Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI

Article *in* Neuroradiology · November 2008

DOI: 10.1007/s00234-008-0463-x · Source: PubMed

9 authors, including:

Benoît Magnin
Centre Hospitalier Universitaire de Clermont-Ferrand
28 PUBLICATIONS **637** CITATIONS
SEE PROFILE

Lilia Mesrob
French Institute of Health and Medical Research
31 PUBLICATIONS **519** CITATIONS
SEE PROFILE

Mélanie Pélégrini-Issac
Sorbonne Université
118 PUBLICATIONS **4,298** CITATIONS
SEE PROFILE

Olivier Colliot
L'Institut du Cerveau et de la Moelle Épinière
285 PUBLICATIONS **6,095** CITATIONS
SEE PROFILE

Neuroradiology
DOI 10.1007/s00234-008-0463-x

DIAGNOSTIC NEURORADIOLOGY

Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI

Benoît Magnin · Lilia Mesrob · Serge Kinkingnéhun · Mélanie Pélégrini-Issac · Olivier Colliot · Marie Sarazin · Bruno Dubois · Stéphane Lehéricy · Habib Benali

Received: 13 May 2008 / Accepted: 15 September 2008

Abstract
Purpose We present and evaluate a new automated method based on support vector machine (SVM) classification of whole-brain anatomical magnetic resonance imaging to discriminate between patients with Alzheimer's disease (AD) and elderly control subjects.
Materials and methods We studied 16 patients with AD [mean age±standard deviation (SD)=74.1±5.2 years, mini-mental score examination (MMSE)=23.1±2.9] and 22 elderly controls (72.3±5.0 years, MMSE=28.5±1.3). Three-dimensional T1-weighted MR images of each subject were automatically parcellated into regions of interest (ROIs). Based upon the characteristics of gray matter extracted from each ROI, we used an SVM algorithm to classify the subjects and statistical procedures based on bootstrap resampling to ensure the robustness of the results.
Results We obtained 94.5% mean correct classification for AD and control subjects (mean specificity, 96.6%; mean sensitivity, 91.5%).
Conclusions Our method has the potential in distinguishing patients with AD from elderly controls and therefore may help in the early diagnosis of AD.

Keywords Alzheimer's disease · Diagnosis · Magnetic resonance image · Support vector machine · Sensitivity · Specificity

B. Magnin · M. Pélégrini-Issac (✉) · H. Benali
UMR-S 678, Inserm,
Paris 75013, France
e-mail: Melanie.Pelegrini@imed.jussieu.fr

B. Magnin · L. Mesrob · S. Kinkingnéhun · M. Sarazin · B. Dubois · S. Lehéricy
UMR-S 610, Inserm,
Paris 75013, France

B. Magnin · L. Mesrob · S. Kinkingnéhun · M. Pélégrini-Issac · M. Sarazin · B. Dubois · S. Lehéricy · H. Benali
Faculté de médecine Pitié-Salpêtrière,
UMPC Univ Paris 06,
Paris 75013, France

B. Magnin · L. Mesrob · S. Kinkingnéhun · M. Pélégrini-Issac · O. Colliot · M. Sarazin · B. Dubois · S. Lehéricy · H. Benali
IFR 49,
Gif-sur-Yvette 91191, France

S. Kinkingnéhun
e(ye)BRAIN,
94400 Vitry-sur-Seine, France

O. Colliot
UPR 640 LENA, CNRS,
Paris 75013, France

M. Sarazin · B. Dubois
Department of Neurology,
Pitié-Salpêtrière Hospital,
Paris 75013, France

S. Lehéricy
Center for NeuroImaging Research–CENIR,
UMPC Univ Paris 06,
Paris 75013, France

S. Lehéricy
Department of Neuroradiology,
Pitié-Salpêtrière Hospital,
Paris 75013, France

H. Benali
UNF/CRIUGM, Université de Montréal,
Montréal, QC H3W 1W5, Canada

Introduction

Dementia is a growing health problem, and Alzheimer's disease (AD) is the leading cause of dementia in the elderly accounting for 50–60% of all cases [1–3]. AD patients benefit from early cholinesterase inhibitors [4, 5] and would consequently gain from early and accurate diagnosis of AD. In recent years, the early clinical signs of AD have been extensively investigated, leading to the concept of amnestic mild cognitive impairment (MCI) [6, 7].

Besides neuropsychological examination, structural imaging is increasingly used to support the diagnosis of AD. Pathological studies have shown that neurodegeneration in AD begins in the medial temporal lobe, successively affecting the entorhinal cortex, the hippocampus, the limbic system, then extending toward neocortical areas [8]. Therefore, there has been considerable effort put on the detection of medial temporal lobe atrophy (MTA), and particularly in the hippocampus, the entorhinal cortex, and the amygdala [9, 10]. MTA has been evaluated using visual rating scales, linear or volumetric measurements, and voxel-based approaches. Overall, the sensitivity and specificity of hippocampus measurements for distinguishing AD patients from healthy aged subjects have been evaluated to range from 80% to 95% [11–15]. Evaluation of MTA for diagnostic purposes in AD has limitations however. MTA measurements are much less efficient in the pre-dementia conditions such as amnestic MCI [11, 16–19]. Atrophy in early stages of AD is not confined to the hippocampus or the entorhinal cortex. Other areas are affected in AD patients and MCI patients as well [20]. Whole-brain methods for characterizing brain atrophy may therefore be more efficient in differentiating AD and MCI patients who will evolve toward AD from healthy control subjects. To be considered useful by clinicians, such methods must provide an individual predictive diagnosis.

Recent classification methods have been developed that allow an individual class prediction. Among them, machine-learning techniques have been proposed to distinguish magnetic resonance (MR) images from two groups of subjects (e.g., patients vs. healthy subjects) [21, 22]. All these techniques require a training population, i.e., well-characterized subjects (for instance healthy subjects and patients with known diagnosis), in order to categorize new subjects, who belong to the so-called test population, into one of the classes the subjects of the training population belong to. They also require one or more feature parameters to differentiate the two groups under study.

In particular, support vector machines (SVM, [23]) have recently been used to help distinguish AD subjects from elderly control subjects using anatomical MR imaging (MRI) [24–26]. Classification methods have also been recently applied to the classification of MCI subjects compared to control subjects [27, 28] or to help differentiate AD from frontotemporal lobar degeneration [25, 29]. Even though feature parameters can be determined from the whole brain [24, 25], these parameters may have no relevant physiopathological interpretation [25], or only a selected set of most discriminating voxels or regions are eventually used to classify the subjects [24].

In this paper, our purpose is to individually classify AD patients and healthy elderly control subjects by using a whole-brain MR image analysis. We use a histogram analysis of the MR images to extract feature parameters. Thereby, we focus on characteristics of the distribution of the gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), which intuitively makes sense when dealing with neurodegenerative diseases in general and AD in particular.

More specifically, the method parcellates the subjects' brain MRI into regions of interest (ROIs). For each ROI, a histogram analysis of the distribution of the intensity in the voxels is performed, which identifies the respective contributions of GM, WM, and CSF in the region. From this analysis, a parameter is extracted that characterizes the subject and has a physiological meaning, since it represents the relative weight of GM compared to WM and CSF. Finally, based upon this parameter estimated in all the ROIs of the whole brain, the method classifies the subjects with an SVM algorithm. Most importantly, robust classification results are ensured even for small groups, since we use statistical procedures based on bootstrap resampling.

Materials and methods

Patients and subjects

The study was approved by a regional ethics committee. Sixteen patients with AD and 22 healthy aged control subjects (CS) were recruited. All patients and control subjects signed an informed consent form after the nature of the procedures was fully explained. Sixteen patients with AD were recruited from 1998 to 2002. Inclusion criteria were the following: (1) fulfillment of the National Institute of Neurological and Communication Disorders and Stroke/Alzheimer Disease and Related Disorders Association criteria for probable AD [30]; (2) Clinical Dementia Rating scale (CDR) ≥ 0.5 [31] (all patients but two had CDR=1, two patients had CDR=0.5); (3) live in the community; (4) no clinical or neuroimaging evidence of focal lesions including brain tumor, subdural haematoma, and central nervous system infection; (5) no cortical or subcortical vascular lesions on MRI visible on the structural image; (6) no medical conditions that may interfere with cognitive performance or follow-up; and (7) no history of depression.

All but two patients were treated with inhibitors of acetylcholinesterase. All patients underwent the same procedure. Twenty-two control subjects matched for age were also included. All control subjects underwent a neuropsychological assessment including the Mini Mental State Examination (MMSE) [32] and verbal fluency tests [33]. All control subjects had an MRI examination using the same scanner and the same procedure as AD patients. Clinical and demographic data for the included subjects are reported in Table 1.

Onset dating and features

Disease duration was estimated based on the interview of the patient and the informants. Disease onset was also estimated using a standardized onset interview technique that systematically queries the earliest date of manifestation of specific disease symptoms and the latest point at which these symptoms were not present [34].

Neurological examination

Clinical neurological examination included selected items from the Unified Parkinson's Disease Rating Scale [35] in order to rate extrapyramidal signs. Record of medical events and current treatment was also reported. In addition, a global examination was performed by the CDR [31].

Neuropsychological assessment

All AD patients were tested at inclusion using a standardized neuropsychological battery. Cognitive tests were selected to assess a broad range of cognitive abilities commonly affected in AD. The total duration of the cognitive examination was approximately 2 h and included the MMSE [32] and modified MMSE [36] for global cognitive efficiency, free and cued recall test with control of encoding for verbal episodic memory [37], semantic memory tests including verbal and non-verbal modalities assessment [38], Benton Visual Retention test for visual memory [39], limb praxis [40], pictures copy of Rosen [41], and pictures copy from the modified MMSE for constructive apraxia, pictures naming [42], and Boston Diagnosis Aphasia Examination [43] to evaluate the phonological, morphological, syntactic, and semantic aspects in the processes of language comprehension and language production, verbal fluency tests [33], the Frontal Assessment Battery, which explores conceptualization, mental flexibility, motor programming, sensitivity to interference, inhibitory control, and environmental autonomy [44] for executive functions, and the Cambridge neuropsychological tests automated battery for attentional processing and ability to shift cognitive set [45, 46].

Table 1 Demographic characteristics of the study population

	Control subjects (n_{CS}=22)	Patients (n_{AD}=16)
Male/female	4/18	5/11
Age (years)	72.3±5.0[a]	74.1±5.2[a]
MMSE	28.5±1.3[a]	23.1±2.9[a]
Modified MMSE	–	47.3±5.4[a]
Disease duration (years)	–	4.78±2.04[a]

[a] Mean±standard deviation

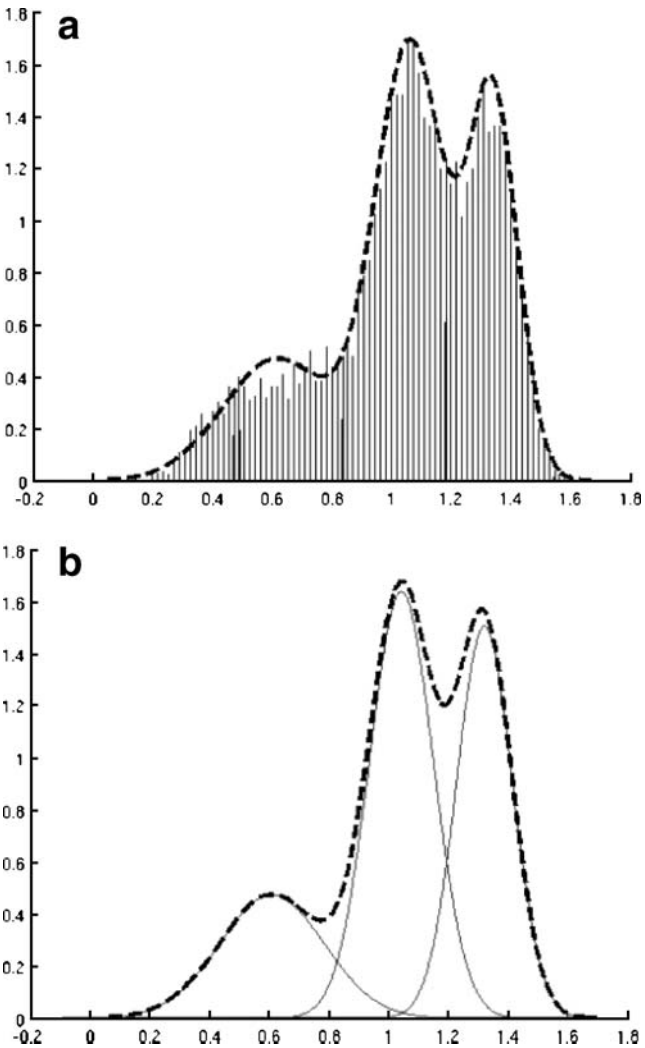

Fig. 1 Histogram modeling and separation. (**a**) Histogram of intensity levels in one ROI of a representative control subject and its modeling (*dashed line*) by a mixture of normal distributions. (**b**) Three normal distributions (*solid lines*) are evidenced in the histogram using the EM algorithm. The *x*-axis represents the intensity. The *y*-axis indicates the probability that voxels with a given level in the ROI belong to the distributions

MR images

MRI examination for all subjects was performed using a 1.5-T system (GE Medical Systems, Milwaukee, WI, USA). Anatomical high-resolution scans were acquired using a coronal three-dimensional fast spoiled gradient echo sequence (repetition time/echo time/flip angle, 23 ms/5 ms/35°, 256×192 matrix; voxel size, 0.86×0.86×1.5 mm^3) using a standard head coil.

Brain parcellation

In order to parcellate the brain into ROIs, we used an anatomically labelled template of the brain previously developed by Tzourio-Mazoyer et al. [47]. This template is based on sulci delineation and includes the whole-brain gray matter, but some white matter areas are not parcellated. We also created a mask to exclude voxels belonging to the skull by using a segmentation routine implemented in the SPM2 software [48] (Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London UK, http://www.fil.ion.ucl.ac.uk/spm/) in conjunction with Matlab™ version 7.0.1 (Mathworks, MA, USA), to obtain probability masks for gray matter, white matter, and CSF. By construction, the labelled template and the mask are both registered to the MNI standard space, which is a close analogue of the Talairach space [49] and was developed by the Montreal Neurological Institute (ICBM, NIH P20 project, Principal Investigator John Mazziotta). Besides, still using the SPM2 software, we computed a nonlinear transformation from the T_1-weighted image of each individual subject to the T_1-weighted MNI template. This nonlinear transformation was computed using an affine registration with 12 degrees of freedom, which seeks to minimize the least-squares distance between the individual T1-weighted image and the template, followed by 16 nonlinear iterations performed using a discrete cosine transform to remove global nonlinear differences between the two images [50]. Then, using the inverse of this nonlinear transformation and the Deformation Toolbox of SPM2 [51], we denormalized the labeled template and the mask from the MNI space to the space of the native individual image. This denormalization process provided a parcellation of the native image into 116 ROIs and a mask excluding the voxels from the skull. We excluded the 26 ROIs of the cerebellum, since some of them were partially cut off by the denormalization process.

Extraction of MRI feature parameter

In each ROI for each subject, we created the histogram of the intensity levels in the voxels. In most regions and all subjects, the histogram showed three modes (Fig. 1a). We modeled the histogram by the distribution

$$\alpha_1 N(\mu_{1,}, \sigma_1^2) + \alpha_2 N(\mu_2, \sigma_2^2) + \alpha_3 N(\mu_3, \sigma_3^2),$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $N(\mu, \sigma^2)$ is the normal distribution of mean μ and of variance σ^2. We used the expectation and maximization (EM) algorithm [52] to separate the mixture of normal distributions (Fig. 1b). The three modes obtained by the EM algorithm correspond to CSF, GM, and WM, respectively (Fig. 2), which proves consistent with the underlying anatomy.

The separation of the modes provided us with nine parameters describing each ROI (μ_i,σ_i, and α_i,i=1,…,3). As gray matter modifications are well known in AD, we focused in our study on the distribution of the gray matter. We therefore chose one feature parameter describing each ROI for each subject: α_2, which is the relative weight of GM compared to WM and CSF.

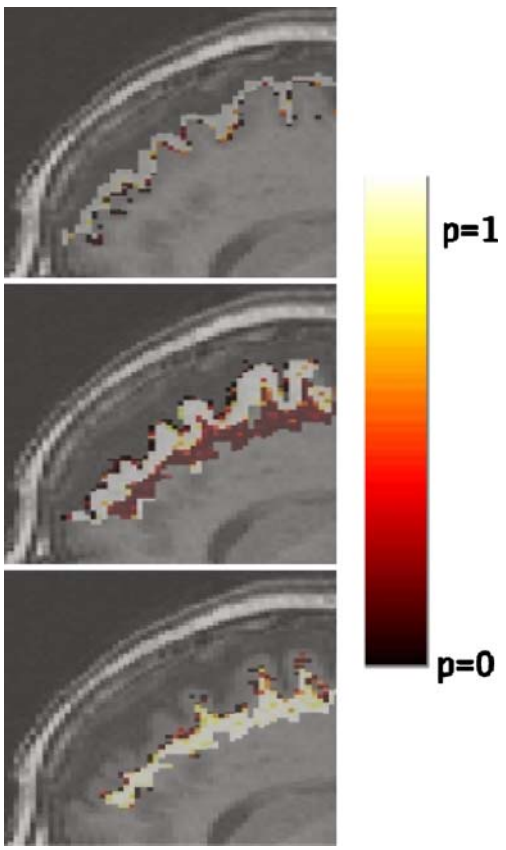

Fig. 2 Localization of voxels belonging to the three modes. *Top* Voxels included in the mode presenting the lowest intensity; *middle* voxels included in the mode presenting the intermediate intensity; *bottom* voxels included in the mode presenting the highest intensity. The *color bar* indicates the probability that voxels belong to a given mode

SVM classifier algorithm

The SVM is a learning machine for two-class classification problems [23]. SVM conceptually implements the idea that vectors are nonlinearly mapped to a very high dimension feature space. In this feature space, a linear separation surface is created to separate the training data by minimizing the margin between the vectors of the two classes. The training ends with the definition of a decision surface that divides the space into two subspaces. Each subspace corresponds to one class of the training data. Once the training is completed, the test data are mapped to the feature space. A class is then assigned to those data depending on which subspace they are mapped to.

Statistical analysis

Discriminating power of the feature parameter

The whole set of 90 ROIs was used to classify the patients. Nevertheless, it was important to evaluate the discriminating power of the feature parameter α_2 and investigate whether and in which brain regions this parameter was indeed able to distinguish AD patients from healthy controls.

Group differences in the value of α_2 in each ROI between AD patients and CS were assessed as follows. For each ROI, we tested whether the distribution of α_2 in CS significantly differed from that in AD patients, by using a two-sample T test. As the number of subjects examined was relatively limited (n_{CS}=22 healthy CS and n_{AD}=16 AD patients), we assumed that this test was not robust enough. We bypassed this difficulty by using a bootstrap method [53], working with the null hypothesis H_0 that there was no difference between the two groups. The initial set of subjects S was divided into S_1 of cardinal n_{CS}, initially containing all the healthy CS, and S_2 of cardinal n_{AD}, initially containing all the AD patients, and the initial value T_0 of the T test was calculated. We then resampled the set S under the hypothesis H_0, creating resampled sets $S_1^{*,1}$ and $S_2^{*,1}$ by drawing with replacement n_{CS} and n_{AD} $s$ubjects from the whole set S. For the ith resampling $S_1^{*,i}$ and $S_2^{*,i}$, we calculated the corresponding value $T^{*,i}$ of the T test. We can perform n resamplings and calculate the percentile corresponding to the initial value T_0 of the T test in the set of values $\{T^{*,i},\, i \in [1,n]\}$, i.e., cardinal $(T^{*,i} \geq T_0)/n$. According to the bootstrap theory [53], this percentile is a good estimate of the p value of the T test if n is large enough. Accordingly, we determined the significance of the T test for each ROI. The significance for all ROIs gives a hint of the discriminating power of α_2 between the AD patients and the CS. To obtain a very precise estimation of those p values, we performed the method with n=400,000 resamplings so that the lowest non-null value that can be obtained is $1/250{,}000 = 2.5 \times 10^{-6}$.

Selection of training and test data using a bootstrap method

We developed a robust procedure for selecting training and test data using a bootstrap method. This was aimed at testing the classification method on a population with known diagnosis. We drew without replacement one subset of each group for test data. The cardinal of those subsets was approximately 25% of the cardinal of the initial groups. The remaining subjects were used as training data to create a separation surface in order to classify the test data. We repeated the drawing of the test data 5,000 times, so that each subject was classified using various combinations of training data. As a result, we obtained a mean rate of

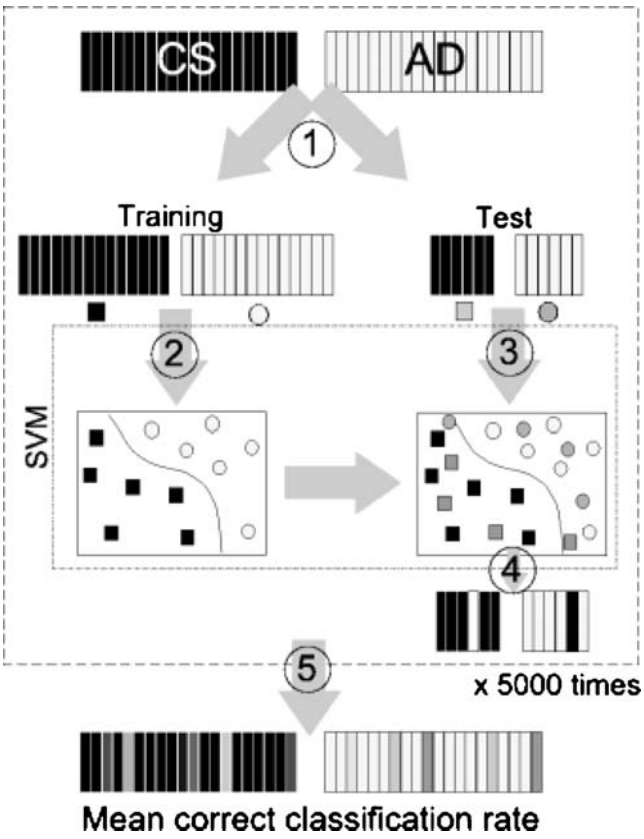

Fig. 3 Method for data selection to perform classification on one population of known diagnosis. *1* Random drawing of subsets from each group. The largest subsets are training data, the smallest test data. *2* SVM transforms the training data to a very high-dimension feature space. In this space, a separation surface is built that divides the space into two subspaces. *3* The test data are mapped to the same high dimension space. *4* Depending on the subspace they are mapped to, a class is then assigned to the test data. *5* The process is repeated 5,000 times. The classification results are averaged, resulting in a mean correct classification rate for each subject

correct classifications for each subject. This method is summarized on Fig. 3. Similarly, we also developed a second classification procedure that can be used to classify new subjects using a population with known diagnosis as the training data.

Choice of SVM parameters

In our method, the two classes were "AD patient" or "CS"; each subject was represented by a vector of the α_2 values in all 90 ROIs. A kernel function needs to be chosen for SVM: It reflects the nonlinear mapping from the input space to the feature space. We chose a commonly used radial basis function of the form $(x,y) \rightarrow K(x,y) = e^{\gamma|x-y|^2}$. We determined the optimal values of two constants: γ, width of the radial basis function, and C, an input parameter for the SVM algorithm, which represents the error/trade-off parameter that adjusts the importance of the separation error in the creation of the separation surface. We proceeded by using a grid search: Using the bootstrap procedure for training and test data selection as described above, we performed classifications for the MRI dataset with (γ,C) varying along a grid. The performance of the classification for a given value of (γ,C) was evaluated by computing the mean accuracy across all subjects of the mean classification rates.

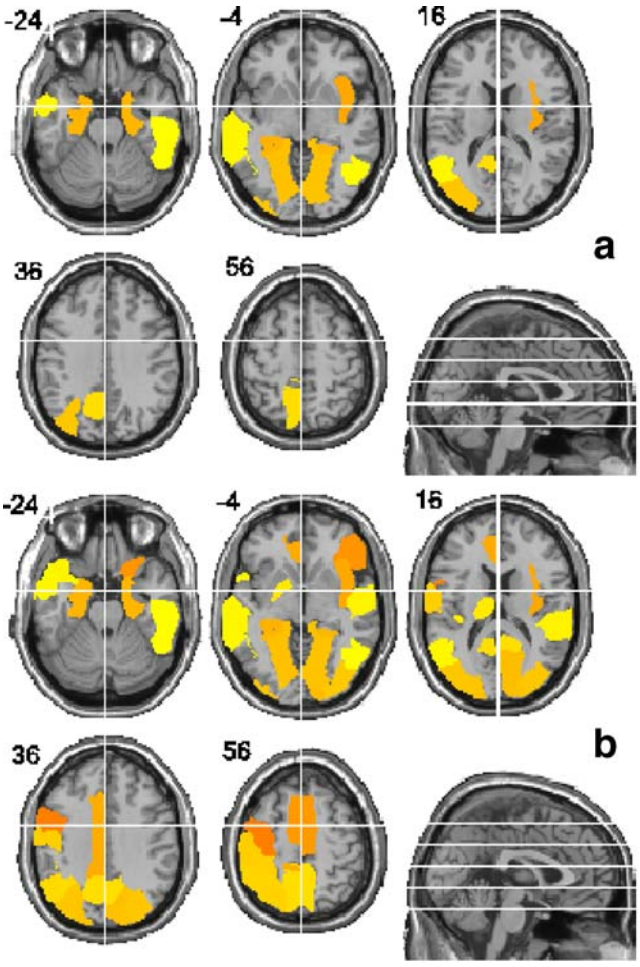

Fig. 4 ROIs that best discriminate patients with Alzheimer's disease and control subjects. **a** ROIs with $0<p<10^{-4}$ and **b** ROIs with $p<10^{-2}$. The ROIs are superimposed on the MNI structural template

Table 2 ROIs that best discriminate patients with Alzheimer's disease and control subjects

ROI	Bootstrap estimate of the p value
Left parahippocampal gyrus	$<2.5\times10^{-6}$
Right parahippocampal gyrus	$<2.5\times10^{-6}$
Left hippocampus	$<2.5\times10^{-6}$
Right lingual gyrus	5×10^{-6}
Left lingual gyrus	5×10^{-6}
Right insula	5×10^{-6}
Left middle occipital gyrus	5×10^{-6}
Left precuneus	3×10^{-5}
Left middle temporal gyrus	3.75×10^{-5}
Left inferior temporal gyrus	4.75×10^{-5}

Results

Most significantly different ROIs

Ten ROIs had a p value less than 10^{-4} and 34 ROIs had a p value less than 10^{-2} (Fig. 4). These small p values indicate that there was a significant difference between AD patients and CS in the characteristics of gray matter in these ROIs as measured by the parameter α_2. The most significant ROIs included regions classically affected in AD, such as the hippocampus or the parahippocampal gyrus. Those ROIs are listed in Table 2.

Classification results

The mean specificity of the method was 96.6%, and the mean sensitivity was 91.5%, with an overall mean accuracy of 94.5%. Most subjects had a very high mean correct classification rate, whereas only few subjects had a low correct classification rate (34/38 subjects with a mean correct classification rate greater than 90%). Classification results for all subjects are shown in Table 3.

Discussion

In this study, we developed an automated whole-brain analysis method that is able to classify patients with Alzheimer's disease from elderly healthy subjects; it is based on structural MRI and works automatically using defined regions of interest.

Feature parameter

We designed a feature parameter that showed a discriminating power between patients with AD and healthy CS, and we used this feature parameter to create a classifier that can predict whether a single subject belongs to the AD or the CS group. As gray matter modifications are well known in AD, it makes sense to focus on the distribution of the gray matter for the choice of the feature parameter and, consequently, on the parameters (μ_2,σ_2 and α_2) estimated by the Gaussian mixture model. All three parameters convey some information that could be used either separately or in combination as representative features for the ROIs in the classification algorithm. In this paper, we chose to use only α_2 for two main reasons. Firstly, α_2 has a physiological meaning in that it is the relative weight of GM compared to WM and CSF. Therefore, it intuitively reflects the modifications that are well known to be induced by AD, namely, a decrease in GM compared to WM and CSF. Secondly, this parameter is more invariant with respect to the acquisition parameters, as opposed to μ_2 and α_2, which are more likely to depend on the contrast in the images. We expect that using α_2 alone would make the comparison of datasets acquired with different sequences (e.g., coming from different clinical centers) more robust.

Classification method

The SVM classifier involves a nonlinear mapping from the input parameter space to the feature space. The nonlinear mapping was performed using a radial basis function. We tried linear kernels first, but the obtained classification rates were lower than those obtained with the radial function.

Bootstrap resampling methods were used both to select test and training data to evaluate the classification accuracy of our method and to estimate the SVM parameters. For this latter purpose, alternative approaches have been proposed in the literature, for instance, cross-validation methods such as leave-one-out [24, 25] or *n*-fold cross-validation [26]. For instance, Vemuri et al. used a technique in which patients and controls were first divided in two groups (training and testing sets), the training group being in turn divided into three subgroups used for feature selection and model optimization and one subgroup used for testing. However, the authors acknowledge that this leads to perform the classification on much smaller groups, and they repeated their method ten times and averaged the results to increase the robustness of their classification.

In our study, since a nonlinear mapping of the parameters is performed, bootstrap is the method of choice, as leave-one-out has been shown to be less robust than bootstrap when applied to nonlinear cases **(**[53], p. 146**)**.

Classification results

The results with the proposed classifier yielded a sensitivity of 91.5% and a specificity of 96.6%, for a 94.5% mean correct classification rate. The MR images of all subjects were of similar quality. However, the SVM classified four subjects with a correct classification rate less than 90%, as shown in Table 3, which may be explained by clinical considerations. Patient 1, who was never classified correctly, had a very particular clinical evolution. At the time the MRI was performed, the scores for this patient were CDR=1 and MMSE=19, which happened to be the lowest MMSE of all the subjects. The initial clinical examination was very suggestive of AD: starting with amnestic symptoms and an initially normal neurological examination (in particular, no extrapyramidal syndrome). Four years after the scan, the patient showed a rapid cognitive decline with myoclonus and an extrapyramidal syndrome combined with posture disorders. This clinical evolution suggests a variant with Lewy bodies of AD. The patient died in 2008, but no autopsy was performed. This very particular clinical evolution may explain why the classification method failed to consider this subject as a typical AD patient and therefore never classified her as an AD patient. Patient 2 was one of the two patients included with CDR=0.5. These two patients had a neuro-

Table 3 Classification results for healthy control subjects and AD patients

Control subject	Correct classification rate (%)	Patient	Correct classification rate (%)
CS 1	68	Patient 1	0
CS 2	75	Patient 2	86
CS 3	90	Patient 3	90
CS 4	94	Patient 4	91
CS 5	100	Patient 5	98
CS 6	100	Patient 6	99
CS 7	100	Patient 7	99
CS 8	100	Patient 8	100
CS 9	100	Patient 9	100
CS 10	100	Patient 10	100
CS 11	100	Patient 11	100
CS 12	100	Patient 12	100
CS 13	100	Patient 13	100
CS 14	100	Patient 14	100
CS 15	100	Patient 15	100
CS 16	100	Patient 16	100
CS 17	100		
CS 18	100		
CS 19	100		
CS 20	100		
CS 21	100		
CS 22	100		

logical follow-up during 8 years, and the diagnosis of dementia was certain according to usual clinical criteria. They were therefore eventually included in the analysis, although their CDR value was 0.5 at the time of inclusion in 1998. Incidentally, the second patient was always classified correctly in our analysis (patient 8 in Table 3). Concerning controls subjects CS 1 and CS 2, the values of the feature parameter α_2 in the left hippocampus and the left parahippocampal gyrus ROIs (in which α_2 significantly discriminates AD patients from controls, see Table 2) were closer to the values found for AD patients than to those found for the controls; this may therefore explain why the algorithm failed to classify these subjects as normal although they had no diagnosis of dementia.

Measurement of medial temporal lobe atrophy

Other classification methods have been proposed in Alzheimer's disease. The estimation of medial temporal lobe atrophy was efficient to discriminate patients with AD from healthy control subjects [11–15]. Overall, studies based on manual hippocampal volumetry reported classification rates between 80% and 95% [11–15]. However, the discriminating power of MTA measurements was lower in pre-dementia conditions such as MCI. Studies based on manual segmentation of the hippocampus reported classification rates of 60–74% for MCI patients [11, 16–19]. Whole-brain methods for characterizing brain atrophy, such as the one developed in our study, may be more efficient in differentiating AD and MCI patients that will convert to AD from healthy controls. Future studies will help determine the efficacy of these classification methods in MCI patients.

MTA measurements were also not successful in differentiating AD from other neurodegenerative dementias. Indeed, MTA has been reported in other dementias including fronto-temporal dementia and dementia with Lewy bodies, although the medial temporal lobe is relatively preserved in the latter pathology [54, 55]. Pathological studies and whole-brain structural imaging methods have suggested that neurodegenerative diseases have a specific pattern of gray matter atrophy [56, 57]. Therefore, whole-brain methods for characterizing brain atrophy may be more appropriate for differentiating AD from other neurodegenerative dementias.

SVM-based methods in Alzheimer's disease classification

Other whole-brain SVM-based classification methods have been proposed recently. Fan et al. [24] used a pattern classification method to obtain a 94.3% correct classification rate between AD patients (mean MMSE, 23.07) and control subjects. Their results were quite comparable to ours; however, a different method was used. In particular, the regions from which the features were extracted were calculated for each training data and therefore varied for each sample, while in the present study, the same regions from a single labeled anatomical template were used for all subjects. The SVM-based classification method proposed by Vemuri et al. [26] was based on MRI features that were different from those used in the present study: The features parameters were initially all the voxels from the segmented MRI and were subsequently reduced to a subset of voxels that best distinguished AD from CS. The studied AD patients were at a more advanced stage than in the present study (the mean MMSE was 20, compared to 23.1 in our group), and the classification was slightly less accurate (the accuracy was 86% when using only MRI data and 89.3% when adding demographic information, compared to 94.5% in our study). Klöppel et al. [25] also proposed a method for classifying AD patients and elderly control subjects using SVM and achieved 95% correct classification on a group of patients with 16.7 mean MMSE. The correct classification rate was up to 89% on a group of patients at an earlier stage of the disease (mean MMSE, 23.5), quite similar to that in the present study; however, their results were slightly less accurate than ours. It should be noted, however, that a thorough comparison between our study and others is not easy since different datasets were used, and the quality of imaging data is obviously an important factor. Nevertheless, on the basis of reported correct classification rates and MMSE values, it can be concluded that the method proposed in this paper performs very well compared to existing classification approaches.

Possible improvements of the proposed method

The separation of the modes in each ROI provided us with nine parameters describing the ROI. The current method involves only one parameter from the GM mode. Inclusion of parameters from the WM mode in the classification procedure did not improve the results. The robustness of this classification method on MR examinations with various acquisition parameters remains to be tested. This would enable one to perform classification on data coming from different clinical centers.

The use of bootstrap and resampling methods ensured a statistical robustness to the procedure. This allowed us to obtain significant results on relatively small MRI datasets. Nevertheless, the results of our study require confirmation in much larger groups of participants.

The classifier was developed using a specific procedure to parcellate the brain into ROIs. We then performed parameter extraction at a regional scale, thus making the method robust to registration imperfections and anatomical variations across individuals. The current parcellation yielded good classification results, but it can be improved.

Indeed, the proposed method extracted one feature parameter for each ROI. Therefore, it cannot take variations of the feature parameter within the ROI into account. Consequently, using a parcellation into smaller ROIs could be sensitive to variations of the feature parameter at smaller scales and thus better fit the atrophy pattern of the disease and improve the classification results. However, it is expected that the smaller the ROIs, the less robust the estimation of the feature parameter, as possible registration errors would likely introduce larger errors when estimating the proportion of gray matter from the histogram analysis and hence yield a decrease in sensitivity. A parcellation at a finer scale should then take this trade-off issue into consideration. Improvement may also include other routines to denormalize the labeled anatomical template to the individual MRI or the use of a different labeled template.

As described previously, the current classifier used parameters extracted from structural MRI. Data from other acquisition modalities can be used jointly to anatomical MRI data. Indeed, features extracted from diffusion tensor imaging, functional MRI, or single photon emission computed tomography could be combined or simply concatenated to the features extracted from anatomical MRI in the vector that describes each subject in the SVM.

Conclusion

We have developed a method that is able to classify automatically patients with early Alzheimer's disease from control subjects. This method has a potential for early diagnosis of Alzheimer's disease. The method will be evaluated in MCI patients and for other neurodegenerative diseases, and its robustness will be assessed in patients with images obtained from different MR scanners with various acquisition parameters.

Conflict of interest statement S. Kinkingnéhun has a financial relationship with e(ye)BRAIN. B. Dubois and H. Benali consult for e(ye)BRAIN.

References

1. Brookmeyer R, Gray S, Kawas C (1998) Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. Am J Public Health 88:1337–1342
2. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes PR, Rimmer E, Scazufca M (2005) Global prevalence of dementia: a Delphi consensus study. Lancet 366:2112–2117 doi:10.1016/S0140-6736(05)67889-0
3. Ramaroson H, Helmer C, Barberger-Gateau P, Letenneur L, Dartigues J (2003) Prevalence of dementia and Alzheimer's disease among subjects aged 75 years or over: updated results of the PAQUID cohort. Rev Neurol (Paris) 159:405–411 (in French)
4. Winblad B, Wimo A (1999) Assessing the societal impact of acetylcholinesterase inhibitor therapies. Alzheimer Dis Assoc Disord 13(Suppl 2):S9–S19 doi:10.1097/00002093-199911002-00003
5. DeKosky ST, Marek K (2003) Looking backward to move forward: early detection of neurodegenerative disorders. Science 302:830–834 doi:10.1126/science.1090349
6. Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. J Intern Med 256:183–194 doi:10.1111/j.1365-2796.2004.01388.x
7. 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 RC (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 doi:10.1111/j.1365-2796.2004.01380.x
8. Braak H, Braak E (1995) Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 16:271–278 (discussion 278–284) doi:10.1016/0197-4580(95)00021-6
9. Bastos Leite AJ, Scheltens P, Barkhof F (2004) Pathological aging of the brain: an overview. Top Magn Reson Imaging 15:369–389 doi:10.1097/01.rmr.0000168070.90113.dc
10. Glodzik-Sobanska L, Rusinek H, Mosconi L, Li Y, Zhan J, de Santi S, Convit A, Rich K, Brys M, de Leon MJ (2005) The role of quantitative structural imaging in the early diagnosis of Alzheimer's disease. Neuroimaging Clin N Am 15:803–826 doi:10.1016/j.nic.2005.09.004
11. Xu Y, Jack CR Jr, O'Brien PC, Kokmen E, Smith GE, Ivnik RJ, Boeve BF, Tangalos RG, Petersen RC (2000) Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. Neurology 54:1760–1767
12. Frisoni GB, Laakso MP, Beltramello A, Geroldi C, Bianchetti A, Soininen H, Trabucchi M (1999) Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease. Neurology 52:91–100
13. Laakso MP, Soininen H, Partanen K, Lehtovirta M, Hallikainen M, Hänninen T, Helkala EL, Vainio P, Riekkinen PJS (1998) MRI of the hippocampus in Alzheimer's disease: sensitivity, specificity, and analysis of the incorrectly classified subjects. Neurobiol Aging 19:23–31 doi:10.1016/S0197-4580(98)00006-2
14. Lehéricy S, Baulac M, Chiras J, Piérot L, Martin N, Pillon B, Deweer B, Dubois B, Marsault C (1994) Amygdalohippocampal MR volume measurements in the early stages of Alzheimer disease. Am J Neuroradiol 15:929–937
15. Jack CR Jr, Petersen RC, O'Brien PC, Tangalos EG (1992) MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 42:183–188
16. Pennanen C, Kivipelto M, Tuomainen S, Hartikainen P, Hänninen T, Laakso MP, Hallikainen M, Vanhanen M, Nissinen A, Helkala E, Vainio P, Vanninen R, Partanen K, Soininen H (2004) Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. Neurobiol Aging 25:303–310 doi:10.1016/S0197-4580(03)00084-8
17. Du AT, Schuff N, Amend D, Laakso MP, Hsu YY, Jagust WJ, Yaffe K, Kramer JH, Reed B, Norman D, Chui HC, Weiner MW (2001) Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. J Neurol Neurosurg Psychiatry 71:441–447 doi:10.1136/jnnp.71.4.441
18. De Santi S, de Leon MJ, Rusinek H, Convit A, Tarshish CY, Roche A, Tsui WH, Kandil E, Boppana M, Daisley K, Wang GJ, Schlyer D, Fowler J (2001) Hippocampal formation glucose

metabolism and volume losses in MCI and AD. Neurobiol Aging 22:529–539 doi:10.1016/S0197-4580(01)00230-5
19. Convit A, De Leon MJ, Tarshish C, De Santi S, Tsui W, Rusinek H, George A (1997) Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. Neurobiol Aging 18:131–138 doi:10.1016/S0197-4580(97)00001-8
20. Chetelat G, Baron J (2003) Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. Neuroimage 18:525–541 doi:10.1016/S1053-8119(02)00026-5
21. Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C (2004) Morphological classification of brains via high-dimensional shape transformations and machine learning methods. Neuroimage 21:46–57 doi:10.1016/j.neuroimage.2003.09.027
22. Fan Y, Shen D, Davatzikos C (2005) Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. Med Image Comput Comput Assist Interv Int Conf 8:1–8
23. Cortes C, Vapnik V (1995) Support-Vector Networks. Mach Learn 20:273–297
24. Fan Y, Batmanghelich N, Clark CM, 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 doi:10.1016/j.neuroimage.2007.10.031
25. Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RSJ (2008) Automatic classification of MR scans in Alzheimer's disease. Brain 131:681–689 doi:10.1093/brain/awm319
26. Vemuri P, Gunter JL, Senjem ML, Whitwell JL, Kantarci K, Knopman DS, Boeve BF, Petersen RC, Jack CR Jr (2008) Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies. Neuroimage 39:1186–1197 doi:10.1016/j.neuroimage.2007.09.073
27. Teipel SJ, Born C, Ewers M, Bokde ALW, Reiser MF, Möller H, Hampel H (2007) Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. Neuroimage 38:13–24 doi:10.1016/j.neuroimage.2007.07.008
28. Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM (2008) Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. Neurobiol Aging 29:514–523 doi:10.1016/j.neurobiolaging.2006.11.010
29. Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM (2008) Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 41:1220–1227 doi:10.1016/j.neuroimage.2008.03.050
30. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34:939–944
31. Morris JC (1993) The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43:2412–2414
32. Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198 doi:10.1016/0022-3956(75)90026-6
33. Benton AL (1968) Genuine memory deficits in dementia. Neuropsychologia 6:53–60 doi:10.1016/0028-3932(68)90038-9
34. Sano M, Stern Y, Mayeux R, Hartman S, Devanand DP (1987) A standardized technique for establishing the onset symptoms of probable Alzheimer's disease. J Clin Exp Neuropsychol 9:65
35. Stern Y, Albert M, Brandt J, Jacobs DM, Tang MX, Marder K, Bell K, Sano M, Devanand DP, Bylsma F et al (1994) Utility of extrapyramidal signs and psychosis as predictors of cognitive and functional decline, nursing home admission, and death in Alzheimer's disease: prospective analyses from the Predictors Study. Neurology 44:2300–2307
36. Stern Y, Mayeux R, Sano M, Hauser WA, Bush T (1987) Predictors of disease course in patients with probable Alzheimer's disease. Neurology 37:1649–1653
37. Grober E, Buschke H (1987) Genuine memory deficits in dementia. Dev Neuropsychol 3:13–36
38. Goldblum MC, Gomez CM, Dalla Barba G, Boller F, Deweer B, Hahn V, Dubois B (1998) The influence of semantic and perceptual encoding on recognition memory in Alzheimer's disease. Neuropsychologia 36:717–729 doi:10.1016/S0028-3932(98)00007-4
39. Benton AL (1974) The revised visual retention test: clinical and experimental applications. Psychological Corporation, New York
40. Sirigu A, Cohen L, Duhamel JR, Pillon B, Dubois B, Agid Y (1995) A selective impairment of hand posture for object utilization in apraxia. Cortex 31:41–55
41. Mayeux R, Rosen W (1983) The dementias. Raven, New York
42. Deloche G, Hannequin D (1997) Test de dénomination orale d'images D080. Les Editions du Centre de Psychologie Appliquée, Paris
43. Kaplan E, Goodglass H, Weintraub S (1983) The Boston Naming Test. Lea and Febiger, Philadelphia
44. Dubois B, Slachevsky A, Litvan I, Pillon B (2000) The FAB: a Frontal Assessment Battery at bedside. Neurology 55:1621–1626
45. Dorion AA, Sarazin M, Hasboun D, Hahn-Barma V, Dubois B, Zouaoui A, Marsault C, Duyme M (2002) Relationship between attentional performance and corpus callosum morphometry in patients with Alzheimer's disease. Neuropsychologia 40:946–956 doi:10.1016/S0028-3932(01)00150-6
46. Robbins TW, James M, Owen AM, Sahakian BJ, Lawrence AD, McInnes L, Rabbitt PM (1998) A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. J Int Neuropsychol Soc 4:474–490 doi:10.1017/S1355617798455073
47. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289 doi:10.1006/nimg.2001.0978
48. Ashburner J, Friston K (2003) Image segmentation. In: Frackowiak R, Friston K, Frith C, Dolan R, Price C, Zeki S, Ashburner J, Penny W (eds) Human brain function, 2nd edn. Academic, San Diego, pp 695–706
49. Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Thieme Medical Publisher, New York
50. Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. Hum Brain Mapp 7:254–266 doi:10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G
51. Ashburner J, Andersson JL, Friston KJ (2000) Image registration using a symmetric prior—in three dimensions. Hum Brain Mapp 9:212–225 doi:10.1002/(SICI)1097-0193(200004)9:4<212::AID-HBM3>3.0.CO;2-#
52. Redner R, Walker H (1984) Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 26:195–239 doi:10.1137/1026034
53. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York
54. Burton EJ, Karas G, Paling SM, Barber R, Williams ED, Ballard CG, McKeith IG, Scheltens P, Barkhof F, O'Brien JT (2002) Patterns of cerebral atrophy in dementia with Lewy bodies using voxel-based morphometry. Neuroimage 17:618–630 doi:10.1016/S1053-8119(02)91197-3

55. Barber R, McKeith IG, Ballard C, Gholkar A, O'Brien JT (2001) A comparison of medial and lateral temporal lobe atrophy in dementia with Lewy bodies and Alzheimer's disease: magnetic resonance imaging volumetric study. Dement Geriatr Cogn Disord 12:198–205 doi:10.1159/000051258
56. Burton EJ, McKeith IG, Burn DJ, Williams ED, O'Brien JT (2004) Cerebral atrophy in Parkinson's disease with and without dementia: a comparison with Alzheimer's disease, dementia with Lewy bodies and controls. Brain 127:791–800 doi:10.1093/brain/awh088
57. Ballmaier M, O'Brien JT, Burton EJ, Thompson PM, Rex DE, Narr KL, McKeith IG, DeLuca H, Toga AW (2004) Comparing gray matter loss profiles between dementia with Lewy bodies and Alzheimer's disease using cortical pattern matching: diagnosis and gender effects. Neuroimage 23:325–335 doi:10.1016/j.neuroimage.2004.04.026