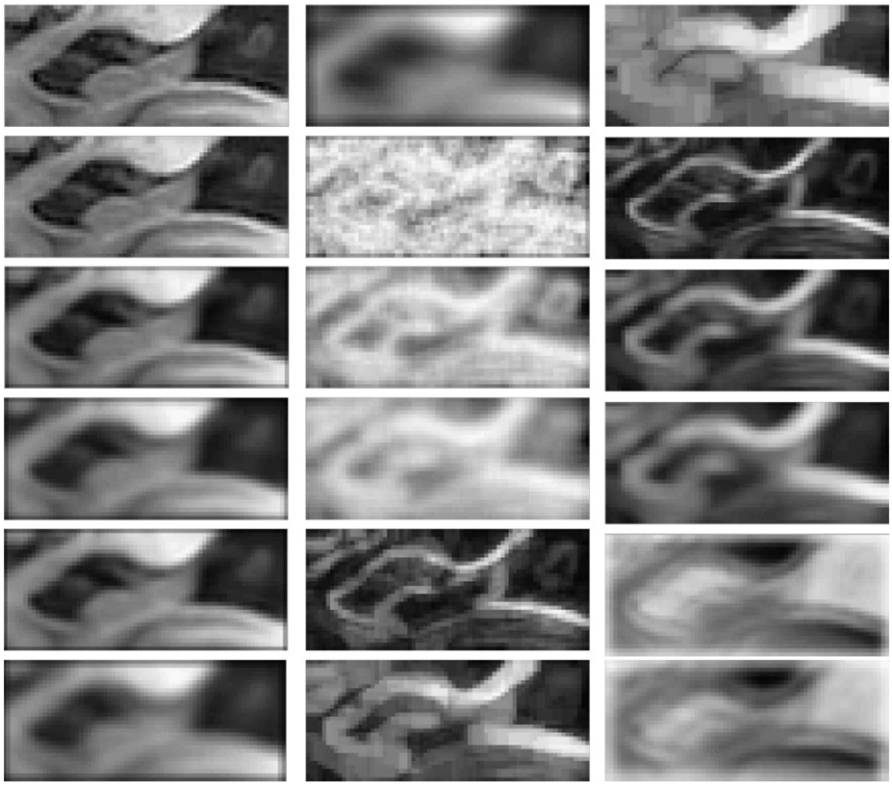

Fig. 6. Sample VOI section (right hippocampus of an AD subject, sagittal cross-section) after applying intensity and texture filters. Left column, from top: original image, Gaussian $3\times3\times3$, Gaussian $5\times5\times5$, Gaussian $7\times7\times7$, average $3\times3\times3$, average $5\times5\times5$. Central column, from top: average $7\times7\times7$, entropy $3\times3\times3$, entropy $5\times5\times5$, entropy $7\times7\times7$, range $3\times3\times3$, range $5\times5\times5$. Right column, from top: range $7\times7\times7$, std. deviation $3\times3\times3$, std. deviation $5\times5\times5$, std. deviation $7\times7\times7$, Mexican hat $9\times9\times9$, Mexican hat $11\times11\times11$.

Our implementation includes several checkpoints (on the raw MRI, VOI extraction and VOI content) to ensure data quality throughout the process. These checkpoints verify incoming data and filter consistency, so that MRI images and/or VOIs which do not meet satisfactory quality requirements are automatically discarded. Our checks found about 3% of the initial raw dataset to be affected either by insufficient data quality or image registration problems. Numbers in Table 1 describe the dataset once the poor quality images were dropped out.

Results

The first result comes directly from the RF classifier. For visualization purposes, a smoothed, thresholded IFM overlaid on a representative structural MRI scan is shown in Fig. 8. The plotted IFM is the superposition of all thresholded IFMs computed on the 9 VOIs. It shows that the relevant information is squeezed into some decidedly small areas within the VOIs.

Relevant locations mostly match those found in literature (Hinrichs et al., 2009; McEvoy et al., 2009; Misra et al., 2009; Morra et al., 2009; Pelaez-Coca et al., 2011; Hinrichs et al., 2011; Li et al., in press; Plant et al., 2010) and they are rather precisely pinpointed. This characteristic is due to the lack of smoothing in all our pre-processing, VOI extraction and region analysis, whereas most works in literature rely on deformable registration and Gaussian convolution. See Table 4 for a detailed list of the main relevant regions.

Classification

Once the classifier was trained and tested on the CRTL/AD cohorts, we applied it to the MCI population. The CTRL population is first compared with the mild AD subjects and then with the prodromal AD subjects (i.e. the MCI-C).

The averaged output from the SVM classifiers is taken as Classification Index (CI) and plotted. Its values range from 1 (Normalcy) to −1 (AD). Fig. 9 summarizes its performance on the training/testing set (in cross-validation) and on the trial set.

The area under the ROC curve (AUC) is ≃0.97 for the CTRL vs. AD, with a sensitivity ≃ 89% at specificity ≃94%; and AUC ≃0.92 for the CTRL vs. MCI-C, with a sensitivity ≃89% at specificity ≃80%. Finally, the MCI-NC/MCI-C are compared and their AUC is found to be ≃0.74, with a sensitivity ≃72% and specificity ≃65%.

It is also noteworthy to see how well the single VOI performs in comparison to the combined volume analysis. We trained and tested new classifiers separately on each single VOI (in a similar fashion as described in Section 2) and looked at their cross-validation performance, as well as their discrimination ability on the trial set.

Table 5 summarizes these findings and compares the performance on the individual VOIs to the whole analysis.

Clearly some VOIs exhibit better than average results but no single VOI consistently matches (or exceeds) the performance of the combined analysis on both training/testing and trial set.

As expected, the rolandic areas in the 2 control VOIs (see Tables 2 [b] and 5) yield a weaker discrimination than the other VOIs, particularly on the trial set. The only discrimination value within these two VOIs comes from temporal lobe structures included at their borders. Fig. 10 shows a section of the 2 control VOIs and the related IFM. It is apparent that the IFM is mostly clustered in areas at the VOIs border, where it is adjacent to the temporal lobe structures. The general brain atrophy extension and magnitude in AD subjects makes the information coming from that IFM still relevant for the CTRL/AD discrimination but it is not sufficient to tell MCI-C from MCI-NC.

We checked whether genotype distribution on misclassified MCI subjects differed significantly from the original population. A χ^2 test found both misclassified MCI-NC and MCI-C to be not significantly

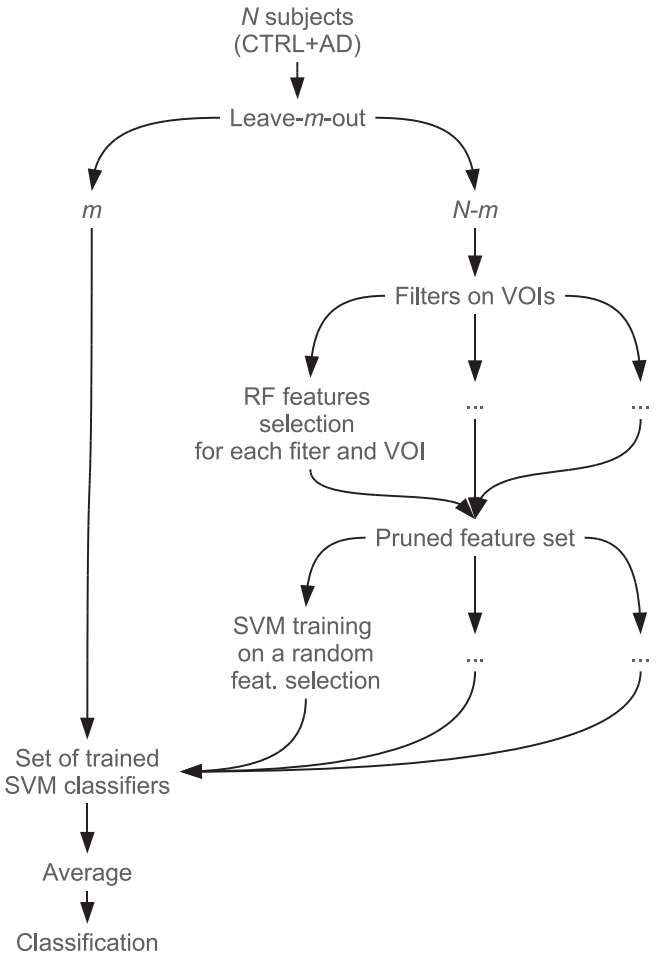

Fig. 7. Training and testing flowchart. Schematic representation of the steps to get the Classification Index on the CTRL + AD set.

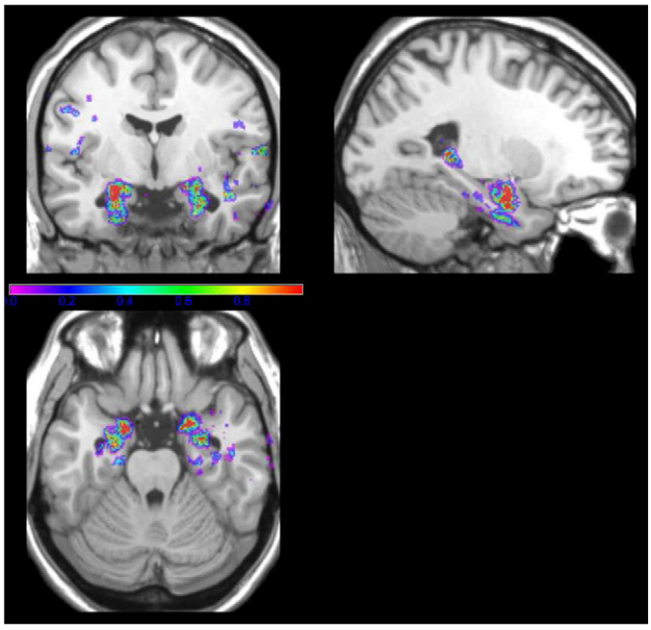

Fig. 8. Important features map (IFM) sections superimposed on the MNI reference image. Color scale is proportional to the normalized IFM value.

Table 4
Main anatomical structures captured by the thresholded IFM, Brodmann Areas are within parentheses. These locations are the only ones involved in the computing of the classification index. A graphical representation is given in Fig. 8.

VOI n.	Relevant locations
1	Superior temporal gyrus (BA 38), para-hippocampal gyrus, right uncus hippocampi
2	Left precentral gyrus (Frontal Lobe, BA 6), left superior temporal gyrus (BA 38), Left parahippocampal gyrus (BA 28), left uncus hippocampi (BA 20)
3	Right superior temporal gyrus (BA 38)
4	Left superior temporal gyrus (BA 38)
5	Right middle temporal gyrus (BA 21), right inferior temporal gyrus (BA 20)
6	Left middle temporal gyrus (BA 21)
7	Left superior temporal gyrus (BA 38)
8	Right superior temporal gyrus (BA 22), right superior temporal gyrus (BA 38)
9	Left superior temporal gyrus (BA 38)

different from the original MCI-NC/MCI-C population ($p = 0.82$ and $p = 0.11$ respectively).

In addition, we tested whether CI classification on MCI subjects performs significantly better than a random classifier with a McNemar's χ^2 test. Setting the significance level to $\alpha = 0.05$ we found $\chi^2 \simeq 7.62$ ($p = 0.003$).

We also checked how well our VOI-based analysis correlates with the MMSE score and we found the Pearson's correlation to be $r = 0.31$ ($p < 10^{-3}$) on the trial set.

Discussion and conclusions

This study proposes a computational neuro-anatomic method to quantify local patterns of brain atrophy in a large sample of cognitively normal individuals and in patients with prodromal or mild AD. The general approach presented here was validated on a trial set of a heterogeneous MCI population, and its results in terms of cohort discrimination are comparable to those found in recent works, where applied methodologies mostly concerned whole-brain pattern analysis (Karas et al., 2008; Davatzikos et al., in press; Wang et al., 2010; Fan et al., 2008). Similarly to these studies, we give individual patient classification in addition to the group analysis.

Unlike previous studies our approach is based on local analysis, where the interesting volumes encompass only a small fraction of the whole brain. In addition, we approached analysis using affine and rigid registration only, although it relies on regional atlases to achieve high spatial accuracy among subjects.

It is still unclear whether a whole-brain analysis delivers better insight on MCI to AD conversion, although some studies advocate the whole-brain approach (Misra et al., 2009; Cuingnet et al., 2011).

As shown in Table 5, some VOIs (n. 1, 4 and 5) perform nearly as well as the whole analysis on the training set in cross-validation, but they come short on the trial set. This could suggest that the diversity in the MCI population cannot be fully taken into account with a single region and a more sophisticated approach is called for.

On the other hand, CTRL/AD discrimination seems to be easier to attain even with a single VOI analysis, as a more advanced AD state leaves unequivocal marks in more than a few regions. As expected, CTRL and AD are very well differentiated whereas the MCI cohort is confirmed to be more heterogeneous even in terms of CI.

The complexity of the atrophy pattern paired to the histopathological studies of β-amyloid plaques and neurofibrillary tangles deposition (Braak et al., 1996), indicates a widespread distribution of brain neuro-degeneration during the progression of AD. This implies that to examine volumes of a small number of structures (typically the hippocampus and the entorhinal cortex) may not be sophisticated enough for separating the CTRL, MCI and AD (let alone

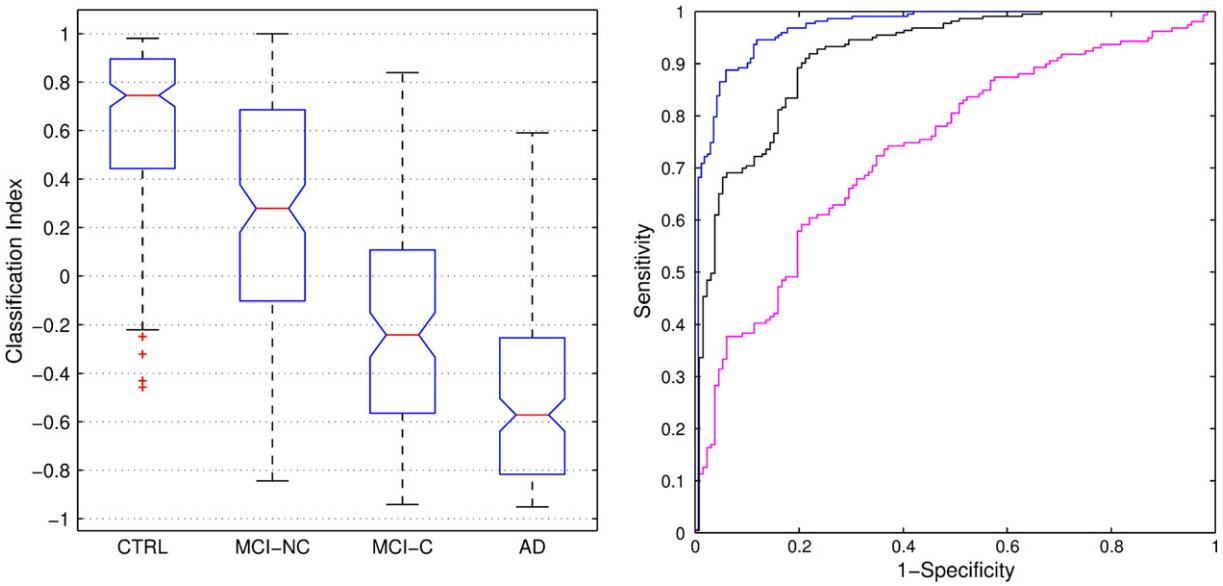

Fig. 9. Boxplot representation and ROC curve of the classification index (CI) for the CTRL/AD, CTRL/MCI-C and the MCI-NC/MCI-C cohorts. CI values on the training set are computed in cross-validation (20-fold). Boxplot notches represent the 95% confidence level for the median value (red horizontal line). The Area under the ROC curve for the CTRL/AD is 0.97 (blue line) with a sensitivity ≃89% and specificity ≃94%. The performance on the prodromal-AD (CTRL/MCI-C) is AUC = 0.92 (black line) with a sensitivity ≃89% and specificity ≃80%. The performance on the trial set (MCI-NC/MCI-C) gives an AUC = 0.74 (magenta line) with a sensitivity ≃72% and specificity ≃65%.

MCI-C and MCI-NC) with clinically adequate sensitivity and specificity, as found in neuroimaging literature (Dickerson, 2001; Walhovd et al., 2010a; Dixon et al., 2002; Liu et al., 2010).

Similarly to Davatzikos et al. (in press) a classification index checked against clinical assessment and prognosis is derived. In their work they obtain an MCI-NC/MCI-C performance of AUC = 0.66 on an ADNI MCI population very similar to the oneused in this work; the performance improves (AUC = 0.734) on a subset of 120 MCI patients.

Cuingnet et al. (2011) compared several different methods for the MCI-C/MCI-NC separation on a dataset from the ADNI database and confirmed that the MCI-C/MCI-NC discrimination is a very difficult task, even with rather different analysis approaches: only four methods proved to be slightly more predictive than a random pick. Two of them, described in Vemuri et al. (2008) and Fan et al. (2007), make use of information from the whole brain volume; other approaches, discussed in Klöppel et al. (2008) and Chupin et al. (2009), are basedon gray matter intensity-based classification and from hippocampus related information, respectively. The performance is provided in terms of a working point identified by the sensitivity and specificity values and the McNemar test is used to assess the reliability of the methods with respect to a random choice.

Table 5
Classifier performance on the training set (results in cross-validation) and the testing set when only the information from a single VOI is taken into consideration. Numbers differ among VOIs, and none consistently matches the whole 9-VOI analysis (first line) on the whole dataset. Results for the CTRL/AD cohorts come from the *k*-fold cross-validation; MCI subjects were never used for training.

VOI n.	Main structure	AUC		
		CTRL/AD	CTRL/MCI-C	MCI-NC/MCI-C
All	–	0.97	0.92	0.74
1	Hippocampus (r)	0.93	0.91	0.68
2	Hippocampus (l)	0.92	0.88	0.67
3	Amigdala (r)	0.91	0.88	0.66
4	Amigdala (l)	0.93	0.91	0.69
5	Mid. inf. temp. gyr. (r)	0.96	0.94	0.71
6	Mid. inf. temp. gyr. (l)	0.91	0.83	0.65
7	Insula (l)	0.89	0.86	0.69
8	Rolandic (r)	0.81	0.68	0.59
9	Rolandic (l)	0.82	0.73	0.57

None of the methods is fully satisfactory, suggesting that the complexity of the disease and the inter-individual differences are hardly taken into account with biomarkers coming form a single technique. However, the results may also suggest that many of the MCI-NC are likely to become MCI-C in the near future and that some are progressing to dementia due to causes other than AD.

Another recent work by Costafreda et al. (2011) developed an automatic hippocampal shape analysis procedure for the prediction of MCI-C, and compares MCI-NC/MCI-C prediction abilities among several other methods, from fully manual to fully automated. Their method scores a 77% sensitivity with an accuracy of 80% calculated on a follow-up of one year.

In addition, they compare their method to several other studies in literature. Follow-up time, sample size and criteria differs among studies and results are not directly comparable to our work, for which the MCI-NC/MCI-C accuracy is ≃68%. Given our sample size of 302 MCI and follow-up time of 24 months and the fully automatic

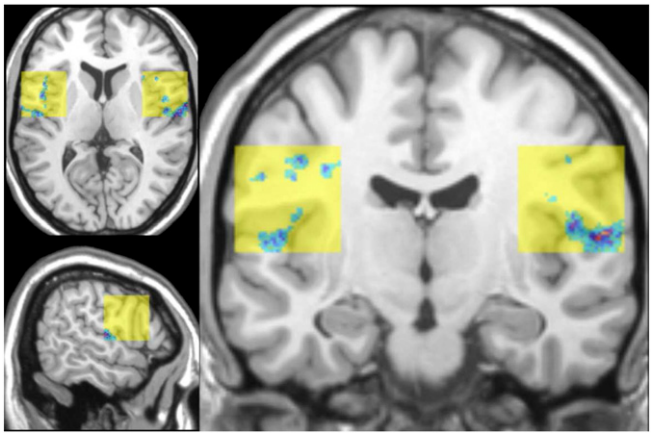

Fig. 10. Rolandic VOIs section (yellow) and the related IFM (in color). The IFM position within the VOIs shows that the relevant information mostly comes from areas bordering on the temporal lobe.

characteristic of our analysis, we believe our results to be competitive (McEvoy et al., 2009; Plant et al., 2010; Ferrarini et al., 2009).

Longitudinal studies have shown that although the majority of MCI subjects converting to AD dementia does so in the first two years, a non negligible part of them may convert in the following years (Bennett et al., 2002). Thus, a 2-year follow-up period may be insufficient to ensure adequate clinical separation between MCI-C and MCI-NC. This is probably the main reason to explain the lower accuracy achieved in the MCI-C vs. MCI-NC comparison, with respect to the MCI-C vs. CTRL comparison.

The CTRL/MCI-C curve shows a good performance in discriminating prodromal AD and indicates that CI scores computed on baseline scans already suggest the trend to AD conversion. Due to the heterogeneous composition of the MCI cohort, the CI score on MCI-NC vs MCI-C comes out less accurate although it still exhibits a fairly marked separation between converters/non-converters.

The issue of non-converter MCI patients cannot be solved without long-term follow-up but we know that MRI changes can be already detected at least three years before the diagnosis of AD (Whitwell et al., 2007). Moreover, local cortical thinning in temporal lobes has been reported in asymptomatic subjects with high brain amyloid load (Dickerson et al., 2009). This evidence, together with those coming from the other main biomarkers (such as FDG-PET and cerebrospinal fluid assays), suggests that signs of neurodegeneration may be detected several years (i.e., 5–10) before the onset of objective cognitive signs or symptoms (Clifford R Jack et al., 2010). Thus, we may suppose the non-converter MCI population with a limited follow-up time (i.e., 2–3 years) is highly heterogeneous, containing patients who will never convert as well as 'late' converters to AD or other dementias. These 'late' converters are actually in an early pre-Alzheimer state, which may have reduced the accuracy of MRI in distinguishing between MCI 'rapid' converters and non-converters.

The general strategy found in the most recent studies is to combine biomarkers from different sources, typically structural, functional and biochemical. Indeed, our results also demonstrated that sole measurements of the hippocampus and the entorhinal cortex are not sufficient to tell MCI-converters from the MCI-non converters efficiently, particularly when the follow-up time is limited, while they seemaccurate enough for the discrimination between CTRL and AD groups (see Table 5).

Further methodological considerations

The combination of bootstrapped RF, SVM and feature thresholding methods is not new in literature (Ebina et al., 2011; Larios et al., 2010; Saeys et al., 2008; Singh et al., 2008; Waske and Van Der Linden, 2008; Waske et al., 2010).

More specifically, Ebina et al. (2011) used an SVM trained on 25 optimal features, which were selected from a set of 3000 features using a RF algorithm complemented with a stepwise feature selection. Their results on prediction sensitivity using RF + SVM were over 19.9% higher than those of control SVM predictors trained with non-optimized features, strongly suggesting the efficiency of feature selection method. Larios et al. (2010) applied RF selection of Haar-based features to a SVM with a non-linear kernel, combining efficient low-level feature evaluation with discriminative learning for pattern recognition and image classification.

The importance of feature selection was checked against direct SVM classification without the IFM thresholding step, that is direct SVM classification of the whole feature space without any prior feature selection via RF. Due to the high number of features involved in the 9 VOI analysis, we compared the classification performances on VOI n.1 only (the VOI around the right hippocampus). The training/testing procedure followed the same path as depicted in Fig. 7, taking only one VOI into consideration and without the RF step.

Results indicate that direct SVM classification performed slightly worse. On the training/testing set in cross-validation (CTRL + AD) we got $AUC_{SVMdirect} = 0.87$ vs. $AUC_{RF+SVM} = 0.93$, and the performance on the trial set (MCI-NC + MCI-C) was $AUC_{SVMdirect} = 0.66$ vs. $AUC_{RF+SVM} = 0.68$.

As classification is computed by averaging the output of several SVM classifiers—each trained on a subset of randomly chosen features—it is reasonable that some SVM results be impaired when including confounding features in their training. Therefore a smaller *AUC* value is to be expected because in the whole unpruned set, many features may be confounding or only marginally relevant for the CRTL/AD discrimination.

On the same VOI (right hippocampus) we also tested how well would a single SVM classifier perform (still without any prior RF feature selection). A single SVM was trained on CTRL + AD and tested on MCI-NC/MCI-C cohorts. Results were disappointing, yielding a mere $AUC_{singleSVM} = 0.52$ between MCI-NC and MCI-C.

We believe the poor performance is due to the high number of features (even when considering one VOI only) compared to the rather small number of training/test data. In this condition, the cardinality of the data set does not allow a correct evaluation on the large number of free parameters of the classifier, and that exposes it to overtraining effects.

We explored also another approach, that is using a RF step only without any SVM. In this approach a class of boot-strapped RF classifiers are used to directly provide the classification and not the feature selection.

As playing ground, we still tested the performance of this RF-only approach on VOI n.1 and the results should be compared with the second row of Table 5. This test required some optimization to tune the number of RF instances and the number of features to consider. Results show that a set of RF classifiers perform already rather well, yielding an $AUC_{RFonly} = 0.90$ on the CTRL/AD set in cross-validation and $AUC_{RFonly} = 0.64$ on the trial set (MCI-NC + MCI-C).

We then tested the relevance of image filters, which we used to enhance intensity and texture characteristics and to provide a larger feature space. According to the recent work of Cuingnet et al. (2011) the voxel-DIRECT approaches were competitive with other methods including cortical thickness measures and hippocampal volume and shape. As baseline comparison, the CI performance was checked against the application of filters to VOIs, that is with a feature set computed with the null filter only (first line of Table 3).

Analysis was performed on the 9 VOIs and the $AUC_{null\text{-}filter}$ compares directly to the numbers given on the first line of Table 5 ("All"). We had $AUC_{null\text{-}filter} = 0.95$ vs. $AUC_{with\text{-}filters} = 0.97$ on the CTRL + AD subjects in cross-validation, and the performance on the trial set (MCI-NC + MCI-C) was $AUC_{null\text{-}filter} = 0.71$ vs. $AUC_{with\text{-}filters} = 0.74$. These results show that voxel intensity alone is already a very good feature.

Acknowledgments

This research was supported by Istituto Nazionale di Fisica Nucleare (INFN), Italy, under the project MAGIC-5 (Medical Application on a Grid Infrastructure Connection), a joint research project involving researchers from 6 different INFN sites in Italy: Genova, Torino, Pisa, Bari, Napoli and Lecce.

This research was also supported by grants to LR, PB and ME from Università degli Studi di Genova, Italy.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc., F.

Hoffman-La Roche, Schering-Plough, Synarc, Inc., as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

References

Archer, K.J., Kimes, R.V., 2008. Empirical characterization of random forest variable importance measures. Comput. Stat. Data Anal. 52, 2249–2260.

Bennett, D., Wilson, R., Bienias, J., Aggarwal, N., Mendes De Leon, C., Morris, M., Schneider, J., Evans, D., 2002. Cognitive activity and incident AD in a population-based sample of older persons. Neurology 59, 1910–1914.

Braak, H., Braak, E., Yilmazer, D., De Vos, R., Jansen, E., Bohl, J., 1996. Pattern of brain destruction in Parkinson's and Alzheimer's diseases. J. Neural Transm. 103, 455–490.

Breiman, L., 2001. Random forest. Mach. Learn. 45, 5–32.

Bylander, T., 2002. Estimating Generalization Error on Two-Class Datasets Using Out-of-Bag Estimates. Mach. Learn. 48, 287.

Calvini, P., Chincarini, A., Gemme, G., Penco, M.A., Squarcia, S., Nobili, F., Rodriguez, G., Bellotti, R., Catanzariti, E., Cerello, P., De Mitri, I., Fantacci, M.E., 2009. Automatic analysis of medial temporal lobe atrophy from structural MRIs for the early assessment of Alzheimer's disease. Med. Phys. 36, 3737.

Castleman, K., Schulze, M., Wu, Q., 1998. Simplified design of steerable pyramid filters, Circuits and Systems. Proceedings of the 1998 IEEE International Symposium, vol. 5. ISCAS '98, pp. 329–332.

Chupin, M., Grardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehricy, S., Benali, H., Garnero, L., Colliot, O., 2009. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from adni. Hippocampus 19, 579–587.

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

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Costafreda, S.G., Dinov, I.D., Tu, Z., Shi, Y., Liu, C.Y., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Wahlund, L.O., Spenger, C., Toga, A.W., Lovestone, S., Simmons, A., 2011. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. Neuroimage 56, 212–219.

Cuingnet, R., Gérardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage 56 (2), 766–781.

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

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., in press. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging., doi:10.1016/j.neurobiolaging.2010.05.023.

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., 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain J. Neurol. 132, 2048–2057.

Dickerson, B., 2001. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. Neurobiol. Aging 22, 747–754.

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, 497–510.

Dixon, R.M., Bradley, K.M., Budge, M.M., Styles, P., Smith, A.D., In, R., 2002. Longitudinal quantitative proton magnetic resonance spectroscopy of the hippocampus in Alzheimer's disease. Brain J. Neurol. 125, 2332–2341.

Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDADRDA criteria. Lancet Neurol. 6, 734–746.

Dubois, B., Feldman, H.H., Jacova, C., Cummings, J.L., DeKosky, S.T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N.C., Galasko, D., 2010. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 9.

Duchesne, S., 2006. MICCAI brainstem segmentation protocol. From Statistical Atlases to Personalized Models (Workshop). MICCAI Society, Copenhagen, Denmark.

Ebina, T., Toh, H., Kuroda, Y., 2011. DROP: an SVM domain linker predictor trained with optimal features selected by random forest. Bioinformatics 27, 487–494.

Fan, Y., Shen, D., Gur, R., Gur, R., Davatzikos, C., 2007. Compare: classification of morphological patterns using adaptive regional elements. IEEE Trans. Med. Imaging 26, 93–105.

Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C., 2008. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39, 1731–1743.

Fennema-Notestine, C., Hagler Jr., D.J., Mcevoy, L.K., Fleisher, A.S., Wu, E.H., Karow, D.S., Dale, A.M., 2009. Structural MRI biomarkers for preclinical and mild Alzheimer's disease. Hum. Brain Mapp. 30, 3238–3253.

Ferrarini, L., Frisoni, G.B., Pievani, M., Reiber, J.H.C., Ganzola, R., Milles, J., 2009. Morphological hippocampal markers for automated detection of Alzheimer's disease and mild cognitive impairment converters in magnetic resonance images. J. Alzheimers Dis. JAD 17, 643–659.

Ferreira, L.K., Diniz, B.S., Forlenza, O.V., Busatto, G.F., Zanetti, M.V., in press. Neurostructural predictors of Alzheimer's disease: a meta-analysis of VBM studies. Neurobiol. Aging., doi:10.1016/j.neurobiolaging.2009.11.008.

Freeborough, P.a., Fox, N.C., 1998. MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease. IEEE Trans. Med. Imaging 17, 475–479.

Frisoni, G.B., Sabattoli, F., Lee, a.D., Dutton, R.a., Toga, a.W., Thompson, P.M., 2006. In vivo neuropathology of the hippocampal formation in AD: a radial mapping MR-based study. Neuroimage 32, 104–110.

Guyon, I., Vapnik, V., Barnhill, J., Stephen, W., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.

Healy Jr., D.M., Weaver, J.B., 1992. Two applications of wavelet transforms in magnetic resonance imaging. IEEE Trans. Inf. Theory 38, 840–860.

Hilton, M., Ogden, T., Hattery, D., Eden, G., Jawerth, B., 1996. Wavelet denoising of functional mri data.

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M.K., Johnson, S.C., 2009. Spatially augmented LPBoosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48 (1), 138–149.

Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., 2011. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage 55 (2), 574–589.

Holland, D., Brewer, J.B., Hagler, D.J., Fenema-Notestine, C., Dale, A.M., 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc. Natl. Acad. Sci. U.S.A. 106 (49), 20954–20959.

Karas, G., Sluimer, J., Goekoop, R., van der Flier, W., Rombouts, S.a.R.B., Vrenken, H., Scheltens, P., Fox, N., Barkhof, F., 2008. Amnestic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease. AJNR Am. J. Neuroradiol. 29, 944–949.

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

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

Lao, Z., 2004. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21, 46–57.

Larios, N., Soran, B., Shapiro, L.G., Martinez-Munoz, G., Lin, J., Dietterich, T.G., 2010. Haar Random Forest Features and SVM Spatial Matching Kernel for Stonefly Species Identification. PR International Conference on Pattern Recognition. IEEE, pp. 2624–2627.

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, 995–1001.

Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., Shen, D., in press. Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. Neurobiol. Aging., doi:10.1016/j.neurobiolaging.2010.11.008.

Liu, Y., Paajanen, T., Zhang, Y., Westman, E., Wahlund, L.O., Simmons, A., Tunnard, C., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Evans, A., Spenger, C., Lovestone, S., Soininen, H., 2010. Analysis of regional MRI volumes and thicknesses as predictors of conversion from mild cognitive impairment to Alzheimer's disease. Neurobiol. Aging 31, 1375–1385.

Mazziotta, L., 1995. A probabilistic atlas of the human brain: theory and rationale for its developement. Neuroimage 2, 89–101.

McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M., 2009. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology 251, 195–205 PMID: 19201945.

Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage 44, 1415–1422 PMID: 19027862.

Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Toga, A.W., Jack, C.R., Schuff, N., Weiner, M.W., Thompson, P.M., 2009. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Neuroimage 45, S3–S15 PMID: 19041724.

Mosconi, L., Brys, M., Glodzik-Sobanska, L., De Santi, S., Rusinek, H., de Leon, M.J., 2007. Early detection of Alzheimer's disease using neuroimaging. Exp. Gerontol. 42, 129–138.

Nowak, R., 1999. Wavelet-based rician noise removal for magnetic resonance imaging. IEEE Trans. Image Process. 8, 1408–1419.

Pelaez-Coca, M., Bossa, M., Olmos, S., 2011. Discrimination of AD and normal subjects from MRI: anatomical versus statistical regions. Neurosci. Lett. 487 (1), 113–117.

Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M., 2003. A versatile wavelet domain noise filtration technique for medical imaging. IEEE Trans. Med. Imaging 22, 323–331.

Plant, C., Teipel, S.J., Oswald, A., Bhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based

on MRI for the prediction of Alzheimer's disease. Neuroimage 50, 162–174 PMID: 19961938.
Portilla, J., Strela, V., Wainwright, M., Simoncelli, E., 2003. Image denoising using scale mixtures of gaussians in the wavelet domain. IEEE Trans. Image Process. 12, 1338–1351.
Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. Cereb. Cortex 10, 433–442.
Rex, D.E., Ma, J.Q., Toga, A.W., 2003. The LONI Pipeline Processing Environment. Neuroimage 19, 1033–1048.
Risacher, S.L., Shen, L., West, J.D., Kim, S., McDonald, B.C., Beckett, L.a., Harvey, D.J., Jack, C.R., Weiner, M.W., Saykin, A.J., 2010. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. Neurobiol. Aging 31, 1401–1418.
Saeys, Y., Abeel, T., Peer, Y.V.D., 2008. LNAI 5212 - Robust Feature Selection Using Ensemble Feature Selection Techniques. Lecture Notes in Computer Science, Vol. 5212/2008, 09/2008, pp. 313–325.
Schiavi, E., Hernandez, C., Hernandez, J.A., 2004. Fully 3d wavelets mri compression. Biological and Medical Data Analysis. Lecture Notes in Computer Science, vol. 3337. Springer, Berlin / Heidelberg, pp. 9–20.
Seber, G., 1984. Multivariate Observations. Wiley.
Shen, D., Moffat, S., Resnick, S.M., Davatzikos, C., 2002. Measuring size and shape of the hippocampus in MR images using a deformable shape model. Neuroimage 15, 422–434.
Singh, R.K., Naik, S.K., Gupta, L., Balakrishnan, S., Pai, K.M., 2008. Hybrid SVM—Random Forest Classification System for Oral Cancer Screening using LIF Spectra. International Conference on Pattern Recognition (ICPR 2008), Tampa, Florida, USA, December 2008.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 (Suppl. 1), S208–S219.
Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43, 1947–1958.
Thompson, P.M., Hayashi, K.M., Zubicaray, G.D., 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, 994–1005.
Thompson, P.M., Hayashi, K.M., Zubicaray, G.I.D., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., Toga, A.W., 2004. Mapping hippocampal and ventricular change in Alzheimer disease. Neuroimage 22, 1754–1766.
Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jr, C.R.J., 2008. Alzheimer's disease diagnosis in individual subjects using structural mr images: validation studies. Neuroimage 39, 1186–1197.
Vemuri, P., Wiste, H.J., Weigand, S.D., Knopman, D.S., Trojanowski, J.Q., Shaw, L.M., Bernstein, M.A., Aisen, P.S., Weiner, M., Petersen, R.C., Jack, C.R., 2010. Serial MRI and CSF biomarkers in normal aging, MCI, and AD. Neurology 75, 143–151.
Walhovd, K.B., Fjell, a.M., Brewer, J., McEvoy, L.K., Fennema-Notestine, C., Hagler, D.J., Jennings, R.G., Karow, D., Dale, a.M., 2010a. Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. AJNR Am. J. Neuroradiol. 31, 347–354.
Walhovd, K.B., Fjell, a.M., Dale, a.M., McEvoy, L.K., Brewer, J., Karow, D.S., Salmon, D.P., Fennema-Notestine, C., 2010b. Multi-modal imaging predicts memory performance in normal aging and cognitive decline. Neurobiol. Aging 31, 1107–1121.
Wang, Z., Bovik, A., 2009. Mean squared error: love it or leave it? a new look at signal fidelity measures. Signal Process. Mag. IEEE 26, 98–117.
Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612.
Wang, Y., Fan, Y., Bhatt, P., Davatzikos, C., 2010. High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. Neuroimage 50, 1519–1535.
Waske, B., Van Der Linden, S., 2008. Classifying multilevel imagery from SAR and optical sensors by decision fusion. IEEE Trans. Geosci. Remote Sens. 46, 1457–1466.
Waske, B., Van Der Linden, S., Benediktsson, J.A., Rabe, A., Hostert, P., 2010. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. IEEE Trans. Geosci. Remote Sens. 48, 2880–2889.
Whitwell, J.L., Przybelski, S.A., Weigand, S.D., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R., 2007. 3D maps from multiple mri illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to Alzheimer's disease. Brain 130, 1777–1786.
Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., Petersen, R.C., 2004. Mild cognitive impairment—beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment. J. Intern. Med. 256, 240–246.
Yoo, T.S., Ackerman, M.J., Lorensen, W.E., Schroeder, W., Chalana, V., Aylward, S., Metaxas, D., Whitaker, R., 2002. Engineering and algorithm design for an image processing Api: a technical report on ITK–the Insight Toolkit. Stud. Health Technol. Inform. 85, 586–592.
Zhu, X., Ambroise, C., Mclachlan, G.J., 2006. Selection bias in working with the top genes in supervised classification of tissue samples. Stat. Method. 3, 29–41.