Information Sciences 181 (2011) 903–916

Contents lists available at ScienceDirect

Information Sciences

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

^{18}F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis ☆

I.A. Illán [a], J.M. Górriz [a,*], J. Ramírez [a], D. Salas-Gonzalez [a], M.M. López [a], F. Segovia [a], R. Chaves [a], M. Gómez-Rio [c], C.G. Puntonet [b], the Alzheimer's Disease Neuroimaging Initiative

[a] Department of Signal Theory, Networking and Communications, University of Granada, 18071 Granada, Spain
[b] Department of Computer's Architecture and Technology, University of Granada, 18071 Granada, Spain
[c] Nuclear Medicine Service, Virgen de las Nieves Hospital, Granada, Spain

ARTICLE INFO

Article history:
Received 20 January 2010
Received in revised form 4 September 2010
Accepted 22 October 2010

Keywords:
Alzheimer's disease (AD)
Computer aided diagnosis
Principal component analysis (PCA)
Independent component analysis (ICA)
Support vector machine (SVM)
Supervised learning
FDG-PET

ABSTRACT

Finding sensitive and appropriate technologies for non-invasive observation and early detection of Alzheimer's disease (AD) is of fundamental importance to develop early treatments. In this work we develop a fully automatic computer aided diagnosis (CAD) system for high-dimensional pattern classification of baseline ^{18}F-FDG PET scans from Alzheimer's disease neuroimaging initiative (ADNI) participants. Image projection as feature space dimension reduction technique is combined with an eigenimage based decomposition for feature extraction, and support vector machine (SVM) is used to manage the classification task. A two folded objective is achieved by reaching relevant classification performance complemented with an image analysis support for final decision making. A 88.24% accuracy in identifying mild AD, with 88.64% specificity, and 87.70% sensitivity is obtained. This method also allows the identification of characteristic AD patterns in mild cognitive impairment (MCI) subjects.

1. Introduction

Over 4 million people in the US have probable AD, the 10% of the population over the age of 65 years (3% of those 65–74 years old and 45% over 85 years old) [20], and a very substantial number have other dementias [61]. The incidence of dementia is expected to double during the next 20 years, thus becoming an enormous public health problem [9]. No existing treatment has yet been shown to slow the progression of AD but a large number of potential treatments are under development. For emergence of effective treatments for AD as well as performance of prevention trials, it is of fundamental importance to identify those subjects at high risk for cognitive decline and dementia at the earliest stage possible. At the present time, a certain AD diagnostic can only be reached after autopsy [31], therefore requiring developments of non-invasive methods for early diagnosis. Extensive progress has been achieved to this end, but an accurate and early diagnosis of the AD still remains a challenge.

The positron emission tomography (PET) scans being used within this work are an example of non-invasive, three-dimensional functional imaging modalities measuring the brain's rate of glucose metabolism with the tracer [^{18}F] Fluorodeoxyglucose. In AD, characteristic brain regions show decreased glucose metabolism, specifically bilaterally regions in the temporal

☆ Data used in the preparation of this article were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf).

* Corresponding author. Fax: +34 958243271.
E-mail address: gorriz@ugr.es (J.M. Górriz).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.10.027

and parietal lobes, posterior cingulate gyri and precunei, as well as frontal cortex and whole brain in more severely affected patients [17,23,24,11,19,47,49,50,34,33,39,1,53,40]. Lower glucose metabolism has been correlated with dementia severity, either by assessing the correlation between performance on measures of overall cognitive function, such as the mini-mental state exam (MMSE), and cerebral metabolic rate for glucose (CMRgl) reduction over all study participants, or by comparing subgroups based on measures of disease severity, such as the MMSE, clinical dementia rating (CDR), or Global Deterioration Scale score [11,50,24,65].

Data used in the preparation of this article were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database (www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5 year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, 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 Principle Investigator of this initiative is Michael W. Weiner, M.D., VA Medical Centre and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the US and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55–90, 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 (for more information see www.adni-info.org).

A vast amount of the AD literature suggests that early manifestations of AD occur in a prodromal stage, years before the symptoms of the disease appear and are clinically detectable, allowing for other non-invasive techniques for detection as nuclear imaging. The examination of the predictive abilities of nuclear imaging with respect to AD in this early stage has been widely studied, through visual assessments performed by experts [8,15,33,66], or by means of voxel-wise statistical analysis as SPM, NEUROSTAT & 3D-SSP, ANOVA or MANCOVA [25,18,36,64,71,63,45,54]. Recently, a new branch of emerging research has demonstrated that Machine Learning techniques may also be powerful analysis tools of brain imaging, with recent works adapting state-of-art computer vision techniques in MRI for early AD diagnosis [21,16], or Support Vector Machine classification in single photon emission computerized tomography (SPECT) analysis [26,42,35,44,12,62]. This decision taking oriented framework usually considers as one observation all the voxels in a single scan, allowing for a regional or global brain image analysis, contrary to voxel-wise statistical tools. On the other hand, it suffers from the *curse of dimensionality* problem. This major problem, associated with pattern recognition systems, occurs when the number of available features for designing the classifier is very large compared with the number of available training examples.

One of the most successful and prolific solutions to the small sample size problem is based in the eigenimage approach of Turk and Pentland [73], originally applied to the related problem of face detection. Principal component analysis (PCA) or independent component analysis (ICA) eigenimage decomposition allows to detect underlying patterns that differentiate individuals within a population, even if the differences are subtle. This is the case in AD diagnosis, where eigenimage analysis has been successfully adapted in several ways [35,42,44] for SPECT imaging and for PET imaging [43]. From these works one may conclude that a reduced number of components or *eigenbrains* are necessary to describe the information about Alzheimer's disease encoded in the images. However, some eligibility criteria must be chosen in order to select those discriminant components, based on some *a priori* knowledge of the underlying data structure.

The novelty of the approach presented in this work resides in a CAD construction in the eigenimage framework for PET imaging, which does not require for an eigenbrain selection (see Section 2.3), avoiding the introduction of subjective information in the process, and guaranteeing the inclusion of characteristic AD patterns at the same time. This last fact allows for a meaningful eigenimage decomposition, and a *image analysis* support for the final decision making. These advantages are achieved by searching for characteristic vectors representing AD, subsequently demixing with PCA and ICA, and using projections for growing a SVM classifier. PCA and ICA have been proven to be useful in a vast variety of fields [5,72,22,29,3,57]. On the other hand, SVMs have attracted recent attention from the pattern recognition community due to a number of theoretical and computational merits derived from Statistical Learning Theory [74]. These techniques have also been successfully used in a large number of applications, including pedestrian detection [10], time-series forecasting [30], speech processing [60,28] and medical imaging diagnosis [59,58,41,26,27,35]. The proposed method, tested on 401 ADNI PET images, is developed with the aim of reducing the subjectivity in visual interpretation of these scans by clinicians, thus providing an objective methodology for improving the accuracy of diagnosing Alzheimer's disease in its early stage.

2. Materials and methods

2.1. Subjects

Baseline FDG PET data from 401 ADNI participants, acquired from Siemens, general electric (GE), Philips, Siemens HRRT and BioGraph HiRez PET scanners, were collected from the ADNI Laboratory on NeuroImaging (LONI, University of California, Los Angeles) website (http://www.loni.ucla.edu/ADNI/).

Participant's enrolment was conditioned to some eligibility criteria. General inclusion/exclusion criteria were as follows:

- Normal control subjects: MMSE scores between 24–30 (inclusive), a CDR of 0, non-depressed, non MCI, and non-demented. The age range of normal subjects will be roughly matched to that of MCI and AD subjects. Therefore, there should be minimal enrolment of normals under the age of 70.
- MCI subjects: MMSE scores between 24–30 (inclusive), a memory complaint, have objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.
- Mild AD: MMSE scores between 20–26 (inclusive), CDR of 0.5 or 1.0, and meets NINCDS/ADRDA criteria for probable AD.

Therefore, FDG PET data was separated into three different classes: 97 normal control (NC) (age range: 76.7 ± 5.2 (mean ± standard deviation)), 209 MCI (age range: 76.0 ± 7.7), and 95 AD (age range: 77.3 ± 7.4). From the 209 MCI, ADNI sites provide information about those who are stable after 2 years follow-up, and those who have converted into AD, which results as 45 MCI converters (MCI-c) and 164 MCI non-converters (MCI-nc). 182 images from the dataset are held-out for test, while 219 are used for training and cross-validation.

For posterior analysis, the data was arranged into three different groups. For testing early diagnosis capabilities, MCI was considered a pre-stage of AD [52,53,21], and the following groups where designed:

- **Group 1** All the database images are considered. Both AD and MCI patients are labelled as positive, and Normal controls as negative.
- **Group 2** Only AD (positive) and Normal controls (negative) patient images are considered.
- **Group 3** Only MCI (positive) and Normal controls (negative) patient images are considered.

2.2. Image database pre-processing

FDG PET scans were acquired according to a standardized protocol. A 30 min dynamic emission scan, consisting of 65 min frames, was acquired starting 30 min after the intravenous injection of 5.0 ± 0.5 mCi of ^{18}F-FDG, as the subjects, who were instructed to fast for at least 4 h prior to the scan, lay quietly in a dimly lit room with their eyes open and minimal sensory stimulation. Data were corrected for radiation–attenuation and scatter using transmission scans from Ge-68 rotating rod sources and reconstructed using measured-attenuation correction and image reconstruction algorithms specified for each scanner (http://www.loni.ucla.edu/ADNI/Data/ADNIData.shtml). Following the scan, each image was reviewed for possible artifacts at the University of Michigan and all raw and processed study data was archived.

Subsequently, the images were normalized through a general affine model, with 12 parameters [76] using the SPM5 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm5). After the affine normalization, the resulting image was registered using a more complex non-rigid spatial transformation model (see first step in Fig. 1). The non-linear deformations to the Montreal Neurological Imaging (MNI) Template were parametrized by a linear combination of the lowest-frequency components of the three-dimensional cosine transform bases [4]. A small-deformation approach was used, and regularization was by the bending energy of the displacement field, ensuring that the voxels in different FDG-PET images refer to the same anatomical positions in the brains. After spatial normalization, an intensity normalization was required in order to perform direct images comparisons between different subjects. The intensity of the images was normalized to a value I_{max}, obtained

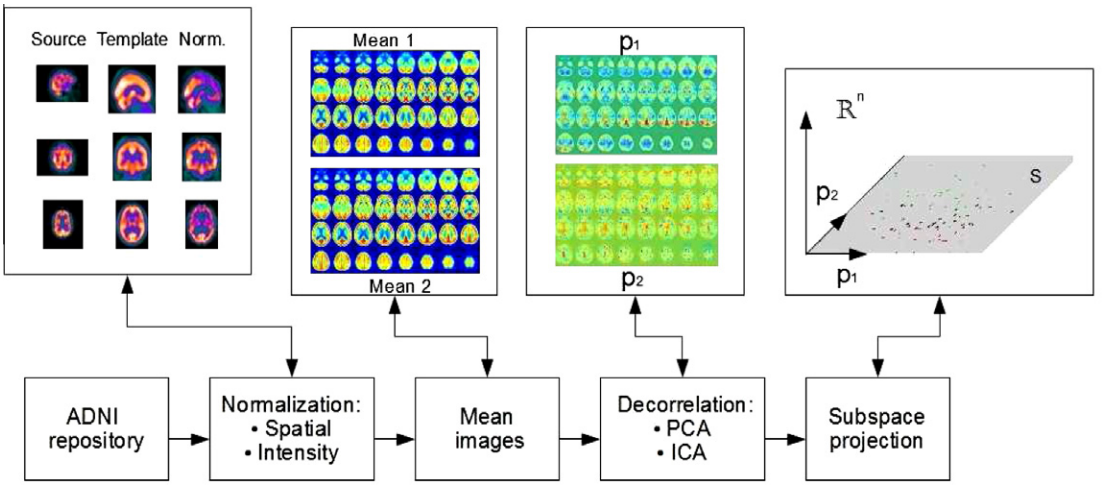

Fig. 1. Preprocessing steps to extract features from raw data.

averaging the 0.1% of the highest voxel intensities exceeding a threshold. The threshold was fixed to the 10th bin intensity value of a 50 bins intensity histogram, for discarding most low intensity records from outside-brain regions, and preventing image saturation.

2.3. Feature extraction

Brain images of this study are 3D images $\mathbf{t}_i$ of size $M = 79 \times 95 \times 69 \sim 5 \cdot 10^5$ voxels. The approach taken in this work for feature extraction is schematically shown in Fig. 1. It starts by rearranging each normalized image to a vector of dimension M, thus considering each patient image as a vector of the euclidean space $\mathbf{t}_i \in \mathbb{R}^M$ belonging to the class AD, MCI or NC. Thus, images are considered as a vectors in a high dimensional vectorial space, the 'brain images space'. In this space, most of the dimensions represent voxels that are not relevant for AD discrimination, and a dimensional reduction is required.

A dimensional reduction is achieved by searching for a *discriminative* subspace S of $\mathbb{R}^M$, and project the raw data onto this subspace. This subspace S is spanned by a basis of *characteristic vectors* $\{\mathbf{p}_k\}$ that represent each class. Projecting an image vector $\mathbf{t}_i$ onto S determines its coordinates in this subspace. This coordinates are used as features, since they give a measure of the similarity between the image $\mathbf{t}_i$ and the class $\mathcal{C}_k$, represented by $\mathbf{p}_k$.

We present three different ways of constructing this subspace S. The first immediate hint is to obtain a class average image vector for each class, as a way of characterizing these classes and use them as a basis of S. Specifically let the full 3D brain image database be $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_N$, each pertaining to a class $\mathcal{C}_1 = \text{AD}$, $\mathcal{C}_2 = \text{MCI}$, $\mathcal{C}_3 = \text{NC}$. A within class average brain image $\mathbf{p}$ can be defined as:

$$\mathbf{p}_k = \frac{1}{N_k} \sum_{\{\mathbf{t}_i \subset \mathcal{C}_k\}} \mathbf{t}_i, \quad k = 1, 2, 3, \tag{1}$$

where N_k denotes the number of images in the class $\mathcal{C}_k$. These average images form a vector basis, spanning the subspace S of the 'brain images space'. For example, in the case of group 2, a two element basis $\mathbf{p}_{AD}$, $\mathbf{p}_{NC}$ will be obtained. The projection of an image vector into this subspace will weigh the contribution of each mean image vector in describing it, which may be used as features for classification. But these projection coefficients may be highly correlated (see Fig. 4). This suggests the use of the average images to construct a new image basis, from whose projection coefficients become separable. This can be achieved with the use of PCA and ICA.

2.4. PCA

PCA is a standard technique for extracting the most significant features from a dataset, frequently used to reduce the raw data to a subset of features that contains the largest amount of variance. It is based on a linear transformation acting on a zero mean dataset, that diagonalizes its covariance matrix. The resulting eigenvectors are a new set of uncorrelated variables, whose variance is represented by its eigenvalue. The use of PCA within this work is justified not as a dimensional reduction technique but it is applied to the average images vector obtained by Eq. (1). Therefore, we introduce PCA as a procedure for extracting a *decorrelated* eigenvector basis from the previous defined average image vectors. Similar frameworks in which a reduced number of sources are obtained from a linear combination of the mixtures have also been used in other fields as face recognition problem [5], or functional magnetic resonance imaging (fMRI) [46].

Applying PCA to the set $\mathbf{p}_k$ defined in Eq. (1) can be seen as a simple basis change in S, due to its linear character. This transformation changes to an eigenvector basis set of image vectors whose correlation is minimized. We refer to them as 'eigenbrains' because of their brain like appearance (see Appendix A.1 for details). The number of vectors in this new basis is the same as the number of classes used to define the average images. Therefore, none of the eigenbrains are neglected, independently from its variance eigenvalue. The projection of the image vectors onto the eigenbrains space, i.e. the subspace spanned by the eigenbrains, will determine the coordinates of each image vector in this subspace. It is expected that this projection produce a pattern more suitable for class separation than the projection onto the average image space, due to the eigenbrains decorrelation.

2.5. ICA

PCA only takes into account pair-wise relationships between voxels of the brain images. It seems reasonable to expect that important information may be contained in the high-order relationships among voxels, given in terms of cumulants, for which ICA is a sensitive method. The ICA transformation is used for capturing group-differences from high order voxel relations, generating from the original average images sources $\mathbf{p}_k$ defined in Eq. (1) a new set of statistically independent components. As the PCA-eigenbrains, all these latent variables will be used to construct a subspace of the brain image space, where the essence of the class differences is enhanced upon projecting the image vectors onto it. For achieving the ICA transformation, we made use of FastICA [55] which is an iterative fixed-point algorithm, with the contrast function chosen to be a cubic polynomial (see Appendix A.2 for details). FastICA has been proven to be a reliable algorithm in other fields, as fMRI analysis, if compared to other algorithms as Infomax or JADE [69]. The stronger constraint of independence of ICA over PCA, theoretically implies a greater capability of class prominent features detection.

2.6. Classification

Once the basis vector $\{\mathbf{p}_k\}$ of the subspace S is obtained, a set of features can be extracted to train a SVM with labelled data. The training data is obtained from the raw data as:

$$(x_k)_i = \mathbf{p}_k \mathbf{t}_i, \quad i = 1, \ldots, N, \; k = 1, 2, \ldots, K. \tag{2}$$

These coefficients $(x_k)_i$ are the coordinates of the ith-brain image $\mathbf{t}_i$ in the subspace spanned by the average images (for $\mathbf{p}_k$ being the average image vectors basis), the eigenbrains (for $\mathbf{p}_k$ being the eigenbrains basis), or the independent components (for $\mathbf{p}_k$ being the ICs basis). We used these coordinates values x as N K-dimensional training vectors:

$$\mathrm{x}_i = [x_1, x_2, \ldots, x_K]_i, \quad i = 1, 2, \ldots, N, \tag{3}$$

each of them with its corresponding class label $y_i \in \{\pm 1\}$. In practice, $K = 2$ when using two average images from the classes AD and NC, $K = 3$ when using three average images from the classes AD, MCI and NC, and $K = 4$ when using four average images from the classes AD, MCI-c, MCI-nc and NC. Therefore, only a 2, 3 or 4-dimensional vector x encodes *all* the information of the PET brain scans relevant for AD discrimination.

The SVM 'learns' from the labelled training data obtaining a hyperplane that separates the data into the two classes, maximizing the margin between them (see Appendix A.3 for details). Once it has been obtained, new samples with unknown labels can be categorized. A feature vector is obtained from the test image using Eqs. (2) and (3), and it is classified according to its sign in Eq. (A.7), once the hyperplane is determined. Also, samples with known labels may be used to test the reliability of the method, with some cross-validation strategy.

3. Results

The classification performance of our approach is tested in three steps: training, cross-validation and test. Cross validation is achieved by means of the leave-one-out method, a technique that iteratively holds out a subject for test, while training the classifier with the remaining subjects, so that each subject is left out once. Different experiments were performed by considering the three groups described in Section 2.1, and by varying the feature vectors dimension, given by K in Eq. (3), from 2 to 4. This dimension reduction was achieved using different number of average vectors to construct the feature vectors, considering NC and non-NC average vectors ($K = 2$); NC, AD and MCI ($K = 3$); and NC, AD, MCI-c, and MCI-nc ($K = 4$). Once the average vectors were obtained from the training set, each image was processed using Eq. (2) to obtain the feature vectors used to grow the SVM classifier. After the optimum parameters were obtained, the classification performance was tested in a held out set of images.

Fig. 2 shows the results achieved using PCA to decorrelate the average vectors, while Fig. 3 displays the results obtained with ICA, for different SVM kernels: linear and RBF, and the different groups. Table 1 expands the classification results in the

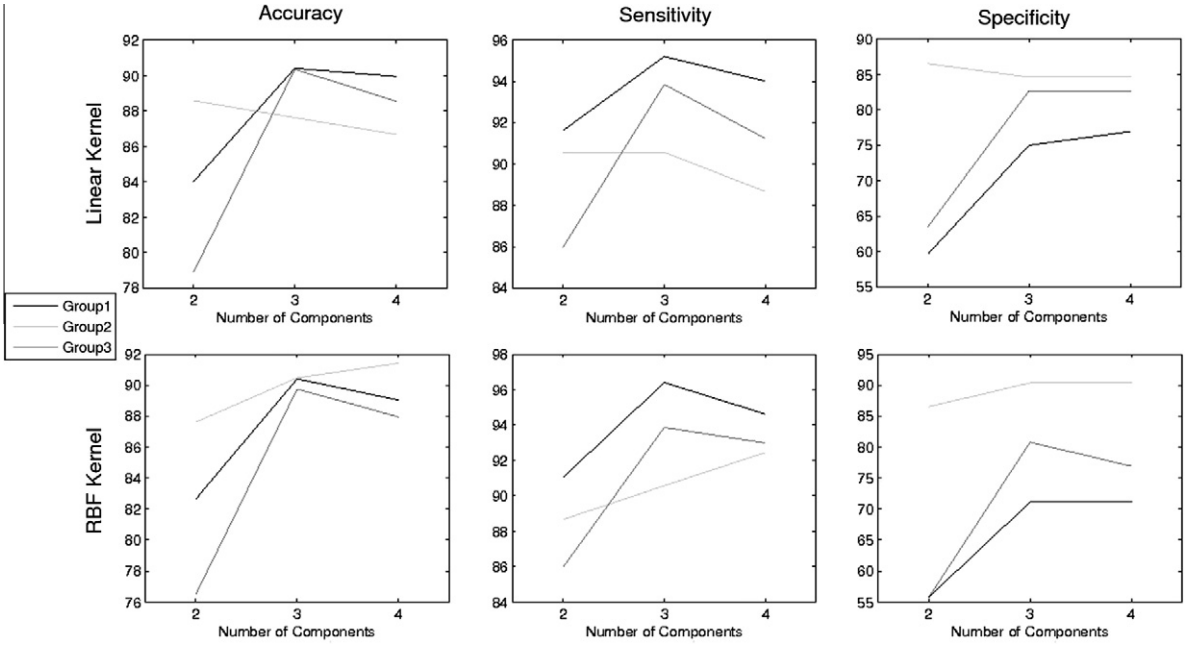

Fig. 2. PCA results of classification using: linear kernel (upper row), RBF kernel (lower row). Varying the number of components, accuracy results (first column), sensitivity results (second column), specificity results (third column), expressed in %.

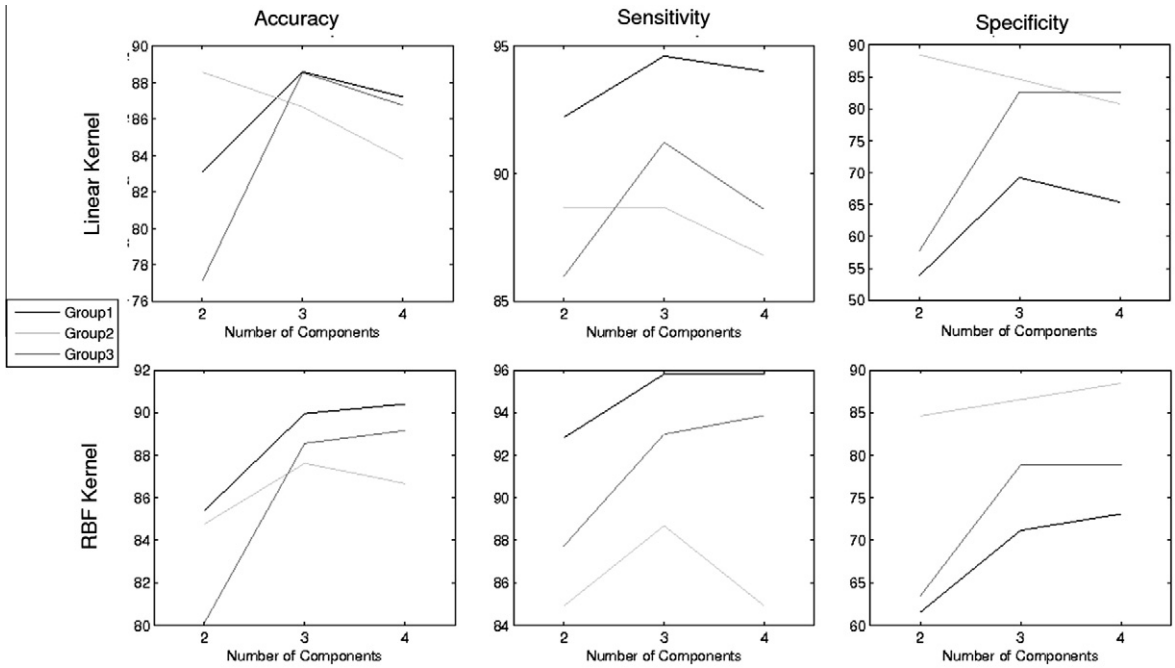

Fig. 3. ICA results of classification using: linear kernel (upper row), RBF kernel (lower row). Varying the number of components, accuracy results (first column), sensitivity results (second column), specificity results (third column), expressed in %.

Table 1
Statistical measures of performance of PCA and ICA feature selection methods, for the three sample groups, and using four components. Voxel-as-features (VAF) results are reported as reference, using two classifiers: *k*-nearest-neighbour (*k*nn with $k = 5$) and SVM linear (no kernel is used due to peaking phenomenon). Comparisons are established with PCA-fisher [35,42–44] for the same number of components.

	VAF		PCA		ICA		PCA-f	
	*k*nn	Linear kernel	RBF kernel	Linear kernel	RBF kernel	Linear kernel	RBF kernel	Linear kernel
Group 1								
Accuracy (%)	65.38	74.73	77.47	74.73	75.82	76.37	73.63	74.18
Specificity (%)	61.36	25.00	34.09	15.91	34.09	20.45	0.00	15.91
Sensitivity (%)	66.67	90.58	91.30	93.48	89.13	94.20	97.10	92.75
Positive likelihood	1.7255	1.2077	1.3853	1.1116	1.3523	1.1843	0.9710	1.1030
Negative likelihood	0.5432	0.3768	0.2551	0.4099	0.3188	0.2834	1.0000	0.4555
Group 2								
Accuracy (%)	65.88	81.18	88.24	82.35	87.06	75.29	67.06	65.88
Specificity (%)	88.64	84.09	88.64	70.45	86.36	59.09	75.00	63.64
Sensitivity (%)	41.46	78.05	87.80	95.12	87.80	92.68	58.54	68.29
Positive likelihood	3.6488	4.9059	7.7268	3.2195	6.4390	2.2656	2.3415	1.8780
Negative likelihood	0.6604	0.2610	0.1376	0.0692	0.1412	0.1238	0.5528	0.4983
Group 3								
Accuracy (%)	56.03	68.79	68.09	67.38	70.21	68.79	65.96	68.79
Specificity (%)	63.64	29.55	34.09	15.91	40.91	18.18	9.09	25.00
Sensitivity (%)	52.58	86.60	83.51	90.72	83.51	91.75	91.75	88.66
Positive likelihood	1.4459	1.2291	1.2670	1.0789	1.4132	1.1214	1.0093	1.1821
Negative likelihood	0.7452	0.4536	0.4838	0.5832	0.4032	0.4536	0.9072	0.4536

held out set for the case of $K = 4$, once the previously described steps, projection, PCA and ICA have been applied to the raw data. Three parameters that measure the performance of the classification task are: the Accuracy, the Specificity and the Sensitivity. The definitions are:

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}, \quad \text{PL} = \text{Sensitivity}/(1 - \text{Specificity}),$$

$$\text{Sensitivity} = \frac{Tp}{Tp + Fn}, \quad \text{NL} = (1 - \text{Sensitivity})/\text{Specificity},$$

$$\text{Specificity} = \frac{Tn}{Tn + Fp},$$

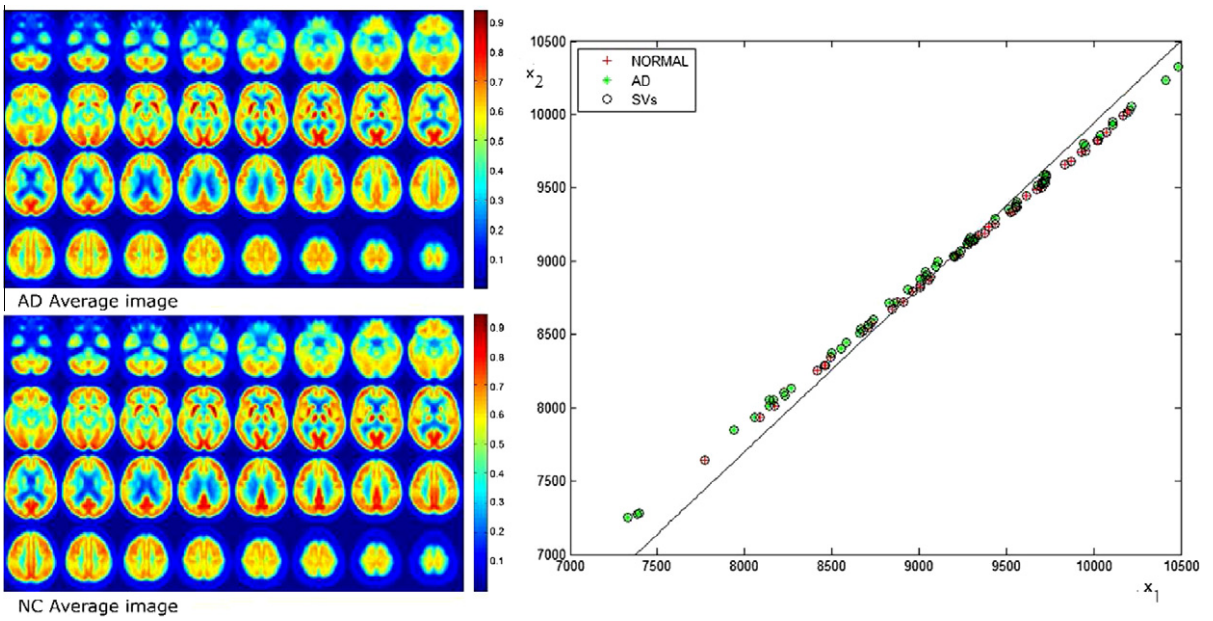

Fig. 4. Linear kernel SVM classifier (support vectors (SVs) are marked with circles, see legend) and training vectors $x_i = [\mathbf{\Omega}_1, \mathbf{\Omega}_2]_i$ (right) obtained projecting each image onto the NC and AD mean images (left, transaxial slices).

where *Tp*, *Tn*, *Fp*, and *Fn* denote true positives, true negatives, false positives, and false negatives, respectively. Positive and negative likelihoods (PL/NL) are also displayed as a measure of the positive and negative predictive value of the method, given its prevalence independence. Results using PCA and ICA are compared to VAF [68,67] and PCA-f [35,42,43], selecting the best performance scenarios for kernel SVM and knn classifiers.[1]

The training process is showed in detail in Figs. 4–6, where the case of $K = 2$ is analysed in Group 2. The x_1 and x_2 values, obtained from Eq. (3) by projecting each image vector into the subspace spanned by the AD (including MCIs) and NC average images, respectively, are represented for the cases of projection, PCA and ICA, together with their eigenimages.

Fig. 4 shows the x_1 and x_2 values obtained when simply projecting each image into the AD and NC mean images. Regardless of intensity differences of the mean images in some brain regions as the posterior cingulate gyri, these are subtle differences. Therefore, the x's coefficients are similar, revealing a high correlation between them. This allows to state that mean images are not suitable vectors for classification, and justifies the use of PCA and ICA for extracting an independent or decorrelated vector basis from them.

Fig. 5 shows how the application of PCA on the mean images allows a class separation, and an efficient classifier construction. The separation of the training samples into the two classes is given by the x_2 value (the Eigenbrain 2 coordinate), while the value of x_1 (the Eigenbrain 1 coordinate) plays no role in the classifier construction. Concretely, positive values of x_2 correspond with AD affected subjects, while negative ones with NC. The interpretation of this fact may be guided by the eigenbrains appearance. Eigenbrain 1 has a normal PET brain image appearance, while Eigenbrain 2 is an implausible one, indeed complementary to Eigenbrain 1. Eigenbrain 1 serves as basis to construct any database brain image sample, by contributing with the basis aspects of a functional brain image, while Eigenbrain 2 pinpoints the disease affected areas which are class-discriminative. Specifically, it highlights the posterior cingulate gyri and precunei, as well as the temporo-parietal region, both considered as typically affected by glucose hypometabolism in the AD [13,48,70,51]. A closer look shows that it also selects small thalamus regions, which has never been described as a relevant region for diagnosis. This has to be interpreted as a spatial normalization failure: when comparing AD and NC mean images, a spatial mismatch is found between thalamus regions. Since PCA coordinates group the maximum variance, adjacent thalamus regions are selected to construct Eigenbrain 2, but are small enough to introduce negligible information into the feature selection. To summarize, a sample is taken as AD when its x_2 value is positive, independently from its x_1 value, therefore complementing Eigenbrain 1 image with lower intensity values in the cingulum, precunei and temporo-parietal regions. On the other hand, a sample is considered as NC when its x_2 value is negative, adding to Eigenbrain 1 a normal glucose metabolism pattern to the previous mentioned regions.

The class separability improvement due to ICA is shown in Fig. 6. A similar analysis to the PCA one can be made, since the independent component (IC) 1 is similar to the Eigenbrain 2, but now IC2 introduces some discriminative power, contrary to Eigenbrain 1. This fact is reflected in the classifier slope, revealing that a combination of x_1 and x_2 values are necessary to determine the image class. The main difference between IC2 and Eigenbrain 1 is the extremely low intensity values of

[1] Detailed results concerning the comparison of different kernels and classifiers can be found on [44].

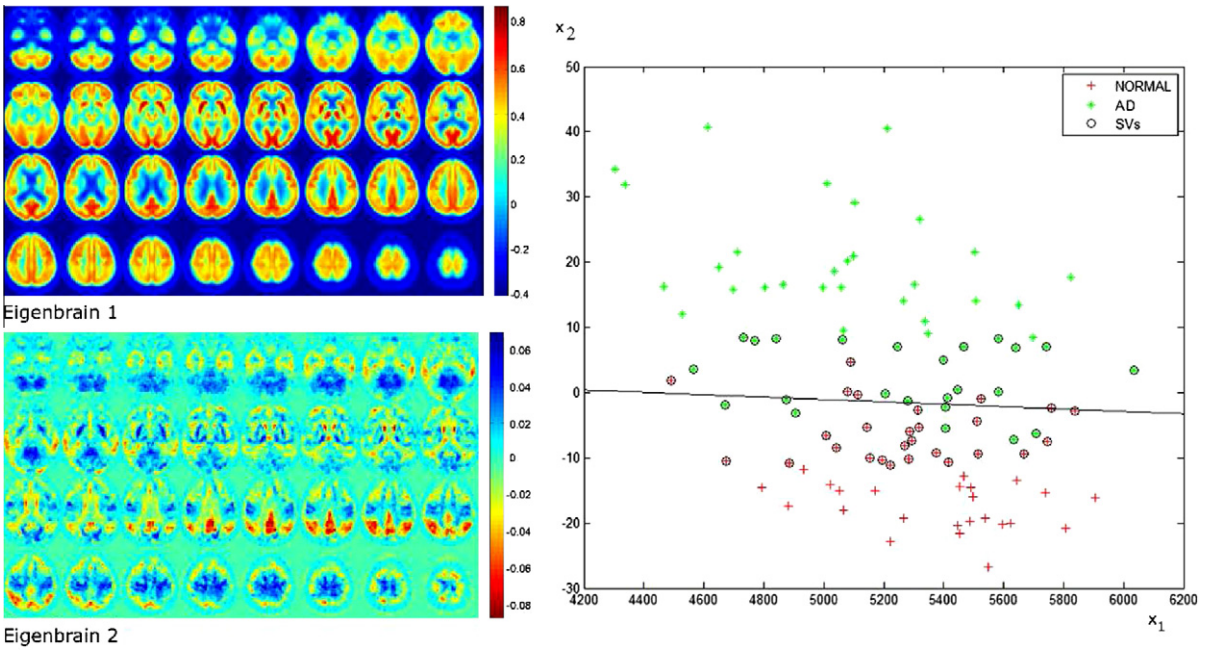

Fig. 5. Linear kernel SVM classifier (support vectors are marked with circles) and training vectors $x_i = [x_1, x_2]_i$ (right) obtained projecting each image onto the PCA Eigenbrain images space (left, transaxial slices).

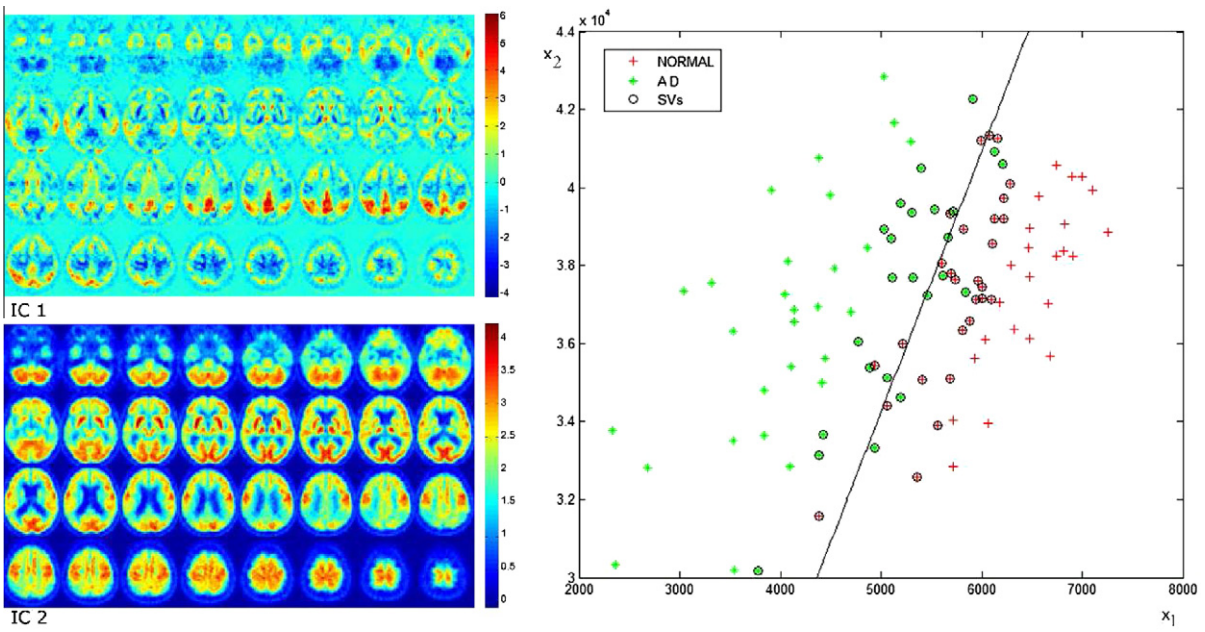

Fig. 6. Linear kernel SVM classifier (support vectors are marked with circles) and training vectors $x_i = [x_1, x_2]_i$ (right) obtained projecting each image onto the IC images space (left, transaxial slices).

cingulum, temporo-parietal regions as well as lower intensity values in the frontal cortex. The last fact is related to more severely AD affected patients, and therefore implies that a combination of low x_1 and x_2 values (the lower left corner) represent severely AD affected subjects, while high x_1 and x_2 values (the upper right corner) are more likely to represent NC image patterns, with its intermediate region delimited by the classifier line.

Concerning the different feature extraction techniques, performance measures of PCA and ICA methods are shown on Figs. 2 and 3, respectively, for the three groups. Use of different kernels are also displayed, providing higher sensitivity values with the use of RBF kernels and higher specificity values when using a linear kernel. The election of the kernel to use would

depend on whether higher sensitivity or specificity values are desired, since accuracy values are similar for both kernels. Group 2 shows more stable behaviour and higher specificity values than groups 1 and 3. This may be due to balance condition of the training samples. Group 2 satisfies this condition (with 53 positives and 52 negatives in the training data), but not group 1 and 3 (with 167 positives and 52 negatives, and 114 positives and 52 negatives, respectively) introducing imbalanced training datasets, and therefore lower specificity values when using a small number of components.

These low specificity values may also be explained by the introduction in groups 1 and 3 of MCI patients, whose image pattern of brain atrophy is complex and highly variable, and it evolves in time as the disease progresses [51,18,38]. As the number of components grows, information about the MCI pattern heterogeneity is introduced in the classifier, thus enhancing the specificity values in the mentioned groups. When average images of MCI-c and MCI-nc are added to those of NC and AD ($K = 4$), a 4-dimensional feature vector is produced. Eigenbrains 3 and 4 are obtained in addition to eigenbrains 1 and 2 of Fig. 5. 4-dimensional spaces are not directly representable, but projections in bidimensional spaces can capture relevant information, as the projection in the x_2-x_3 axis given in Fig. 7. This figure illustrates how the information of MCIs is introduced in the classification, showing eigenbrains 2 and 3 (left), and its corresponding coordinates for the training vectors (right) of group 1. In this figure, Eigenbrain 1 has been merged with eigenbrain 2 and 3 in order to allow to identify the brain regions selected by each one. The complementing information introduced by the MCIs is contained in the x_3, since x_1 and x_2 are of the same nature as in Fig. 5. Different colours have been used in order to distinguish between the three classes. The x_3 value allows the identification of MCI-c subjects, which are likely to have a positive value in the eigenbrain 3 projection, corresponding to a soft glucose hypometabolism in temporal lobe areas. Mild AD subjects are more likely characterized by hypometabolism in posterior cingulate gyri and precunei and parietal lobes, while NC appear to have positive x_3 and x_2 values, corresponding to normal metabolism pattern in the previous mentioned areas.

On the other hand, the values of x_2 and x_3 for stable MCI subjects do not take a preferred sign. This fact allows to identify those MCI converters from the stable ones, but is not a sign of classification failure, since relevant classification information is contained in eigenbrain 4 and its projection. Moreover, MCI subject class may incorporate Fronto-Temporal Dementia affected subjects, or reversion cases, that is, subjects that turn out to be controls in the 2 years follow up. Therefore, the spread of x_2 and x_3 for stable MCI suggests only a portion of MCI images present a characteristic AD pattern. On the contrary, MCI-c present similar patterns as those of early AD. Important is to add that eigenbrain 1 accumulate the 99.83% of the variance, while eigenbrains 2, 3 and 4 explain the remaining 0.11%, 0.04% and 0.03%. In other words, differences that allow to indentificate AD and MCI patients from NC are very subtle, and those of MCI-c are minimal.

In Table 1, the case of maximum number of components is reported in detail, because of its highest classification rates. The voxel-as-features (VAF) approach results are reported as reference, since different studies have concluded that this method is, at least, comparable to visual assessments performed by experts [68,67]. The use of kernels in combination with ICA or PCA leads to better performance, due to the small dimensionality of the feature space, being ICA slightly more efficient than PCA. In all the groups, both methods outperform the VAF approximation [68,67], as well as PCA-fisher [35,42,43] using the same number of components.

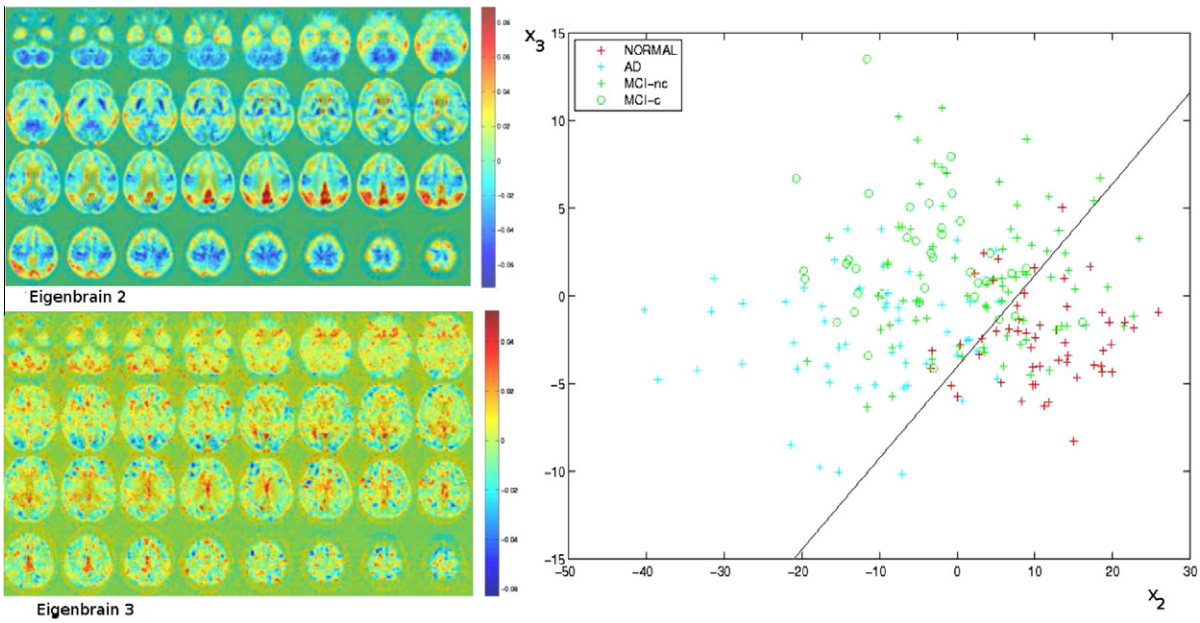

Fig. 7. Linear kernel SVM classifier and some components of training vectors x_2 and x_3 (right) obtained projecting each image onto the third and fourth eigenimages (left, transaxial slices). MCI converters (MCI-c) are represented by circles in contrast with non-converters (MCI-nc).

4. Discussion

The use of imbalanced training datasets of groups 1 and 3 causes two major problems. Firstly, the use of performance parameters as specificity or sensitivity is inappropriate, since they are sample prevalence dependent. For instance, a low specificity value in a imbalanced dataset may not reflect a high false negative rate. This problem is solved using other sample prevalence independent parameters, as positive or negative likelihoods. Secondly, classifiers are limited in their performance when dealing with imbalanced datasets [37,75]. However, SVM have been reported to perform well with moderately imbalanced databases, as ADNI with a ratio of 1–3, even without any modification [2]. SVM is constructed using only feature vectors that are close to the boundary, i.e. the support vectors. This means that SVM is unaffected by non-noisy instances far away from the boundary even if they are huge in number. The election of SVM to manage the classification task may give response to this second problem.

Differences in the images due to factors not related to AD are an important source of variability that may affect classification, as scanner differences or formats. If the approach of [35,42,43] is taken, it is found that first two eigenbrains reflect only differences in the thalamus regions as the biggest source of variation between images, and AD related patterns are a less important source. This may not happen when working with other databases but ADNI, as in [35,42]. This problem may be tackled by extracting the eigenimage basis from intra-class mean images defined in Eq. (1). Therefore, not AD related issues are soften while maintaining AD differences. Moreover, an overfitting to a specific training database is avoided, increasing generalization capability. On the other hand, it is possible to improve PCA-f results on ADNI presented in Table 1 by increasing the number of components up to 8 and/or using some eligibility criteria for the components [43]. The price paid is the inclusion of some *a priori* knowledge about the disease for determining the selection criteria and/or the loss of image analysis support. By using PCA/ICA on Eq. (1), AD related patterns are automatically included in a very reduced number of eigenbrains, without introducing *a priori* information and guaranteeing meaningful representations: graphical displays are possible and partial compensations upon superpositions of positive–negative components are controlled.

When comparing the method presented in this work with other existing in the literature, one should also consider that testing with ADNI database excludes advanced AD affected subjects, whose disease image pattern become more marked. Introducing advanced AD patient images would simplify the classification task. ADNI database represents some ideal laboratory conditions, as subjects are recruited following a restricted clinical criteria to select only early AD related issues. Following the analysis of Fig. 6, advanced AD affected subjects would lie in the left down corner, far away from the boundary, as well as young normals would lie in the upper right corner, again far away from the boundary, thus having small effect in the support vector determination, and therefore in the hyperplane definition. The constraints imposed by ADNI increase the information near the boundary, thus allowing to test the real discrimination capability of the classifier for *early* AD.

On the other hand, ADNI patient diagnostics are not pathologically confirmed, introducing some uncertainly on the subject's labels. Using these labels, allows to test the robustness of the classifier. This should be also considered when comparing to other methods tested on autopsy confirmed AD patients, as [32], on which every classifier is expected to improve its performance. Pooled sensitivity and specificity estimates of 1989–2003 studies using FDG-PET for AD diagnosis are 86%, using visual, ROIs and voxel-based methods [56,52, and references therein], values also outperformed in group 2 results within this work.

4.1. Conclusions and outlook

For purposes of brain image analysis and *early* AD diagnosis assistance by pattern disease discrimination, we proposed a novel CAD based on eigenbrain decomposition, incorporating several key ideas: dimension reduction with projection into a discriminative subspace, feature extraction with representative vector demixing by means of PCA/ICA, and classification with kernel-SVM. This fully automatic construction allows to identify characteristic AD patterns on brain images, while discarding other sources of variability, dealing with the high dimensionality problem in PET images with a low computational cost, and in clinically relevant times. The final decision is supported by an image analysis that emulates the methodology followed by experts in the visual assessment of images, by pinpointing the same specific brain regions, providing potential useful information for physicians when facing a new patient's diagnostic. The robustness and classification capabilities of the method are tested on 401 ADNI patients yielding to 88.24% accuracy in identifying mild AD, with 88.64% specificity, and 87.70% sensitivity. The method also allows the differentiation between those MCI converters after 2 years and the stable ones, incluiding also the identification of characteristic AD patterns in stable MCI subjects.

The approach to the CAD system on 401 subject images from ADNI data presented in this work can be complemented with the remaining images from the database, increasing the sample size (for example MCI-c patients) for statistical analysis. In this work, the probability distribution function of the MCI converters brain image has been estimated using 45 subjects, with only 7 confirmed cases. All committee confirmed conversion are expected to be reported when ADNI project ends, making it possible to have better estimates of the MCI brain image pdf, as well as reducing the reversion cases considered as MCI. Also these tools can be used with other information provided by ADNI, as other markers or modalities, for instance ^{11}C-PIB PET or SPECT images or even to the diagnosis assistance of other diseases.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol–Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F. Hoffman-La Roche, Schering-Plough, Synarc, Inc., and Wyeth, as well as non-profit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the US Food and Drug Administration. Private sector contributions to ADNI are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

This work was partly supported by the MICINN under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, TEC2007-68030-C02-01 and HD2008-0029 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Project TIC-02566.

Appendix A

A.1. PCA

For applying PCA, it is needed firstly to extract the average of the image set to each 3D brain images $\mathbf{t}_i$, producing a new set $\mathbf{h}_i = \mathbf{t}_i - \mathbf{t}$ with $n = 1,2,\ldots,N$. On this set, a PCA transformation is composed by M orthogonal vectors $\mathbf{u}_i$, such that

$$\lambda_i = \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{u}_i^T\mathbf{h}_n\right)^2 \tag{A.1}$$

is maximum, subject to the constrain

$$\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij}, \tag{A.2}$$

where δ_{ij} is the Kronecker delta. The resulting $\mathbf{u}_i$ and λ_i are the eigenvectors and eigenvalues, respectively, of the covariance matrix:

$$\mathbf{C} = \frac{1}{N}\sum_{i=n}^{N}\mathbf{h}_n\mathbf{h}_n^T = \frac{1}{N}\mathbf{A}\mathbf{A}^T, \tag{A.3}$$

where $\mathbf{A} = [\mathbf{h}_1,\ldots,\mathbf{h}_N]$. We will refer to this orthogonal eigenvector basis $\{\mathbf{u}_i\}$, $i = 1,\ldots,M$ as eigenbrains. To obtain them, it is necessary to diagonalize the $M \times M$ covariance matrix, which for brain images would be approximately a $5 \cdot 10^5 \times 5 \cdot 10^5$ matrix. The computational complexity of the diagonalization process can be significantly reduced by diagonalizing the matrix $\widehat{\mathbf{C}} = 1/N\mathbf{A}^T\mathbf{A}$, which size is $N \times N$, with $N \ll M$ [73]. This allows to obtain N of the M eigenvectors $\mathbf{u}_n$ of $\mathbf{C}$, from the eigenvectors $\mathbf{v}_n$ of $\widehat{\mathbf{C}}$ as $\mathbf{u}_n = \mathbf{A}\mathbf{v}_n$, $n = 1,\ldots,N$. The obtained eigenbrains span a new subspace which we refer to as the "eigenbrain space".

A.2. ICA

The task in ICA is to find a solution to the noiseless blind source separation problem [6,14,7], which can be stated as follows: let $\mathbf{X}$ be an observed random vector and $\mathbf{A}$ a full rank matrix such that:

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \tag{A.4}$$

where the source signals $\mathbf{S}$ are commonly assumed to be stochastically independent: $p_s(\mathbf{s}_1,\ldots,\mathbf{s}_n) = p_{s1}(\mathbf{s}1)\ldots p_{sn}(\mathbf{s}_n)$. Thus, ICA recovers both the sources $\mathbf{s}_j$ and the mixing process using the independence assumption, that implies decorrelation of cumulants to all orders. In the linear case, the latter task consists of finding the mixing matrix $\mathbf{A}$. A popular approach is to find a demixing or separating matrix $\mathbf{W}$ so that variables $\mathbf{y}_j$ in $\mathbf{Y} = \mathbf{W}\mathbf{X}$ are estimates of $\mathbf{s}_j$ up to scaling and permutation. In the deflationary approach the sources are estimated one by one, by finding a column vector $\mathbf{w}_j$ (this will be stored as a row of $\mathbf{W}$) such that $\mathbf{y}_j = \mathbf{w}_j^T\mathbf{X}$ is an estimate of $\mathbf{s}_j$. Hence $\mathbf{W}$ is an estimate of the (pseudo) inverse of $\mathbf{A}$ up to scaling and permutation of the rows of $\mathbf{W}$.

The estimation of the independent components and the mixing matrix is done with the help of FastICA [55], which is an iterative fixed-point algorithm with the following update for $\mathbf{w}$:

$$\mathbf{w} \leftarrow E\{\mathbf{X}g(\mathbf{w}^T\mathbf{X})\} - E\{g'(\mathbf{w}^T\mathbf{X})\}\mathbf{w}, \tag{A.5}$$

where $\mathbf{w}$ is one of the rows of the demixing matrix $\mathbf{W}$, and g is the derivative of the contrast function, chosen to be a cubic polynomial. After each iteration step defined on Eq. (A.5), $\mathbf{w}$ is normalized to have unit norm, ensuring that the rows $\mathbf{w}_j$ of the demixing matrix are orthogonal.

A.3. SVM

SVM is a machine learning algorithm that separates a given set of binary labelled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyper-plane). The objective is to build a function $f : \mathbb{R}^K \rightarrow \{\pm 1\}$ using training data, consisting of K-dimensional patterns x_i and class labels y_i:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N) \in (\mathbb{R}^K \times \{\pm 1\}), \tag{A.6}$$

so that f will correctly classify new examples $(\mathbf{x}, y)$. When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques using the *kernel trick*, so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space that is mapped to a linearized higher-dimensional space [74]. In this way the decision function f can be expressed in terms of the *support vectors* only:

$$f(\mathbf{x}) = \mathrm{sign}\left\{\sum_{i=1}^{N_S} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + w_0\right\}, \tag{A.7}$$

where $K(\cdot,\cdot)$ is the kernel function, α_i is a weight constant derived form the SVM process and $\mathbf{s}_i$ are the support vectors [74]. Common kernels that are used by SVM practitioners for the non-linear feature mapping are:

- Polynomial

$$K(\mathrm{x}, \mathrm{y}) = [\gamma(\mathrm{x} \cdot \mathrm{y}) + c]^d. \tag{A.8}$$

- Radial basis function (RBF)

$$K(\mathrm{x}, \mathrm{y}) = \exp\left(-\gamma \|\mathrm{x} - \mathrm{y}\|^2\right). \tag{A.9}$$

as well as the linear kernel, in which $K(\cdot,\cdot)$ is simply a scalar product.

References

[1] G.E. Alexander, K. Chen, P. Pietrini, S.I. Rapoport, E.M. Reiman, Longitudinal PET evaluation of cerebral metabolic decline in dementia: a potential outcome measure in Alzheimer's disease treatment studies, The American Journal of Psychiatry 159 (5) (2002) 738–745. PMID: 11986126.
[2] R. Akbani, S. Kwek, N. Japkowicz, Applying support vector machines to imbalanced datasets, in: Proceedings of the 15th European Conference on Machine Learning (ECML), 2004, pp. 39–50.
[3] B. Apolloni, S. Bassis, A. Brega, Feature selection via Boolean independent component analysis, Information Sciences 179 (22) (2009) 3815–3831.
[4] J. Ashburner, K.J. Friston, Nonlinear spatial normalization using basis functions, Human Brain Mapping 7 (4) (1999) 254–266.
[5] M. Bartlett, J. Movellan, T. Sejnowski, Face recognition by independent component analysis, IEEE Transactions on Neural Networks 13 (6) (2002) 1450–1464.
[6] A.J. Bell, T.J. Sejnowski, An information–maximization approach to blind separation and blind deconvolution, Neural Computation 7 (6) (1995) 1129–1159.
[7] E. Bingham, Advances in Independent Component Analysis with Applications to Data Mining, Ph.D. Thesis, Helsinki University of Technology, 2003.
[8] H. Braak, E. Braak, diagnostic criteria for neuropathologic assessment of Alzheimer's disease, Neurobiology and Aging 18 (4) (1997) S85–S88.
[9] R. Brookmeyer, S. Gray, C. Kawas, Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset, American Journal of Public Health 88 (9) (1998) 1337–1342.
[10] X. Cao, Y. Xu, D. Chen, H. Qiao, Associated evolution of a support vector machine-based classifier for pedestrian detection, Information Sciences 179 (8) (2009) 1070–1077.
[11] T.N. Chase, N.L. Foster, P. Fedio, R. Brooks, L. Mansi, G.D. Chiro, Regional cortical dysfunction in Alzheimer's disease as determined by positron emission tomography, Annals of Neurology 15 (Suppl.) (1984) S170–S174. PMID: 6611118.
[12] R. Chaves, J. Ramírez, J.M. Górriz, M.M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting, Neuroscience Letters 461 (2009) 293–297.
[13] J.J. Claus, F. van Harskamp, M.M.B. Breteler, E.P. Krenning, I. de Koning abd, J.M. van der Cammen, A. Hofman, D. Hasan, The diagnostic value of SPECT with tc 99 m HMPAO in Alzheimer's disease. A population-based study, Neurology 44 (3) (1994) 454–461.
[14] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (3) (1994) 287–314
[15] J.L. Cummings, H.V. Vinters, G.M. Cole, Z.S. Khachaturian, Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities, Neurology 51 (Suppl. 1) (1998) S2–S17.
[16] C. Davatzikos, S. Resnick, X. Wu, P. Parmpi, C. Clark, Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI, NeuroImage 41 (4) (2008) 1220–1227.
[17] M.J. de Leon, S.H. Ferris, A.E. George, B. Reisberg, D.R. Christman, I.I. Kricheff, A.P. Wolf, Computed tomography and positron emission transaxial tomography evaluations of normal aging and Alzheimer's disease, Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism 3 (3) (1983) 391–394. PMID: 6603463.
[18] A. Drzezga, N. Lautenschlager, H. Siebner, M. Riemenschneider, F. Willoch, S. Minoshima, M. Schwaiger, A. Kurz, Cerebral metabolic changes accompanying conversion of mild cognitive impairment into Alzheimer's disease: a PET follow-up study, European Journal of Nuclear Medicine and Molecular Imaging 30 (8) (2003) 1104–1113. PMID: 12764551.
[19] R. Duara, C. Grady, J. Haxby, M. Sundaram, N.R. Cutler, L. Heston, A. Moore, N. Schlageter, S. Larson, S.I. Rapoport, Positron emission tomography in Alzheimer's disease, Neurology 36 (7) (1986) 879.
[20] D. Evans, H. Funkenstein, M. Albert, P. Scherr, N. Cook, M. Chown, L. Hebert, C. Hennekens, J. Taylor, Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported, Journal of the American Medical Association 262 (18) (1989) 2551–2556.
[21] Y. Fan, N. Batmanghelich, C.M. Clark, C. Davatzikos, Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline, NeuroImage 39 (4) (2008) 1731–1743.
[22] F. Fink, K. Worle, P. Gruber, A.M. Tome, J.M. Gorriz, C.G. Puntonet, E.W. Lang, Ica analysis of retina images for glaucoma classification, in: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, pp. 4664–4667.

[23] N.L. Foster, T.N. Chase, P. Fedio, N.J. Patronas, R.A. Brooks, G.D. Chiro, Alzheimer's disease: focal cortical changes shown by positron emission tomography, Neurology 33 (8) (1983) 961.
[24] N.L. Foster, T.N. Chase, L. Mansi, R. Brooks, P. Fedio, N.J. Patronas, G.D. Chiro, Cortical abnormalities in Alzheimer's disease, Annals of Neurology 16 (6) (1984) 649–654. PMID: 6335378.
[25] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, W. Penny, Statistical Parametric Mapping: The Analysis of Functional Brain Images, Academic Press, 2007.
[26] J.M. Górriz, A. Lassl, J. Ramírez, D. Salas-Gonzalez, C.G. Puntonet, E.W. Lang, Automatic selection of ROIs in functional imaging using Gaussian mixture models, Neuroscience Letters 460 (2) (2009) 108–111.
[27] J.M. Górriz, J. Ramírez, A. Lassl, D. Salas-Gonzalez, E.W. Lang, C.G. Puntonet, I. Álvarez, M. López, M. Gómez-Río, Automatic computer aided diagnosis tool using component-based svm, in: 2008 IEEE Nuclear Science Symposium Conference Record, 2008, pp. 4392–4395.
[28] J.M. Górriz, J. Ramírez, J.C. Segura, C.G. Puntonet, An effective cluster-based model for robust speech detection and speech recognition in noisy environments, Journal of the Acoustical Society of America 120 (1) (2006) 470–481.
[29] J.M. Górriz, C.G. Puntonet, M. Salmerón, F.R. Ruiz, Hybridizing genetic algorithms with ica in higher dimension, Lecture Notes in Computer Science 3195 (2004) 414–421.
[30] J.M. Gorriz, C.G. Puntonet, M. Salmeron, J.J.G. de la Rosa, A new model for time-series forecasting using radial basis functions and exogenous data, Neural Computing and Applications 13 (2) (2004) 101–111.
[31] K. Herholz, D. Perani, C. Morris, The dementias, Informa Health Care (2006).
[32] R. Higdon, N.L. Foster, R.A. Koeppe, C.S. DeCarli, W.J. Jagust, C.M. Clark, N.R. Barbas, S.E. Arnold, R.S. Turner, J.L. Heidebrink, S. Minoshima, A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging, Statistics in Medicine 23 (2004) 315–326.
[33] J.M. Hoffman, K.A. Welsh-Bohmer, M. Hanson, B. Crain, C. Hulette, N. Earl, R.E. Coleman, FDG PET imaging in patients with pathologically verified dementia, Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine 41 (11) (2000) 1920–1928. PMID: 11079505.
[34] V. Ibañez, P. Pietrini, G.E. Alexander, M.L. Furey, D. Teichberg, J.C. Rajapakse, S.I. Rapoport, M.B. Schapiro, B. Horwitz, Regional glucose metabolic abnormalities are not the result of atrophy in Alzheimer's disease, Neurology 50 (6) (1998) 1585–1593. PMID: 9633698.
[35] I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-González, M. López, C.G. Puntonet, F. Segovia, Alzheimer's diagnosis using eigenbrains and support vector machines, IET Electronics Letters 45 (7) (2009) 342–343.
[36] K. Ishii, A.K. Kono, H. Sasaki, N. Miyamoto, T. Fukuda, S. Sakamoto, E. Mori, Fully automatic diagnostic system for early- and late-onset mild Alzheimer's disease using FDG PET and 3D-SSP, European Journal of Nuclear Medicine and Molecular Imaging 33 (5) (2006) 575–583.
[37] N. Japkowicz, The class imbalance problem: significance and strategies, in: Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI), vol. 1, 2000, pp. 111–117.
[38] G. Karas, P. Scheltens, S. Rombouts, P. Visser, R. van Schijndel, N. Fox, F. Barkhof, Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease, NeuroImage 23 (2) (2004) 708–716.
[39] D. Kogure, H. Matsuda, T. Ohnishi, T. Asada, M. Uno, T. Kunihiro, S. Nakano, M. Takasaki, Longitudinal evaluation of early Alzheimer disease using brain perfusion SPECT, The Journal of Nuclear Medicine 41 (7) (2000) 1155–1162.
[40] J.B. Langbaum, K. Chen, W. Lee, C. Reschke, D. Bandy, A.S. Fleisher, G.E. Alexander, N.L. Foster, M.W. Weiner, R.A. Koeppe, W.J. Jagust, E.M. Reiman, Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's disease neuroimaging initiative (ADNI), NeuroImage 45 (4) (2009) 1107–1116.
[41] A. Lassl, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, C.G. Puntonet, E.W. Lang, Clustering approach for the classification of spect images, in: 2008 IEEE Nuclear Science Symposium Conference Record, 2008, pp. 5345–5348.
[42] M. López, J. Ramírez, J.M. Górriz, D. Salas-González, I.A. Illan, F. Segovia, C.G. Puntonet, Automatic tool for the Alzheimer's disease diagnosis using pca and Bayesian classification rules, IET Electronics Letters 45 (8) (2009) 389–391.
[43] M. López, J. Ramírez, J.M. Górriz, D. Salas-González, I.A. Illan, F. Segovia, C.G. Puntonet, Principal component analysis-based techniques and supervised classification schemes for the early detection of the Alzheimer's disease, Neurocomputing (in press), doi:10.1016/j.neucom.2010.06.025.
[44] M.M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, Svm-based cad system for early detection of the Alzheimer's disease using kernel pca and lda, Neuroscience Letters 464 (2009) 233–238.
[45] P. Markiewicz, J. Matthews, J. Declerck, K. Herholz, Robustness of multivariate image analysis assessed by resampling techniques and applied to FDG-PET scans of patients with Alzheimer's disease, NeuroImage 46 (2) (2009) 472–485.
[46] S. Makeig, A.J. Bell, T. ping Jung, T.J. Sejnowski, Independent component analysis of electroencephalographic data, in: Advances in Neural Information Processing Systems, vol. 8, MIT, 1996, pp. 145–151.
[47] E.G. McGeer, R.P. Peppard, P.L. McGeer, H. Tuokko, D. Crockett, R. Parks, H. Akiyama, D.B. Calne, B.L. Beattie, R. Harrop, 18Fluorodeoxyglucose positron emission tomography studies in presumed Alzheimer cases, including 13 serial scans, The Canadian Journal of Neurological Sciences 17 (1) (1990) 1–11. PMID: 2311010.
[48] C. Messa, D. Perani, G. Lucignani, A. Zenorini, F. Zito, G. Rizzo, F. Grassi, A. Del Sole, M. Franceschi, M.C. Gilardi, F. Fazio, High-resolution technetium-99 m-HMPAO SPECT in patients with probable Alzheimer's disease: comparison with fluorine-18-FDG PET, Journal of Nuclear Medicine 35 (2) (1994) 210–216.
[49] S. Minoshima, N. Foster, D. Kuhl, Posterior cingulate cortex in Alzheimer's disease, The Lancet 344 (8926) (1994) 895.
[50] S. Minoshima, K.A. Frey, R.A. Koeppe, N.L. Foster, D.E. Kuhl, A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET, Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine 36 (7) (1995) 1238–1248. PMID: 7790950.
[51] S. Minoshima, B. Giordani, S. Berent, K.A. Frey, N.L. Foster, D.E. Kuhl, Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease, Annals of Neurology 42 (1) (1997) 85–94. PMID: 9225689.
[52] L. Mosconi, Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. FDG-PET studies in MCI and AD, European Journal of Nuclear Medicine and Molecular Imaging 32 (4) (2005) 486–510. PMID: 15747152.
[53] L. Mosconi, W.H. Tsui, K. Herholz, A. Pupi, A. Drzezga, G. Lucignani, E.M. Reiman, V. Holthoff, E. Kalbe, S. Sorbi, J. Diehl-Schmid, R. Perneczky, F. Clerici, R. Caselli, B. Beuthien-Baumann, A. Kurz, S. Minoshima, M.J. de Leon, Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias, Journal of Nuclear Medicine 49 (3) (2008) 390–398.
[54] F. Nobili, D. Salmaso, S. Morbelli, N. Girtler, A. Piccardo, A. Brugnolo, B. Dessi, S.A. Larsson, G. Rodriguez, M. Pagani, Principal component analysis of fdg pet in amnestic mci, European Journal of Nuclear Medicine and Molecular Imaging 35 (12) (2008) 2191–2202.
[55] E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation 9 (1997) 1483–1492.
[56] M.B. Patwardhan, D.C. McCrory, D.B. Matchar, G.P. Samsa, O.T. Rutschmann, Alzheimer disease: operating characteristics of PET- a meta-analysis, Radiology 231 (1) (2004) 73–80.
[57] T. Korenius, J. Laurikkala, M. Juhola, On principal component analysis, cosine and euclidean measures in information retrieval, Information Sciences 177 (22) (2007) 4893–4905.
[58] J. Ramírez, J.M. Górriz, A. Romero, A. Lassl, D. Salas-Gonzalez, M. López, I. Alvarez, M. Gómez-Río, A. Rodríguez, Computer aided diagnosis of Alzheimer type dementia combining support vector machines and discriminant set of features, Information Sciences (in press), doi:10.1016/j.ins.2009.05.012.
[59] J. Ramírez, J.M. Górriz, M. Gómez-Río, A. Romero, R. Chaves, A. Lassl, A. Rodríguez, C.G. Puntonet, F. Theis, E. Lang, Effective emission tomography image reconstruction algorithms for SPECT data, Lecture Notes in Computer Science 5101 (2008) 741–748.
[60] J. Ramírez, P. Yélamos, J.M. Górriz, J.C. Segura, SVM-based speech endpoint detection using contextual speech features, Electronics Letters 42 (7) (2006) 877–879.
[61] K. Ritchie, S. Lovestone, The dementias, The Lancet 360 (9347) (2002) 1759–1766.

[62] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, M.M. López, I. Álvarez, F. Segovia, M. Gomez-Rio, Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines, Neuroscience Letters 461 (2009) 60–64.
[63] E. Salmon, N. Kerrouche, D. Perani, F. Lekeu, V. Holthoff, B. Beuthien-Baumann, S. Sorbi, C. Lemaire, F. Collette, K. Herholz, On the multivariate nature of brain metabolic impairment in Alzheimer's disease, Neurobiology of Aging 30 (2) (2009) 186–197.
[64] N. Scarmeas, C.G. Habeck, E. Zarahn, K.E. Anderson, A. Park, J. Hilton, G.H. Pelton, M.H. Tabert, L.S. Honig, J.R. Moeller, D.P. Devanand, Y. Stern, Covariance pet patterns in early Alzheimer's disease and subjects with cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance, NeuroImage 23 (1) (2004) 35–45.
[65] D.H. Silverman, G.W. Small, C.Y. Chang, Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome, Journal of the American Medical Association 286 (17) (2001) 2120–2127.
[66] S. Ng, V.L. Villemagne, S. Berlangieri, S.T. Lee, M. Cherk, S.J. Gong, U. Ackermann, T. Saunder, H. Tochon-Danguy, G. Jones, C. Smith, G. O'Keefe, C.L. Masters, C.C. Rowe, Visual assessment versus quantitative assessment of 11c-pib pet and 18f-fdg pet for detection of Alzheimer's disease, Journal of Nuclear Medicine 48 (2007) 547–552.
[67] J. Stoeckel, N. Ayache, G. Malandain, P.M. Koulibaly, K.P. Ebmeier, J. Darcourt, Automatic classification of spect images of Alzheimer's disease patients and control subjects, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI, Lecture Notes in Computer Science, vol. 3217, Springer, 2004, pp. 654–662.
[68] J. Stoeckel, G. Malandain, O. Migneco, P.M. Koulibaly, P. Robert, N. Ayache, J. Darcourt, Classification of spect images of normal subjects versus images of Alzheimer's disease patients, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI, Lecture Notes in Computer Science, vol. 2, Springer, 2001, pp. 666–674.
[69] J. Sui, T. Adali, G.D. Pearlson, V.D. Calhoun, An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques, NeuroImage 46 (1) (2009) 73–86.
[70] P.R. Talbot, J.J. Lloyd, J.S. Snowden, D. Neary, H.J. Testa, A clinical role for 99mTc-HMPAO SPECT in the investigation of dementia?, Journal of Neurology, Neurosurgery and Psychiatry 64 (3) (1998) 306–313
[71] S.J. Teipel, R. Stahl, O. Dietrich, S.O. Schoenberg, R. Perneczky, A.L. Bokde, M.F. Reiser, H.-J. Möller, H. Hampel, Multivariate network analysis of fiber tract integrity in Alzheimer's disease, NeuroImage 34 (3) (2007) 985–995.
[72] F.J. Theis, P. Gruber, I.R. Keck, E.W. Lang, Functional mri analysis by a novel spatiotemporal ica algorithm, in: International Conference on Artificial Neural Networks (ICANN), vol. 1, 2005, pp. 677–682.
[73] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[74] V.N. Vapnik, Statistical Learning Theory, John Wiley and Sons, Inc., New York, 1998.
[75] F. Vilariño, P. Spyridonos, J. Vitriá, P. Radeva, Experiments with SVM and stratified sampling with an imbalanced problem: detection of intestinal contractions, in: Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, Springer-Verlag, 2005, pp. 783–791.
[76] R.P. Woods, S.T. Grafton, C.J. Holmes, S.R. Cherry, J.C. Mazziotta, Automated image registration: I. General methods and intrasubject, intramodality validation, Journal of Computer Assisted Tomography 22 (1) (1998) 139–152.