Technology Update

Predictive Models Based on Support Vector Machines: Whole-Brain versus Regional Analysis of Structural MRI in the Alzheimer's Disease

Alessandra Retico, PhD, Paolo Bosco, PhD, Piergiorgio Cerello, PhD, Elisa Fiorina, PhD, Andrea Chincarini, MSc, Maria Evelina Fantacci, PhD

From the Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Pisa, Italy (AR, MEF); Dipartimento di Fisica, Università degli Studi di Genova, Genova, Italy (PB); Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Genova, Italy (PB, AC); Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Torino, Italy (PC, EF); Dipartimento di Fisica, Università di Torino, Torino, Italy (EF); and Dipartimento di Fisica, Università di Pisa, Pisa, Italy (MEF).

ABSTRACT

Decision-making systems trained on structural magnetic resonance imaging data of subjects affected by the Alzheimer's disease (AD) and healthy controls (CTRL) are becoming widespread prognostic tools for subjects with mild cognitive impairment (MCI). This study compares the performances of three classification methods based on support vector machines (SVMs), using as initial sets of brain voxels (ie, features): (1) the segmented grey matter (GM); (2) regions of interest (ROIs) by voxel-wise t-test filtering; (3) parceled ROIs, according to prior knowledge. The recursive feature elimination (RFE) is applied in all cases to investigate whether feature reduction improves the classification accuracy. We analyzed more than 600 AD Neuroimaging Initiative (ADNI) subjects, training the SVMs on the AD/CTRL dataset, and evaluating them on a trial MCI dataset. The classification performance, evaluated as the area under the receiver operating characteristic curve (AUC), reaches AUC = $(88.9 \pm .5)\%$ in 20-fold cross-validation on the AD/CTRL dataset, when the GM is classified as a whole. The highest discrimination accuracy between MCI converters and nonconverters is achieved when the SVM-RFE is applied to the whole GM: with AUC reaching $(70.7 \pm .9)\%$, it outperforms both ROI-based approaches in predicting the AD conversion.

Introduction

Diagnostic early markers of Alzheimer's disease (AD) are widely investigated in order to aid researchers and clinicians in the validation of the effectiveness of new AD treatments, while limiting the duration and cost of clinical trials.[1] Brain atrophy and its evolution are recognized key neuroimaging biomarkers of the AD progression.[2–9]

Brain scans acquired with structural magnetic resonance imaging (MRI) are suitable for quantitative evaluation of brain atrophy; moreover, image analysis algorithms can process these in order to extract useful information and carry out both single-subject analyses and between-group comparisons.

Machine-learning techniques applied to MRI images are increasingly spreading tools that extract information from data

and can predict the pathology progression.[10,11] The introduction of pattern classification and computer vision methods in the neuroimaging field is due to Lao et al.[12] In order to develop an accurate predictor of pathology from a set of volumetric images, the authors highlighted the limitations of voxel-based morphometric methods and the potentiality of support vector machine (SVM)-based decisional systems.

Classification approaches can be particularly useful in studies on the AD pathology, as they predict the conversion to AD of those subjects referred to as affected by mild cognitive impairment (MCI), a transitional state between normal aging and dementia.[13]

Several machine-learning techniques have been implemented so far in neuroimaging studies on AD, such as

principal component analysis and linear discrimination analysis,[14,15] SVM,[16–21] logistic regression,[22] the combination of information extracted from different diagnostic modalities and their classification with SVM,[23] with an automatic learning framework,[24] or by combining multiple weak classifiers to achieve more accurate and robust results.[25]

Many research groups investigating the AD pathology in recent years developed and validated their analyses on data samples provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI),[15,19–26] as reported in the extensive review carried out by the principal investigator on the ADNI initiative Michael W. Weiner and colleagues.[27]

Recent studies by Cuingnet et al[20] and Chu et al[21] on ADNI data were focused on a comparison between whole-brain classification methods and region of interest (ROI)-based approaches, all coupled to feature reduction methods, required by the huge input data size. Both studies concluded that feature selection does not improve the classification accuracy and that whole-brain methods outperform ROI-based ones, unless the ROIs are chosen according to a prior knowledge of the underlying disease. A recent study by Adaszewski et al[28] on longitudinal data of AD, MCI, and control (CTRL) subjects reports that feature selection improves the classification accuracy at early MCI stages, whereas at a later stage whole-brain methods are superior.

We focused our analysis on a comparison between three different methods, in order to select the initial sets of brain voxels to analyze with SVM classifiers. In particular, they are defined as: (1) the segmented grey matter (GM), referred as whole-GM classification; (2) the ROIs, obtained by the voxel-based morphometry (VBM) analysis, ie, voxel-wise t-test filtering, referred as VBM-ROI classification; and (3) the parceled ROIs, chosen according to prior knowledge on the brain involvement in the AD pathology, where the parcellation of the anatomical regions is performed according to Laboratory of Neuro Imaging (LONI) Probabilistic Brain Atlas (LPBA40),[29] thus referred as LONI-ROI classification.

The recursive feature elimination (RFE) method[30] has been implemented in our whole-GM analysis both in order to reduce the data size, and to localize the brain regions that are more involved in the AD pathology. The same approach has been applied also to the VBM-ROI and LONI-ROI classification, so as to evaluate the SVM-RFE potentiality in enhancing the classification performance. Methods (1) and (2), complemented with the RFE procedure, can be considered as data-driven approaches, where different criteria to select the initial set of features are implemented (whole-GM vs. t-test filtered ROIs). They are compared to (3) a prior-knowledge based analysis, where the right and left hippocampi and parahippocampal gyrii are considered as relevant ROIs in the AD pathology.

All the approaches are compared in terms of their ability in correctly distinguishing AD from CTRL subjects and making accurate conversion predictions on the MCI population. The relative performances are compared in terms of the area under the receiver operating characteristic (ROC) curve (AUC), evaluated within cross-validation protocols. The meaning of AUC has been proved to be the probability that a random pair of positive/diseased and negative/nondiseased individuals would be correctly identified by the diagnostic test.[31]

Table 1. Demographic and Mini Mental State Examination (MMSE) Score of the Data Samples

Cohort	Sample Size	Age	Gender (M/F)	MMSE
CTRL	189	76.6 ± 5.1	95/95	29.1 ± .9
MCI-NC	166	75.7 ± 7.3	106/60	27.2 ± 2.4
MCI-C	136	75.1 ± 7.1	80/56	25.2 ± 2.7
AD	144	75.5 ± 7.5	78/66	22.2 ± 3.3

CTRL = control; AD = Alzheimer's disease; MCI-NC = MCI nonconverted to AD.

The analysis is carried out on an MRI dataset extracted from the publicly available ADNI collection.

This paper is structured as follows: a description of the MRI data source and its characteristics is provided; then, a methodological overview for the VBM analysis and the machine-learning SVM procedure is given; the implementation of decisional systems in the whole-GM, VBM-ROI, and LONI-ROI analysis is described, and finally results are discussed and compared to other methods.

Materials and Methods
The ADNI Data Sample

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

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

The analysis reported in this paper has been carried out on the structural MRI data of 635 subjects extracted from the ADNI database. The statistical data of these subjects are summarized in Table 1. The subjects have been divided into two categories: a training/testing 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,

among which 136 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 the availability of baseline and at least 2-years information. Moreover, training subjects were chosen if confirmed to be CTRL/AD at follow-up assessment. The data samples considered in this work are the same analyzed in the paper by Chincarini et al.[19]

MRI Acquisition and Preprocessing

The MRI ADNI data were acquired with 1.5 T scanners. Data were collected across a variety of scanners. Up-to-date information on ADNI eligibility criteria and protocols is available on http://www.adni-info.org. Raw MRI scans converted to the Neuroimaging Informatics Technology Initiative (NIfTI) format 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.[19] Denoized scans were registered onto the Montreal Neurological Institute (MNI) reference[32] with a 12-parameter affine registration and resampled onto a 1-mm^3 isotropic grid. Images were then intensity-normalized as reported in Chincarini et al[19] to achieve good histogram equalization among images coming from different scanners, while ensuring that the average gray levels of the three main cerebral matter contributions are mapped onto those of the MNI reference.

VBM Preprocessing

We conducted a VBM study[33] in order to investigate the differences in the regional volumes of GM between the AD and CTRL groups and between the MCI converted to AD (MCI-C) and MCI nonconverted to AD (MCI-NC) groups, respectively. The T1-weighted volumetric images were analyzed with the latest release of the Statistical Parametric Mapping (SPM) software package, SPM8 (Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ucl.ac.uk/spm), using the VBM protocol with modulation. We implemented the Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL) algorithm.[34] A diffeomorphic warping is implemented to achieve an accurate inter-subject registration with an improved realignment of small inner structures[35] and to generate a study-specific template. The VBM preprocessing was applied as follows: (1) SPM segmentation of brain tissues, ie, GM, white matter (WM), and cerebrospinal fluid (CSF), using the New Segment toolbox, which extends the segmentation procedure described in[36] (2) importing the parameter files produced by the tissue segmentation in the DARTEL procedure[34] to generate a study-specific template and the deformation fields that warp the segmented tissues of each subjects to the DARTEL template; (3) warping to the MNI space of the DARTEL template and of the segmented brain tissues previously aligned to the DARTEL template according to the generated deformation fields; (4) standard smoothing with isotropic Gaussian kernel. After the preprocessing, we obtained smoothed modulated normalized data (in the MNI space) to be used for the statistical analysis. The modulation operation is fundamental to render the final VBM statistics reflecting the local volume dif-
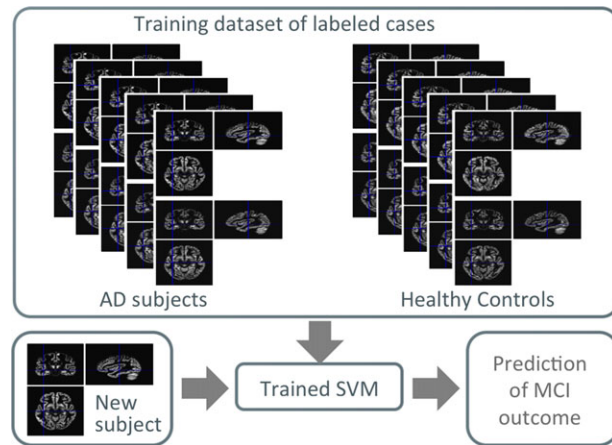
Fig 1. Schematic representation of the SVM training on the AD/CTRL sample and the SVM validation on the MCI sample.

ferences in tissue segments.[33] Modulation allows compensating for the effect of spatial warping that causes volume changes, so that the total amount of GM in the modulated images remains the same as it would be in the original images. The modulated GM segments thus need to be corrected for total intracranial volume (TIV) to take into account brain size variability.

VBM Statistical Analysis

The regional GM volumes were compared between the AD/CTRL and the MCI-C/MCI-NC samples using the VBM-DARTEL analysis. The normalized modulated and smoothed GM image segments in each group were entered into a voxel-wise two-sample *t*-test analysis in SPM8. The conventional VBM analysis was employed using the stringent significance threshold $P < .05$, family-wise error rate (FWE) corrected. An absolute threshold mask of .1 on GM was used to avoid possible edge effects around the border between GM and WM. Age, gender, and the TIV, computed as the sum of the SPM segmented GM, WM, and CSF were entered as covariates in the statistical analysis.

Multivariate Analysis with SVM

We followed the classification approach with SVM proposed in Klöppel et al[16] and in Ecker et al.[37] As opposed to the mass-univariate VBM analysis, the pattern recognition techniques, eg, SVMs, are multivariate and make use of specific inter-regional dependencies to help categorize scans.[12,17]

An SVM[38] is a supervised binary classification method, ie, it requires a training set, used to learn the differences between the two groups, and a validation set to quantify the classification performance on previously unseen data (see the schematic flowchart in Fig 1). In our analysis, each image is treated as a point in the high dimensional R^N space, where the space dimension N is equal to the number of features/voxels in the considered ROI.

The SVM inputs are feature vectors, the sequence of the voxel intensity values of the ROI. The feature vectors belonging either to the category of patients and CTRLs are labeled with "1" and "−1", respectively. The trained SVM classifier

maps the R^N space into R, by assigning to the newly examined cases a one-dimensional (or binary) output. As the number of features/voxels can be very high especially when the whole GM is the ROI to classify, whereas the number of subjects considered in this study is limited to few hundreds, we considered only linear-kernel SVMs to avoid the risk of overfitting data. Training an SVM is a minimization problem where the largest margin hyperplane allowing for an optimal separation of the training examples is identified. The separating hyperplane is defined by a weight vector and an offset, $\underline{w} \cdot \underline{x} + b = 0$, where the weight vector $\underline{w}$ is a linear combination of the support vectors and it is normal to the hyperplane. During the SVM training a free parameter has to be set, the c value, which controls the trade-off between having zero training errors and allowing for misclassifications. The c value has been heuristically estimated in our analysis.

The SVM is trained according to the 20-fold cross-validation (20f-CV) technique. The data are partitioned in 20 folds; one of them is retained as validation data while the others are used to train the classifier. The process is repeated 20 times, ie, until each subsample is used once as validation set. We used in this study SVM-Light software package [39,40] developed by T. Joachims (http://svmlight.joachims.org/), which is freely available for scientific and non-commercial use.

The classification performance is evaluated through the ROC curve,[41] where the sensitivity (true positive rate, ie, the percentage of pathological subjects correctly classified) is plotted against the false positive rate (ie, the percentage of misclassified CTRL subjects). Different ROC curves are compared to each other in terms of the estimated AUC.

The initial selection of brain regions for the three SVM classifications we carried out was operated as follows:

(1) *Whole-GM classification*: We consider as input to the SVM classifier the whole GM segmented volume obtained from the standardized and automated SPM preprocessing.[16,37]
(2) *VBM-ROI classification*: The ROIs that reached the statistical significance in the VBM analysis ($P < .05$, FWE corrected) were used as data-driven reduced input to the SVM classifier.
(3) *LONI-ROI classification*: The (1) and (2) voxel selection methods are complemented with a prior-knowledge based approach, where brain regions that encode interesting information for the AD pathology are selected on the basis of previous studies.[2–4] Usually Neuroradiologists are asked to manually identify the brain regions to be considered. As an alternative, atlas-based parcellation can be implemented to select anatomical regions in a more automated and reproducible way. In the present analysis we defined the ROIs according to the LPBA40[29] (http://www.loni.ucla.edu.Atlases/LPBA40). In particular, brain regions where neurodegeneration is expected in AD have been chosen, ie, the right and left hippocampi and the parahippocampal gyrii.

Discrimination Maps and RFE (SVM-RFE)

The implementation of a linear-kernel SVM allows a direct extraction of the weight vector $\underline{w}$ as an image, which is referred to as discrimination map. The $\underline{w}$ vector, which is normal to the separating hyperplane and indicates the direction along which

the images of the two groups differ most, can be used to generate a map of the most discriminating voxels in the images. As the intensity value reported in each voxel of the considered input images is proportional to the amount of GM in that specific location (the modulation option has been selected in the SPM segmentation), a higher/lower value in the discrimination map indicates that patients have higher/lower GM volume in that specific location with respect to CTRLs.

In order to identify the voxels with the highest discriminating power, we implemented the SVM classification with the RFE procedure (SVM-RFE).[30,42] It is a feature-selection technique that iteratively eliminates features/voxels from the data set, in order to remove as many noninformative features as possible, while retaining features that carry discriminative information. A new SVM classifier is trained at each iteration, thus a new weight map is generated. The selected feature-ranking criterion is the absolute value of each weight vector component $|w_i|$. The features/voxels are iteratively excluded from the dataset with the aim of removing as many noninformative voxels as possible (low $|w_i|$), while retaining those encoding the discriminative information (high $|w_i|$).

The SVM-RFE algorithm is implemented in this study to estimate the SVM classification performance as a function of the number of GM voxels retained as input to the SVM. At each operative point of this curve the discrimination maps can be visualized to localize the set of voxels that encode the between-group discriminant information.

Prediction on the Outcome of MCI Subjects

The SVM-based method we present in this paper is able to provide single-subject classification, ie, once an SVM classifier is trained on a learning dataset, it can be applied on previously unseen data and give a prediction on its class membership. In our analysis the SVMs, trained and validated on the AD/CTRL data sample, are evaluated on a completely independent validation set, constituted by MCI subjects. Also in this case the classifier performance is evaluated using the AUC as figure of merit.

### Correlation between SVM Test Margin and Cognitive Decline

When attempting to set up a useful imaging biomarker, sensitive to the cognitive decline of patients with AD, it is important to test for correlation with the mini mental state examination (MMSE) score.[1] We studied the correlation between the distance of each data point from the SVM optimal hyperplane (we will refer to as the test margin) and the MMSE score of each subject using the Spearman's rank correlation coefficient ρ.[43]

Results
VBM Results

Significant volumetric between-group differences have been found in the AD/CTRL sample with Gaussian smoothing scale s = 3 mm. By contrast, no significant between-group difference has been detected in the MCI-C/MCI-NC analysis at any considered smoothing scale (s = 3, 6, 8 mm). The areas of brain atrophy detected by VBM are shown in Figure 2, where the t statistics of the significantly different areas ($P < .05$, FWE
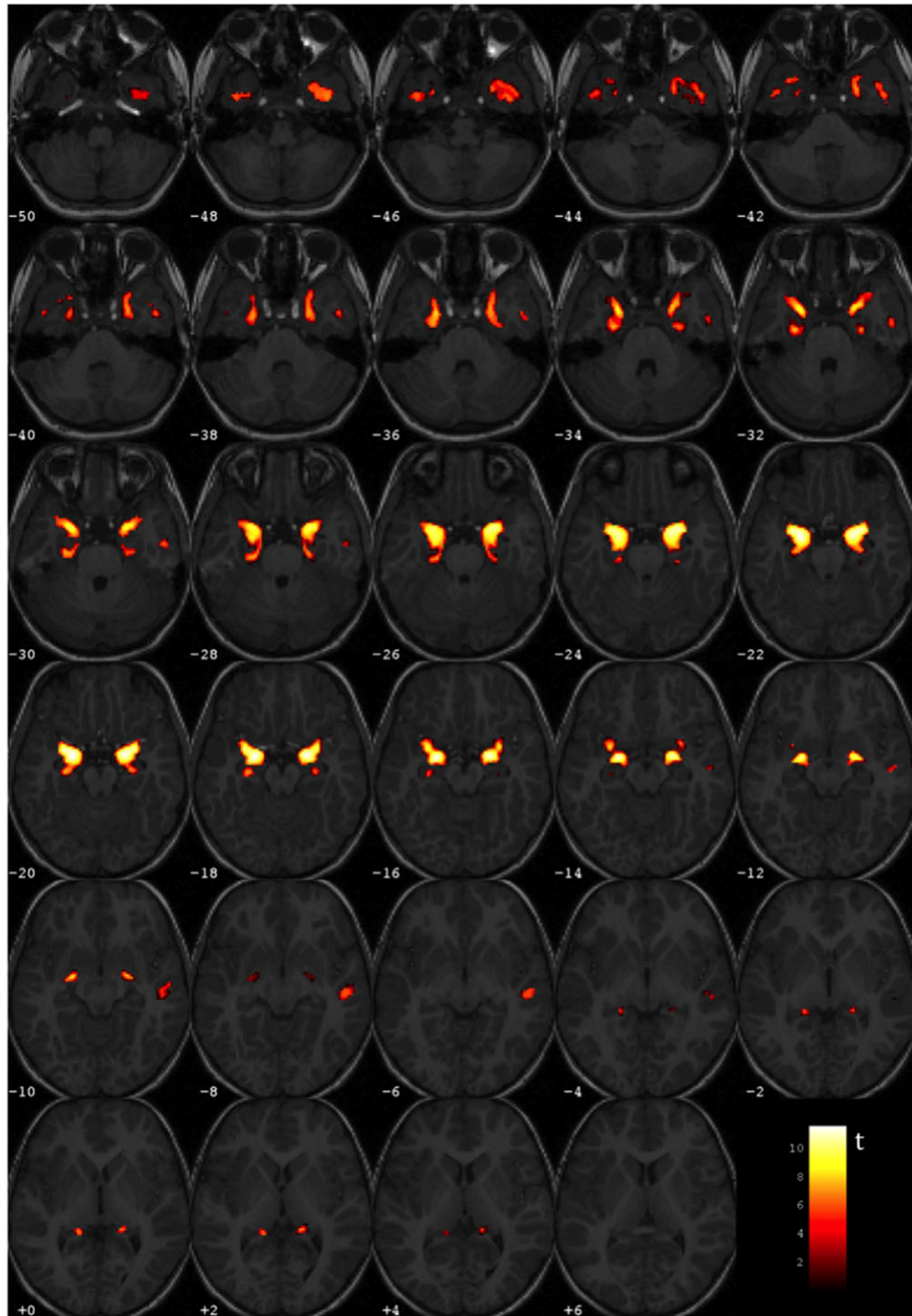
Fig 2. Significant regions detected in the AD vs. CTRL VBM statistical analysis (colored regions reporting the t statistics), overlaid on a representative single-subject T1-weighted MRI (a sequence of axial views is displayed).

corrected) is overlaid on a series of axial views of a representative T1 MRI of a single subject. The effect is localized in the medial temporal lobe, as expected. The VBM cluster coordinates in the MNI reference space and in the standard Talairach and Tournoux space[44] (the conversion being performed with the Talairach Client),[45,46] beside the anatomical description of the corresponding brain areas, are reported in Table 2.

SVM Classification, SVM-RFE, and Discrimination Maps

(1) Whole-GM classification

The SVMs have been trained on the AD/CTRL sample according to the 20f-CV protocol and validated on the MCI cohort. In this paper we address a classification problem where the input is a vector that we call a pattern of n components, which we call features. Thus, a pattern is generated for each

Table 2. Anatomical Structures Detected in the Alzheimer's Disease versus Control Voxel-Based Morphometry Analysis (Data Smoothed with $s = 3$ mm). These Regions are Visible in Figure 2.

Cluster Extent (Voxels)	MNI Coordinates			Talairach Coordinates				Brain Region Localization	
2646	−22	−6	−21	−21	−5	−16	LC	Limbic lobe	Parahippocampal gyrus, amygdala
2892	21	−4	−18	19	−4	−12	RC	Limbic lobe	Parahippocampal gyrus, amygdala
62	−16	−39	1	−16	−38	1	LC	Limbic lobe	Parahippocampal gyrus, BA27
57	18	−36	1	16	−35	2	RC	Limbic lobe	Parahippocampal gyrus, BA27
158	56	−19	−8	51	−19	−4	RC	Temporal lobe	Superior temporal gyrus, BA22

LC = left cerebrum; RC = right cerebrum; BA = Brodmann area.

subject of our dataset, and the *n* features of each input vector are the voxel intensity values of the GM of each subject. It happens that the number of features/voxels is very large (about 6.5×10^5), whereas the number of patterns in the training dataset is limited to about 300. The linear-kernel SVM demonstrated to be able to handle such disproportion between the number of weights to be estimated and the training examples.

The classification error due to the random partitioning of data into the train and test samples was estimated by repeating the 20f-CV procedure 10 times and evaluating the corresponding average AUC and its standard deviation (SD). As shown in Figure 3, the discrimination performance of SVMs trained with all features/voxels of the GM segments is equal to AUC = $(88.9 \pm .5)$% on the AD/CTRL sample, whereas it falls to AUC = $(67.8 \pm .5)$% on the MCI cohort.

The SVM-RFE procedure has been applied and the corresponding AUC variation as a function of the number of retained voxels is shown in Figure 3. The dotted lines above and below the curve highlight the average (AUC ± SD) band.

It can be noticed that, whereas the highest AUC values in the AD/CTRL discrimination are obtained when a large number of GM voxels are retained for the classification, the performance sensibly decreases when retaining less than 10,000 voxels. By contrast, the trend of AUC in the MCI-C/MCI-NC SVM-RFE classification is completely different. We remind that the MCI cohort is used only as validation sample, ie, the SVM classifiers trained on the AD/CTRL data are evaluated on the MCI cases, which did not influence the training process. The reduction in the number of retained voxels leads in this case to a slight optimization of the classification performance, up to AUC = $(70.9 \pm .9)$%, obtained with 8,000 retained voxels.

(2) VBM-ROI classification

The SVM classifier was trained with the grey level information encoded in the ROI identified by the VBM statistical analysis (see Sec. VBM results). About 6×10^3 voxels were significantly different in the AD versus CTRL statistical comparison. The SVM classification, according to the 20f-CV protocol repeated 10 times, achieved AUC = $(85.4 \pm .3)$% in the discrimination between AD and CTRL subjects. The validation performance on the MCI dataset was less effective, with AUC = $(63.7 \pm .2)$%.

In order to optimize the discrimination performance, the SVM-RFE procedure has been applied also in this case. The behavior of AUC as a function of the number of retained voxels is shown in Figure 3 for both the AD/CTRL data and the validation cohort of the MCI-C/MCI-NC subjects. It can be

noticed that despite the SVM-RFE procedure leads in this case to a tiny improvement in the classification performance in both the AD/CTRL and MCI-C/MCI-NC classification, in the latter case the AUC values remain well below 70%.

(3) LONI-ROI classification

The hippocampus and parahippocampal gyrus ROIs extracted from the LPBA40 were coregistered to our images and resliced to be used as masks to the segmented GM of our data sample. The resulting LONI-ROI dataset consisted of vectors of about 14×10^3 features, whose SVM classification provided AUC = $(88.1 \pm .3)$% on the AD/CTRL sample and AUC = $(65.6 \pm .3)$% on the MCI cohort. As shown in Figure 3, the SVM-RFE procedure has almost no effect on the classification performance.

### *Global versus local approaches and discrimination information*

A direct comparison between the classification performance of the whole-GM and the ROI-based methods is shown in Figure 3, which highlights two main results.

I. Before the RFE procedure is applied the whole-GM classification can be directly compared with *t*-test filtering (VBM-ROI) and prior knowledge-based method. The whole-GM classification outperforms both the ROI-based methods, especially in the prediction of the MCI outcome.

II. If the RFE is applied to all three datasets, the following considerations hold: (i) the LONI-ROI approach shows similar performance to the whole-GM method in AD versus CTRL classification, even when classifying only with a few hundred voxels, but its classification accuracy on MCI subjects is not fully satisfactory; (ii) the restrictive choice of classifying only the *t*-test filtered ROIs leads to a limited classification accuracy on both the AD/CTRL and the MCI samples; (iii) the RFE applied to the whole-GM data maximizes the performance in MCI outcome prediction.

The AUC values obtained in the whole-GM SVM-RFE analysis considering 6,000 retained voxels (corresponding approximately to the VBM ROI size) on the AD/CTRL and on the MCI-C/MCI-NC samples correspond to the AUC shown in Figure 4, where the average and SD band over the 10 repetition of the 20f-CV are shown. The 80% sensitivity and 80% specificity values characterize the AD/CTRL sample separation, whereas in the MCI classification the performance is limited to the 70% sensitivity and 62% specificity values.

The discrimination map can be visualized at each step of the SVM-RFE procedure applied to the whole-GM data; in order
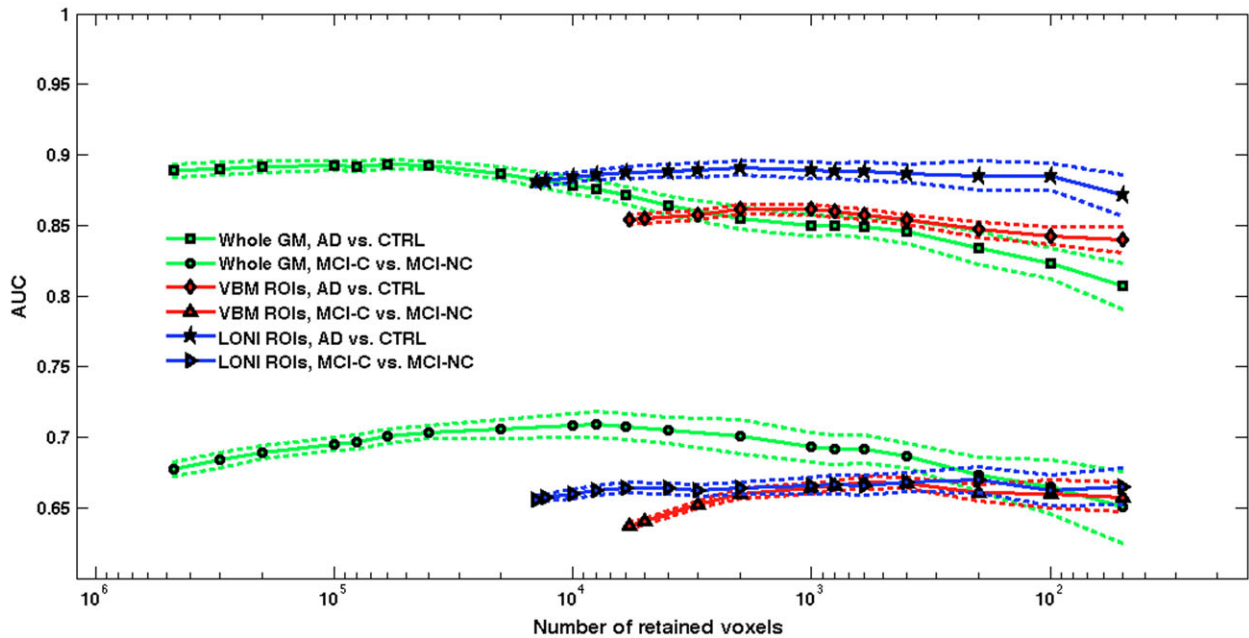
Fig 3. SVM-RFE of whole-GM, VBM-ROI and LONI ROI analyses. The SVM training and testing is performed on the AD/CTRL sample in 20f-CV; an independent validation is carried out on the MCI-C/MCI-NC sample. The average values obtained over 10 repetitions of 20f-CV are shown; the bands around the average curves correspond to two standard deviations.
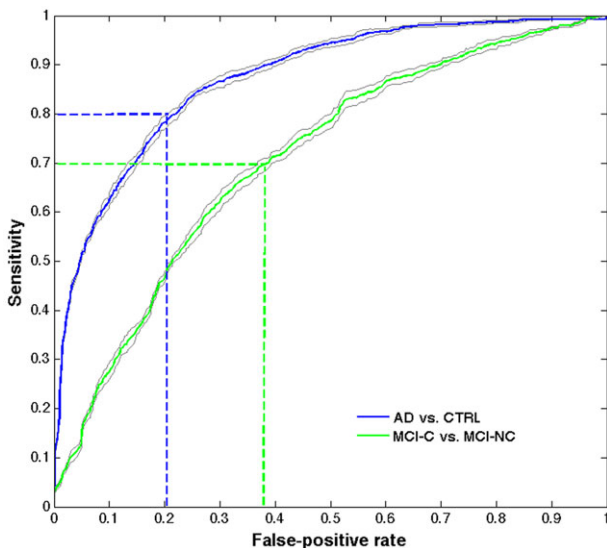

Fig 4. ROC obtained in the SVM-RFE procedure on the AD/CTRL and on the MCI-C/MCI-NC samples, considering 6000 retained voxels (corresponding approximately to the VBM ROI size). The curves averaged over the ten repetition of the 20f-CV are shown, surrounded by the $\pm$ SD band.

to directly compare it to the significant ROIs obtained with the VBM analysis, it has been visualized at the operative point corresponding to 6,000 retained voxels. The computed weight vectors w_r^k are averaged on r and k (where $k = 1, \ldots 20$ refers to the 20 different vectors obtained in the 20f-CV protocol,

and $r = 1, \ldots 10$ refers to the 10 repetitions of the SVM-RFE procedure). The discrimination map is shown in Figure 5, where regions with positive and negative values of w are reported; it is possible to distinguish brain regions where GM is either greater or lower in the patient group with respect to the CTRL group. Since the map was obtained by averaging the 200 different maps generated in the 10 repetitions of the 20f-CV, it is highly stable with respect to the training case variability. The discriminant regions, therefore, reflect the underlying characteristics of the AD pathology for the general population we analyzed.

The more extended and effective discriminant regions are described in terms of coordinates in the MNI reference space and in the standard space of Talairach and Tournoux in Table 3, where the anatomical description of the corresponding brain areas is also provided. It can be noticed that the parahippocampal gyrii and the superior temporal gyrus (BA22) have consistently been found in the VBM statistical univariate analysis. However, the discriminant information that allows the AD/CTRL separation appears not to be mainly localized in the Limbic lobes, as highlighted by the VBM analysis in the VBM-ROI approach, and as a-priori defined through the LONI-ROIs. By contrast, many regions spread over the whole GM contribute to the two-class separation. In addition, the blind validation of this discrimination pattern on the MCI-C/MCI-NC sample has shown enhanced discrimination ability with respect to the SVM classification of the VBM-ROIs and LONI-ROIs. The latter is a major result of the present analysis suggesting that relevant information to make predictions on the MCI population may reside in regions of the brain other than those that are most relevant when the AD pathology has already reached an advanced stage.
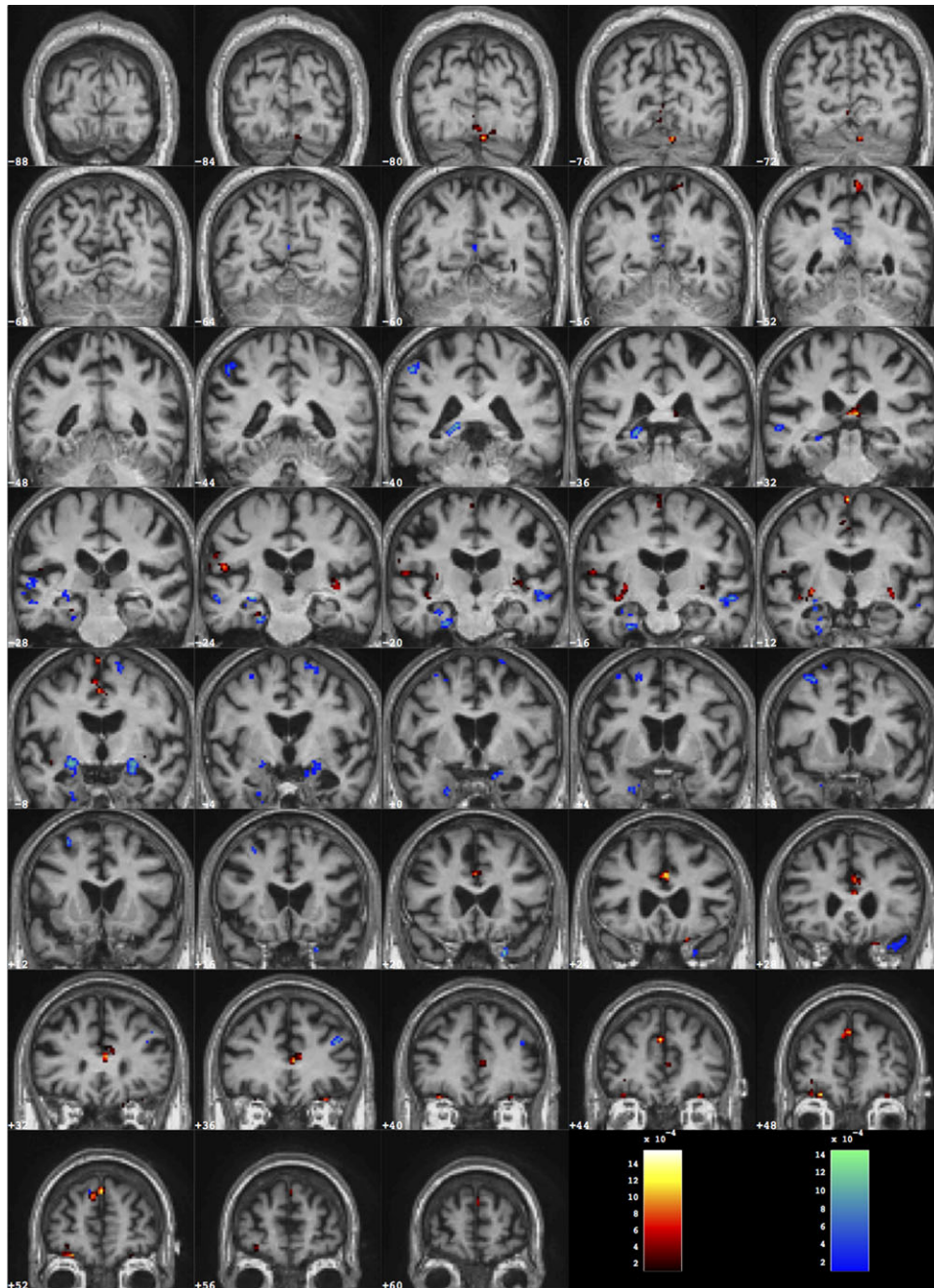
Fig 5. Average discrimination map obtained by training on the AD/CTRL data with the 6000 top ranking SVM weights w_i. The corresponding discrimination average performance is area under the receiver operating characteristic curve (AUC) = (87.1 ± .6)% in 20f-CV repeated ten times on AD/CTRL data, and AUC = (70.7 ± .9)% in independent validation on the MCI-C/MCI-NC cohort.

Correlation of Test Margin with MMSE

We analyzed the correlation of the SVM test margin with the MMSE score for the AD/CTRL and the MCI-C/MCI-NC samples. The test margin has been computed for each subject at the operative point corresponding to 6,000 retained voxels during the 10 repetitions of the SVM-RFE procedure. We found the average Spearman's rank correlation coefficient as: $\rho = .386 \pm .008$ $(P < 10^{-3})$ for the AD/CTRL sample and $\rho = .189 \pm .008$ $(P = 10^{-3})$ for the MCI sample. Figure 6 shows

the scatter plots between the SVM test margins and the MMSE score for both data samples obtained in one of 10 repetitions of the SVM classification. A small noise term has been added to the MMSE score of each subject to make the distributions more visible.

Conclusions and Discussion

We presented in this paper a comparison among three different implementations of SVM classifiers complemented by

Table 3. Anatomical Structures Represented in the Discrimination Maps Obtained with the SVM-RFE Procedures Trained on Alzheimer's Disease/Control Voxel-Based Morphometry and Mild Cognitive Impairment data. These Regions are Visible in Figure 5.

Cluster Extent (Voxels)	MNI Coordinates			Talairach Coordinates			Brain Region Localization			
328	−25	−11	−31	−24	−9	−25	LC	Limbic lobe	Uncus	BA 28
156	32	24	−32	29	23	−22	RC	Frontal lobe	Inferior frontal gyrus	BA 47
212	5	−78	−20	4	−73	−21	RCl	Posterior lobe	Declive	
141	26	36	−24	23	34	−14	RC	Frontal lobe	Middle frontal gyrus	BA 11
224	22	−5	−16	20	−5	−10	RC	Limbic lobe	Parahippocampal gyrus	Amygdala
179	−29	48	−17	−28	45	−7	LC	Frontal lobe	Middle frontal gyrus	BA 11
159	55	−17	−8	50	−17	−4	RC	Temporal lobe	Superior temporal gyrus	BA 22
148	−59	−28	−2	−56	−27	−1	LC	Temporal lobe	Middle temporal gyrus	BA 21
168	−35	−16	−4	−33	−16	−1	LC	Sub-lobar	Lentiform nucleus	Putamen
227	−24	−34	−4	−23	−33	−3	LC	Limbic lobe	Parahippocampal gyrus	BA 27
148	37	−19	2	33	−20	5	RC	Sub-lobar	Claustrum	
392	4	30	19	3	25	24	RC	Limbic lobe	Anterior cingulate	BA 24
116	6	−33	13	4	−34	13	RC	Sub-lobar	Thalamus	Pulvinar
175	−58	−20	17	−55	−22	17	LC	Parietal lobe	Postcentral gyrus	BA 40
186	−5	−55	24	−6	−55	21	LC	Limbic lobe	Posterior cingulate	BA 31
201	−3	50	30	−4	42	35	LC	Frontal lobe	Medial frontal gyrus	BA 6
103	40	36	26	36	29	31	RC	Frontal lobe	Middle frontal gyrus	BA 9
107	−51	−41	50	−49	−44	44	LC	Parietal lobe	Inferior parietal lobule	BA 40
100	−3	−10	52	−4	−16	50	LC	Frontal lobe	Medial frontal gyrus	BA 6
204	−29	7	59	−29	0	57	LC	Frontal lobe	Middle frontal gyrus	BA 6
103	8	−53	68	5	−57	60	RC	Parietal lobe	Precuneus	BA 7
107	18	−4	69	15	−12	66	RC	Frontal lobe	Superior frontal gyrus	BA 6
100	−3	−13	72	−5	−20	67	LC	Frontal lobe	Medial frontal gyrus	BA 6

LC = left cerebrum; RC = right cerebrum; RCL = right cerebellum; BA = Brodmann area; MNI = Montreal Neurological Institute.
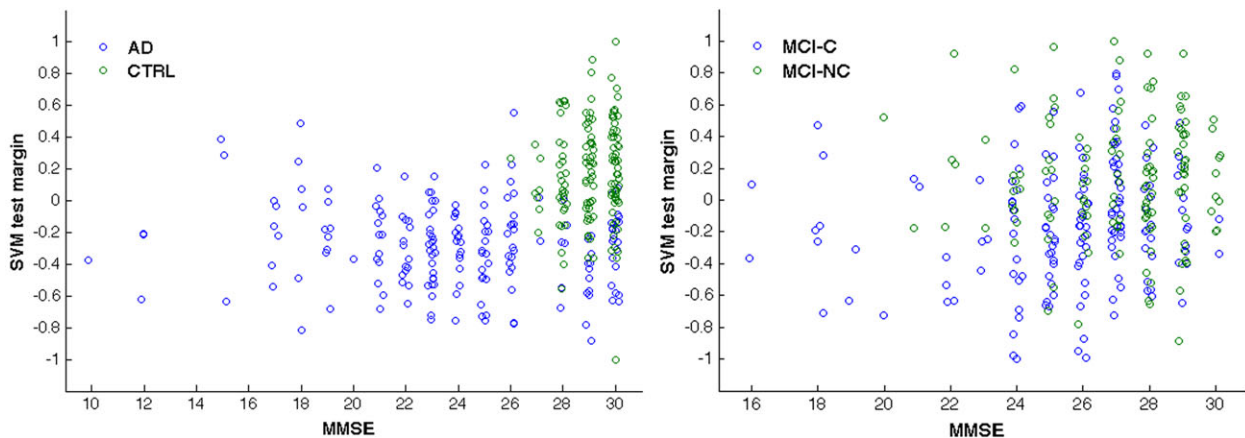

Fig 6. Scatter plots reporting the MMSE score and SVM test margin obtained with SVM trained on AD/CTRL data: AD vs. CTRL (left) and validation on MCI-C versus MCI-NC sample (right). A small noise term has been added to the MMSE score of each subject to make the distributions more visible. The results obtained in one representative out of ten repetitions of the SVM classification are shown.

the RFE feature-selection method and applied to the following data: (1) the whole-GM segmented out of brain MRI; (2) the ROIs selected by *t*-test filtering, ie, those encoding statistically significant between-group differences in AD versus CTRL comparison; and (3) the hippocampus and parahippocampal gyrus ROI parceled according to the LPBA40.[29]

The GM data of either the whole brain (whole-GM method) or the chosen ROIs (VBM-ROI and LONI-ROI methods) have been classified by SVMs, implementing the RFE technique in order to reduce the amount of input data and localize the relevant image information. The global and the local SVM-based techniques are both able to provide single-subject classification,

ie, a prediction on the possible conversion of MCI subjects into AD.

The classification of the whole-GM voxels leads to better performance in terms of AUC with respect to the VBM-ROIs and LONI ROIs classification accuracy on the AD/CTRL sample and especially in the prediction of MCI conversion.

This result is confirmed even when the RFE is applied. In this case the whole-GM approach demonstrated to achieve the best accuracy in MCI-C/MCI-NC separation, ie, AUC = (70.9 ± .9)% with 8,000 retained voxels. Moreover, the AUC values obtained in the prediction of the MCI outcome with the

Table 4. Performance of Recently Developed Machine-Learning Approaches to Alzheimer's Disease Diagnosis

Authors	Publ. Year	Total Number of Cases	AD	CTRL	MCI-C	MCI-NC	Classification	Sens. (%)	Spec. (%)	Accur. (%)	AUC (%)
Casanova et al[22]	2011	98	49	49	–	–	AD/CTRL	82.9	90	85.7	–
Davatzikos et al[26]	2011	356	54	63	69	170	AD/CTRL	–	–	–	73.4
							MCI-C/MCI-NC	–	–	–	66
Hinrichs et al[24]	2011	233	48	66	38	81	AD/CTRL	86.7	96.6	92.4	97.7
							MCI-C/MCI-NC	–	–	–	76.7
Cuingnet et al[20]	2011	509	137	118	106	148	AD/CTRL	81	95	–	–
							MCI-C/MCI-NC	nbtc	nbtc	–	–
Chincarini et al[19]	2011	635	144	189	136	166	AD/CTRL	89	94	–	97
							MCI-C/MCI-NC	72	75	–	74
Cui et al[23]	2011	350	96	111	56	87	MCI-C/MCI-NC	96.4	48.3	67.1	79.6
Liu et al[25]	2012	783	198	229	225	131	AD/CTRL	–	–	90.8	94.7
Cho et al[15]	2012	491	128	160	72	131	AD/CTRL	82	93	–	–
							MCI-C/MCI-NC	63	76	–	–
This paper	2013	635	144	189	136	166	AD/CTRL	80	80	80	87.1 ± .6
							MCI-C/MCI-NC	70	62	66	70.7 ± .9

AD = Alzheimer's disease; CTRL = healthy controls; MCI-C = mild cognitive impairment converted to AD; MCI-NC = MCI nonconverted to AD; AUC = area under the receiver operating characteristic curve; nbtc = no better than chance (as stated by the authors).

whole-GM approach with SVM-RFE outperform the corresponding AUC values obtained with the ROI-based methods for a large range of voxels retained in the SVM-RFE procedure.

The data-driven feature selection operated by the RFE procedure leads to improved performance especially in the prediction of the MCI outcome with the whole-GM approach. It has to be noticed that in this case the data driving the feature selection belong to the AD/CTRL sample, whereas the classification performance improvement refers to the MCI cohort, which is used in this analysis only as validation set.

Despite the classification performance obtained on the MCI population (AUC = 70.7%, sensitivity of 70% and specificity of 62%, and accuracy of 66%) being comparable to values found in recent papers,[15,19,20,22–26] as reported in Table 4, it cannot be considered fully adequate to set up an MRI-based automated tool for the early diagnosis of the AD. A direct comparison is possible with the classification performance obtained in the study by Chincarini et al,[19] which was conducted on the same dataset. The classification accuracy obtained in the present analysis are not as high as those achieved by Chincarini et al,[19] especially in the AD/CTRL separation. This is related to the fact that in the present study we did not carry out any strong optimization of the classification methods, since its goal is the comparison between whole-GM versus preselected ROI classification carried out by using quite straightforward and easily accessible methods of MRI data analysis (eg, VBM preprocessing with SPM tools and SVM analysis with available software packages).

The main result is that the data-driven whole-GM approach based on SVMs is able to find subtle relationships among different brain regions and thus achieve better classification performance in the MCI conversion prediction with respect to decisional systems based on the analysis of preselected ROIs. Cuingnet et al[20] and Chu et al[21] have recently conducted similar comparative analyses on ADNI data. In particular, Cuingnet et al[20] compared 10 different methods and concluded that whole-brain methods are the most powerful in the AD/CTRL classification. However, the authors stated that no classifier ob-

tained significantly better results than chance in the MCI/AD conversion prediction. They also found out that feature selection methods did not improve the classification performance. Chu et al[21] confirmed that data-driven feature selection methods do not perform better than whole-brain approaches (if the dataset is sufficiently populated) and highlighted that only when prior knowledge is used to select ROI (hippocampus and parahippocampal gyrus) it is possible to outperform the whole-brain results. Instead, our work shows that the LONI-ROI approach performs as well as the whole-GM method in AD versus CTRL separation, but it is outperformed by the whole-GM method in the MCI conversion prediction.

Contrary to the results by Cuingnet et al[20] and Chu et al,[21] we found that the data-driven SVM-RFE technique applied to whole-GM data leads to a slight improvement of the classification performance in the MCI-C/MCI-NC discrimination (data considered only as validation set), with AUC = (70.7 ± .9)%. In addition, in our analysis the whole-GM technique, with and without RFE, outperforms the ROI-based ones.

In a recent study by Adaszewski et al[28] the feature selection was shown to improve the classification accuracy at early MCI stages, whereas at a later stage the whole-brain methods are superior. Our analysis on a dataset of subjects at a fixed time point supports the results obtained by Adaszewski et al[28] on a longitudinal dataset of subjects; the feature selection SVM-RFE improves the MCI classification accuracy, whereas it is not necessary to implement it in the AD/CTRL separation, as the performance is maximized by the whole-GM classification.

It also shows that the selected features/voxels in the SVM-RFE method belong to highly distributed clusters in the brain, in agreement with the work of Chu et al.[21] The pattern of highly discriminant voxels that maximizes the predictive power of AD conversion on the MCI population is not confined in the Limbic and Temporal lobes, involving instead a more extended and complex circuit of GM regions.

In conclusion, the analysis reported in this paper shows that higher accuracy in the prediction of MCI conversion to AD can be achieved if the brain is considered as a whole. Several studies

based on whole-brain or multiple-ROI analyses[47–49] support the atrophy of the whole brain as relevant AD biomarker, due to its high capability to differentiate AD from CTRL subjects and to track the disease progression.

References

1. Black SE. The search for diagnostic and progression markers in AD: so near but still too far? *Neurology* 1999;52:1533-1534.
2. Scheltens P, Leys D, Barkhof F, 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 Psychiat* 1992;55:967-972.
3. Karas GB, Scheltens P, et al. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 2004;23(2):708-716.
4. Sluimer JD, van der Flier WM, Karas GB, et al. Whole-brain atrophy rate and cognitive decline: longitudinal MR study of memory clinic patients. *Radiology* 2008;248(2):590-598.
5. Sluimer JD, van der Flier WM, Karas GB, et al. Accelerating regional atrophy rates in the progression from normal aging to Alzheimer's disease. *Eur Radiol* 2009;19(12):2826-2833.
6. Risacher SL, Shen L, West JD, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI). Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiol Aging* 2010;31(8):1401-1418.
7. Chincarini A, Bosco P, Gemme G, et al. Alzheimer's disease markers from structural MRI and FDG-PET brain images. *Eur Phys J Plus* 2012;127(11):135.
8. Ewers M, Walsh C, Trojanowski JQ. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* 2012;33(7):1203-1214.
9. Shaffer JL, Petrella JR, et al. Alzheimer's Disease Neuroimaging Initiative. Predicting cognitive decline in subjects at risk for Alzheimer

10. Klöppel S, Abdulkadir A, et al. Diagnostic neuroimaging across diseases. *Neuroimage* 201261(2):457-463.
11. Orrù G, Pettersson-Yeo W, Marquand AF, et al. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* 2012;36(4):1140-1152.
12. Lao Z, Shen D, Xue Z, Karacali B, et al. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage* 2004;21(1):46-57.
13. Petersen C. Current concepts in mild cognitive impairment. *Arch Neurol* 2001;58(12):1985-1992.
14. Teipel SJ, Born C, Ewers M, et al. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *Neuroimage* 2007;38(1):13-24.
15. Cho Y, Seong JK, Jeong Y, et al. Alzheimer's Disease Neuroimaging Initiative. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage* 2012;59(3):2217-2230.
16. Klöppel S, Stonnington CM, Chu C, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain* 2008;131:681-689.
17. Fan Y, Batmanghelich N, Clark CM, et al. Alzheimer's Disease Neuroimaging Initiative. Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 2008;39(4):1731-1743.
18. Misra C, Fan Y, Davatzikos C. 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* 2009;44(4):1415-1422.
19. Chincarini A, Bosco P, Calvini P. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. *Neuroimage* 2011;58(2):469-480.
20. Cuingnet R, Gerardin E, Tessieras J, et al. Alzheimer's Disease Neuroimaging Initiative. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 2011;56(2):766-781.
21. Chu C, Hsu AL, Chou KH, et al. Alzheimer's Disease Neuroimaging Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 2012;60(1):59-70.
22. Casanova R, Whitlow CT, Wagner B, et al. High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. *Front Neuroinform* 2011;5:22.
23. Cui Y, Liu B, Luo S, et al. Alzheimer's Disease Neuroimaging Initiative. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS One* 2011;6(7):e21896.
24. Hinrichs C, Singh V, Xu G, et al. Alzheimers Disease Neuroimaging Initiative. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage* 2011;55(2):574-589.
25. Liu M, Zhang D, Shen D. Alzheimer's Disease Neuroimaging Initiative. Ensemble sparse classification of Alzheimer's disease. *Neuroimage* 2012;60(2):1106-1116.
26. Davatzikos C, Bhatt P, Shaw LM, et al. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging* 2011;32(12):2322.e19-2322.e27.
27. Weiner MW, Veitch DP, Aisen PS. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement* 2012;8(1 Suppl):S1-S68.
28. Adaszewski S, Dukart J, Kherif F, et al. Alzheimer's Disease Neuroimaging Initiative. How early can we predict Alzheimer's disease

9. disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology* 2013;266(2):583-591.

using computational anatomy? *Neurobiol Aging* 2013;34(12):2815-2826.

29. Shattuck DW, Mirza M, Adisetiyo V, et al. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 2008;39:1064-1080.

30. Guyon I, Weston J, Barnhill S, et al. Gene selection for cancer classification using support vector machines. *Machine Lear* 2002;46:389-422.

31. Green DM, Swets JA. *Signal Detection Theory and Psychophysics.* New York: John Wiley & Sons, Inc., 1966.

32. Mazziotta JC, Toga AW, Evans A, et al. A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* 1996;2:89-101.

33. Ashburner J, Friston KJ. Voxel-based morphometry–the methods. *Neuroimage* 2000;11(6 Pt 1):805-821.

34. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage* 2007;38(1):95-113.

35. Yassa MA, Stark CE. A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *Neuroimage* 2009;44(2):319-327.

36. Ashburner J, Friston KJ. Unified segmentation. *Neuroimage* 2005;26(3):839-851.

37. Ecker C, Rocha-Rego V, Johnston P, et al. MRC AIMS Consortium. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. *Neuroimage* 2010;49(1):44-56.

38. Vapnik V. *The Nature of Statistical Learning Theory.* Berlin: Springer-Verlag ed., 1995.

39. Joachims T. *Making Large-Scale SVM Learning Practical.* In: B Schölkopf, C Burges and A. Smola, eds. Cambridge, MA: MIT Press, 1999.

40. Joachims T. *Learning to Classify Text Using Support Vector Machines.* Kluwer: Springer Science + Business Media, LLC, 2002.

41. Metz CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol* 2006;3(6):413-422.

42. De Martino F, Valente G, Staeren N, et al. Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns. *Neuroimage* 2008;43(1):44-58.

43. Kendall MG. *Rank Correlation Methods.* London: Griffin, 1970.

44. Talairach J, Tournoux P. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System–An Approach to Cerebral Imaging.* New York: Thieme Medical Publishers, 1988.

45. Lancaster JL, Rainey LH, Summerlin CS, et al. Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Hum Brain Mapp* 1997;5:238-242.

46. Lancaster JL, Woldorff MG, Parsons LM, et al. Automated talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 2000;10(3):120-131.

47. Evans MC, Barnes J, Nielsen C, et al. Alzheimer's Disease Neuroimaging Initiative. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. *Eur Radiol* 2010;20(3):674-682.

48. Stonnington CM, Chu C, Klöppel S, et al. Alzheimer Disease Neuroimaging Initiative. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage* 2010;51(4):1405-1413.

49. Vemuri P, Wiste HJ, Weigand SD, et al. Alzheimer's Disease Neuroimaging Initiative. MRI and CSF biomarkers in normal, MCI, and AD subjects: diagnostic discrimination and cognitive correlations. *Neurology* 2009;73(4):287-293.