Neurocomputing 74 (2011) 1260–1271

Contents lists available at ScienceDirect

Neurocomputing

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

Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease

M. López [a], J. Ramírez [a], J.M. Górriz [a,*], I. Álvarez [a], D. Salas-Gonzalez [a], F. Segovia [a], R. Chaves [a], P. Padilla [a], M. Gómez-Río [b], the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Department of Signal Theory, Networking and Communications, University of Granada, Spain
[b] Department of Nuclear Medicine, Hospital Universitario Virgen de las Nieves, Granada, Spain

ARTICLE INFO

Available online 16 October 2010

Keywords:
Alzheimer's disease
PCA
LDA
Supervised learning
Computer-aided diagnosis system

ABSTRACT

In Alzheimer's disease (AD) diagnosis process, functional brain image modalities such as Single-Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) have been widely used to guide the clinicians. However, the current evaluation of these images entails a succession of manual reorientations and visual interpretation steps, which attach in some way subjectivity to the diagnostic. In this work, a complete computer aided diagnosis (CAD) system for an automatic evaluation of the neuroimages is presented. Principal component analysis (PCA)-based methods are proposed as feature extraction techniques, enhanced by other linear approaches such as linear discriminant analysis (LDA) or the measure of the Fisher discriminant ratio (FDR) for feature selection. The final features allow to face up the so-called *small sample size* problem and subsequently they are used for the study of neural networks (NN) and support vector machine (SVM) classifiers. The combination of the presented methods achieved accuracy results of up to 96.7% and 89.52% for SPECT and PET images, respectively, which means a significant improvement over the results obtained by the classical voxels-as-features (VAF) reference approach.

1. Introduction

Alzheimer's disease (AD) is a progressive, degenerative brain disorder that gradually destroys memory, reason, judgment, language, and eventually the ability to carry out even the simplest tasks. Recently, scientists have begun to do research on diagnosing AD with different kinds of brain imaging, trying to diagnose this dementia in its early stage, when the application of the treatment is more effective. Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) scans are two types of non-invasive (i.e., no surgery is required) tests that have been widely used in the AD diagnosis. For both PET and SPECT tests, a small amount of radioactive pharmaceutical is injected into the patient and emission detectors are placed on the brain, providing functional information about the brain activity. However, despite these useful imaging techniques, early detection of AD still remains a challenge since conventional evaluation of these scans often relies on manual reorientation, visual reading and semiquantitative analysis.

Several approaches have been recently proposed in the literature aiming at providing an automatic tool that guides the clinician in the AD diagnosis process [1–5]. These approaches can be categorized into two types: univariate and multivariate approaches. The first family includes statistical parametric mapping (SPM) [6] and its numerous variants. SPM consists of doing a voxelwise statistical test, comparing the values of the image under study to the mean values of the group of normal images. Subsequently the significant voxels are inferred by using random field theory. It was not developed specifically to study a single image, but for comparing groups of images. The latter family is based on the analysis of the images, feature extraction and posterior classification in different classes. Among these techniques, we can find the classical voxels-as-features (VAF) approach for SPECT images [1], which considers the gray image levels as features. Input patterns can be viewed as points in the multidimensional space defined by the input feature measurements. The main problem to be faced up by these techniques is the well-known *small sample size* problem, that is, the number of available samples is much lower than the number of features used in the training step.

Principal component analysis (PCA) corresponds to multivariate approaches and was already applied to functional brain images in [6] in a descriptive fashion, where the impossibility of using this transformation to make any statistical inference is highlighted. However, in this work, a new PCA-based approach is used in

* Corresponding author. Tel./fax: +34 958 243271.
E-mail address: gorriz@ugr.es (J.M. Górriz).

[1] Part of the data used in the preparation of this article were obtained from 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_Manuscript_Citations.pdf).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.06.025

combination with supervised learning methods, which in turn solves the *small sample size* problem since the dimension of the feature space undergoes a significant reduction. Other techniques like linear discriminant analysis (LDA) or the Fisher discriminant ratio (FDR) are combined with PCA to obtain the final features to be used in a classification step.

The task of the supervised learner is to predict the class of the input object after having seen a number of training examples. In this work, support vector machines (SVMs) and neural networks (NN) are trained on the features extracted from the neurological images, and their performances in the classification task we are dealing with are finally discussed.

2. Databases description

2.1. PET

Data used in the preparation of this article were obtained from 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), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (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 Center 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 U.S. and Canada. For more information see www.adni-info.org.

Baseline Fludeoxyglucose (18^F-FDG) PET data from 219 ADNI participants, acquired from Siemens, General Electric (GE), and Philips PET scanners, were collected from the ADNI Laboratory on NeuroImaging (LONI, University of California, Los Angeles). PET data acquired from the Siemens HRRT and BioGraph HiRez scanners were excluded from the primary analysis due to differences in the pattern of FDG uptake. Participants enrollment was conditioned to some eligibility criteria. General inclusion/exclusion criteria—based on measures of disease severity, such as the Mini-Mental State Exam (MMSE) or Clinical Dementia Rating (CDR) were as follows:

- NORMAL control subjects: MMSE scores between 24 and 30 (inclusive), 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 enrollment of normals under the age of 70.
- MCI subjects: MMSE scores between 24 and 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, the 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 and 26 (inclusive), CDR of 0.5 or 1.0, and meets NINCDS/ADRDA [7] criteria for probable AD.

Therefore, FDG PET data were separated into three different classes: Normal Control (NC), Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) images. The patients of this study were divided into 53 AD (age range: 77.2–7.2 (mean–standard deviation)), 114 MCI (age range: 75.1–7.4), and 52 NC (age range: 76.5–4.8). For posterior analysis, the data were arranged into three different groups, in order to label the data into only two different classes:

- Group 1: All the database images are considered. Both AD and MCI patients are labeled 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. SPECT

A SPECT database consisting of 91 patients is also used to evaluate the proposed methods. These images were taken with a PRISM 3000 machine and were reconstructed from projection data by filtered backprojection (FBP) in combination with a Butterworth noise filter.

The SPECT database was initially labeled by experienced clinicians of the Virgen de las Nieves Hospital (Granada, Spain) as NORMAL for subjects without any symptoms of the disease and AD to refer to possible (AD 1), probable (AD 2) or certain (AD 3) patients. We combine the latter three labels and only use two classes: NOR and AD. In total, the database consists of 91 patients: 41 NOR and 50 AD. Table 1 shows other demographic details of the SPECT database.

3. Image preprocessing

All the images were normalized through a general affine model, with 12 parameters [8] using the SPM5 software [9]. After the affine normalization, the resulting image was registered using a more complex non-rigid spatial transformation model. The non-linear deformations to the Montreal Neurological Imaging (MNI) Template were parameterized by a linear combination of the

Table 1
Demographic details of the SPECT images data set.

Label	#Samples	Sex (M/F) (%)	Mean age [range/standard deviation] at scan time
NORMAL	41	32.95/12.19	71.51 [46–85/7.99]
AD 1	27	10.97/18.29	65.29 [23–81/13.36]
AD 2	19	13.41/9.76	65.73 [46–86/8.25]
AD 3	4	0/2.43	76 [69–83/9.90]

AD 1 = possible AD, AD 2 = probable AD, AD 3 = certain AD.

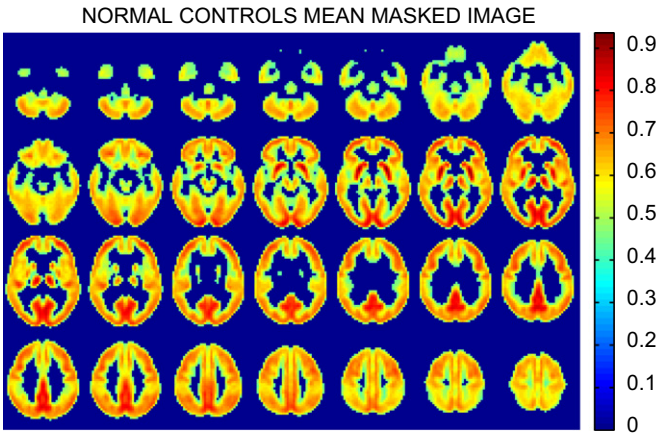


Fig. 1. Mask application to discard voxels corresponding to low activation regions.

lowest-frequency components of the 3D cosine transform bases [10]. A small-deformation approach was used, and regularization was by the bending energy of the displacement field, ensuring that the voxels in different images refer to the same anatomical positions in the brains [11]. The complete registration procedure gives rise to a $69 \times 95 \times 79$ voxel-sized representation of each sample. In order to improve the computational efficiency of the developed algorithms, the dimension of these volumes was reduced to $34 \times 47 \times 39$ by decimating the original 3D volume by a $2 \times 2 \times 2$ factor.

After the spacial normalization, an intensity normalization was required in order to perform direct comparisons between different subjects images. The intensity of the images was normalized to a value *Imax*, obtained 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.

4. Voxel selection

The subsampled brain volumes are made up of $34 \times 47 \times 39 \sim 6 \times 10^4$ voxels. The complete set of these voxels may contain non-useful or redundant information for our concerning classification task, and therefore it is desired to be removed in order to develop efficient classification algorithms. In the feature extraction step (see Section 5) only the meaningful information regarding the detection of the AD will be extracted from the images in a reduced feature space. However, before going on the feature extraction, a previous stage in which voxels of interest are selected may speed up the feature extraction process and make it more efficient. Brain regions characterized by low activity in normal controls as well as voxels located at the borders outside the brain are not considered for the next steps. A mask is constructed by averaging all the normal subjects and selecting the voxels with intensity level above 50% the maximum intensity. This step reduces the number of voxels to $\sim 2 \times 10^4$. Fig. 1 shows the voxels selected after the mask application on the ADNI images. Other methods for selecting voxels of interest are proposed in [12].

5. Feature extraction

After the preprocessing steps, a $34 \times 47 \times 39$ voxel-sized brain representation for each subject is obtained. Let us consider a set of N such brain volumes rearranged into n-dimensional vector forms $\{\mathbf{I}_1,\mathbf{I}_2,\ldots,\mathbf{I}_N\}$, and assume that each image belongs to one of c classes $\{\omega_1,\omega_2,\ldots,\omega_c\}$. If the mask explained above is applied, the input feature space dimension becomes reduced to $n \sim 2 \times 10^4$. This number of features is still much larger than the number of available samples, i.e. $n \gg N$, which is known as the *small sample size* problem. In order to face this problem up, we first apply compression techniques to the images, so the dimension is drastically reduced to a lower dimension $m \sim N$.

5.1. Principal component analysis

Principal component analysis (PCA) [13] has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis—from neuroscience to computer graphics—because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it. This technique and other PCA-based methods have been successfully applied for different image classification purposes [14,15], and specifically for neuroimage classification problems [16,17].

PCA generates an orthonormal basis vector that maximizes the scatter of all the projected samples. After the preprocessing steps, the n remaining voxels for each subject are rearranged into a vector form. Let $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N]$ be the sample set of these vectors, where N is the number of patients. After normalizing the vectors to unity norm and subtracting the grand mean, a new vector set $\mathbf{Y}=[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$ is obtained, where each $\mathbf{y}_i$ represents an n-dimensional normalized vector, $\mathbf{y}_i=(y_{i1},y_{i2},\ldots,y_{in})^t$, $i=1,2,\ldots,N$. The covariance matrix of the normalized vectors set is defined as

$$\mathbf{\Sigma}_Y = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i\mathbf{y}_i^t = \frac{1}{N}\mathbf{Y}\mathbf{Y}^t \tag{1}$$

and the eigenvector and eigenvalue matrices $\mathbf{\Phi}$, $\mathbf{\Lambda}$ are computed as

$$\mathbf{\Sigma}_Y\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda} \tag{2}$$

Note that $\mathbf{Y}\mathbf{Y}^t$ is an $n \times n$ matrix while $\mathbf{Y}^t\mathbf{Y}$ is an $N \times N$ matrix. If the sample size N is much smaller than the dimensionality n, then diagonalizing $\mathbf{Y}^t\mathbf{Y}$ instead of $\mathbf{Y}\mathbf{Y}^t$ reduces the computational complexity [15]

$$(\mathbf{Y}^t\mathbf{Y})\mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Lambda}_1 \tag{3}$$

$$\mathbf{T} = \mathbf{Y}\mathbf{\Psi} \tag{4}$$

where $\mathbf{\Lambda}_1 = diag\{\lambda_1,\lambda_2,\ldots,\lambda_N\}$ and $\mathbf{T}=[\mathbf{\Phi}_1,\mathbf{\Phi}_2,\ldots,\mathbf{\Phi}_N]$. Derived from the *eigenface* concept [15], and due to its still brain-like appearance, the eigenvectors or principal components (PCs) $\mathbf{\Phi}_i$, $i=1,\ldots,N$ of the covariance matrix are called *eigenbrains* [18]. Fig. 2 shows the first and second eigenbrains obtained from the ADNI database when AD and NORMAL subjects are considered in the solution of the eigenvalue problem.

5.2. Linear discriminant analysis

Since the learning set is labeled, it makes sense to use this information to build a more reliable method for reducing the dimensionality of the feature space. Here we argue that using class specific linear methods for dimensionality reduction and simple classifiers in the reduced feature space, one may get better recognition rates than with other multivariate approaches. LDA [19] is an example of a class specific method, in the sense that it tries to "shape" the scatter in order to make it more reliable for classification. Let the between-class scatter matrix be defined as

$$\mathbf{S}_B = \sum_{i=1}^{c} N_i(\boldsymbol{\mu}_i-\boldsymbol{\mu})(\boldsymbol{\mu}_i-\boldsymbol{\mu})^T \tag{5}$$

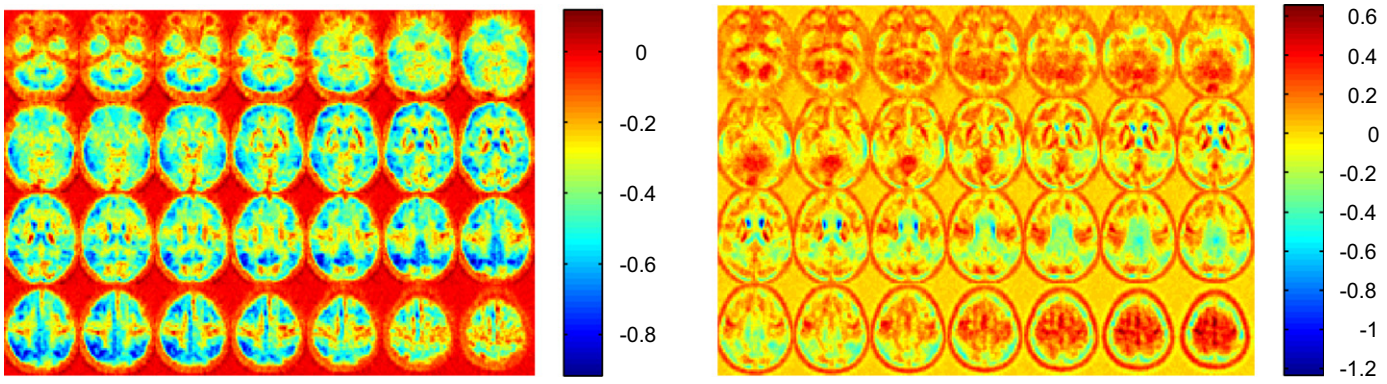


Fig. 2. First (left) and second (right) Eigenbrains extracted from the ADNI database. They represent the principal components where original images will be projected onto to obtain a dimension reduction.

and the within-class scatter matrix be defined as

$$\mathbf{S}_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \boldsymbol{\mu}_i)(\mathbf{x}_k - \boldsymbol{\mu}_i)^T \tag{6}$$

where $\boldsymbol{\mu}_i$ is the mean image of class ω_i, and N_i is the number of samples in class ω_i. If $\mathbf{S}_W$ is nonsingular, LDA chooses the optimal projection $\mathbf{W}_{opt}$ as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|} = [\mathbf{w}_1 \mathbf{w}_2 \ldots \mathbf{w}_l] \tag{7}$$

where $\{\mathbf{w}_i | i = 1,2,\ldots,l\}$ is the set of generalized eigenvectors of $\mathbf{S}_B$ and $\mathbf{S}_W$ corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1,2,\ldots,l\}$, i.e.,

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i \tag{8}$$

Note that there are at most $c-1$ nonzero generalized eigenvalues, and so an upper bound on l is $c-1$, where c is the number of classes [20].

In the image classification problem, one is confronted with the difficulty that the within-class scatter matrix $\mathbf{S}_W \in \mathbb{R}^{n \times n}$ is always singular. This is a consequence of the fact that the rank of $\mathbf{S}_W$ is at most $N-c$, and, in general, the number of images in the learning set N is much smaller than the number of selected features in each image n. This means that it is possible to choose the matrix W_{opt} such that the within-class scatter of the projected samples can be made exactly zero. In order to overcome the drawback of a singular $\mathbf{S}_W$, LDA is usually applied after the PCA transform. Thus, PCA reduces the dimension of the feature space to $m \leq N-c$, and then, the standard LDA transform is applied to reduce the dimension to $l = c-1$ [21].

6. Feature selection

6.1. Fisher discriminant ratio

A description of the databases in terms of variability can be understood by computing the importance of each projection eigenvector or PC. For example, Fig. 3(a) shows the contribution of the first 10 PCs when the whole SPECT database is used to compute the PCA projection axis. The obtained PCs from Eq. (4) are sorted out in decreasing order according to their associated eigenvalue λ. As expected, the first PC holds the maximum information which explains more than 20% of the total variance. However, Fig. 3(b) reveals the actual discrimination power of each

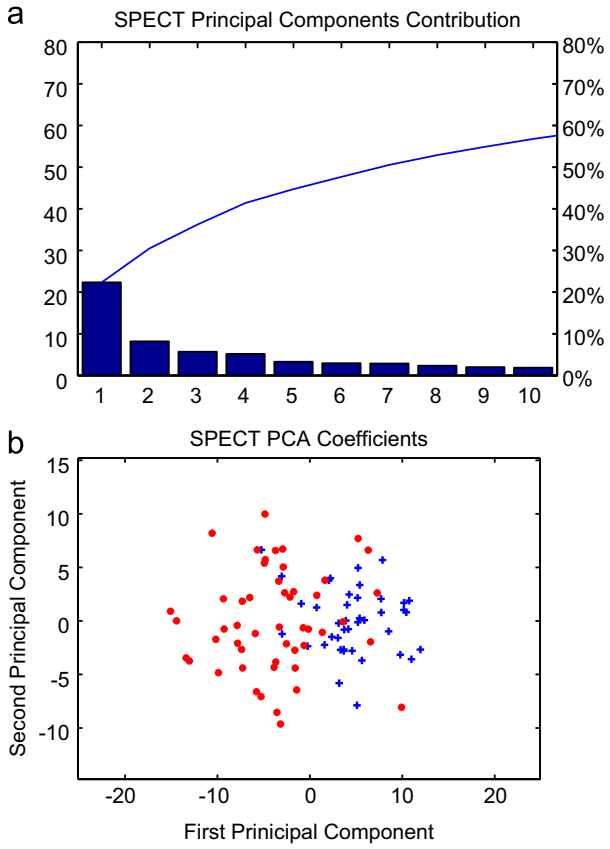


Fig. 3. (a) The percentage of total variance contained in the first 10 PCs for the SPECT database and (b) coefficients after projecting the data onto the first two PCs.

PC. Looking at the first coefficient, we can roughly say that values greater than 0 correspond to NORMAL subjects (represented as blue crosses) whereas negative values are associated to AD patients (depicted as red asterisks). This fact reveals that the variability information held in this coefficient is useful in distinguishing AD from NORMAL subjects. However, this does not occur for the second coefficient, whose 8% of explained variability does not seem to be related to the AD, since it takes values contained in the same range for both classes.

This fact motivates us to sort out the obtained PCA coefficients by using other more useful measure for classification purposes. We propose the Fisher discriminant ratio (FDR) as criteria for rearranging the obtained coefficients instead of their associated eigenvalue. This makes sense since the FDR takes into account the

information of the labels preassigned to the data. It is defined as follows:

$$FDR = \frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2} \tag{9}$$

where μ_i and σ_i denote the i-th class within class mean value and variance, respectively. Thus, in the classification stage, the training set is used to compute the PCs and the FDR values of the resulting PCA projection coefficients are computed. The test sample is projected onto the computed PCs, and the obtained test coefficients are rearranged in decreasing order according to the previous FDR values obtained from the train set. In some cases, this rearrangement will improve the classification results as shown in Section 8. Other metrics to determine the robustness of the PCs used for classification purposes are proposed in [22].

6.2. Slices of interest (SOIs)

PCA can be applied to the entire brain volumes in order to obtain a better description of the whole database. Theoretically the obtained PCs will hold the main differences among the subjects that make the database up, among which hopefully the differences between the two classes samples can be found. However, when dealing with such high dimensional data, PCA is sensitive to any other differences, maybe derived from the normalization process or due to the different origins of the images, in such a way that using the main eigenbrains as projection axes might contain "noisy" differences—that is, not useful for distinguishing the AD information.

This fact motivates us to apply PCA to the brain volume areas of interest. We can then search for the slices of interest (SOI) and discard the rest for the classification task. We explore the slices along the three directions and apply PCA to determine each one's discriminatory capacity. This allows us to pre-locate the most useful voxels for the classification task we are dealing with.

7. Classification

The objective of a binary classifier is to build a function $f: \mathbb{R}^n \longrightarrow \{\pm 1\}$ using training data that is, n-dimensional patterns $\mathbf{x}_i$ and class labels y_i:

$$(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_N,y_N) \in (\mathbb{R}^n \times \{\pm 1\}) \tag{10}$$

so that f will correctly classify new examples $(\mathbf{x},y)$.

7.1. Support vector machines

Support vector machines (SVM) [23] separate binary labeled training data by the hyperplane

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x}+w_0 \tag{11}$$

where $\boldsymbol{w}$ is known as the weight vector and w_0 as the threshold. This hyperplane is maximally distant from the two classes (known as the maximal margin hyperplane).

When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space. If the data are mapped to some other (possibly infinite dimensional) Euclidean space using a mapping $\Phi(\mathbf{x})$, the training algorithm only depends on the data through dot products in such an Euclidean space, i.e. on functions of the form $\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)$. If a "kernel function" K is defined such that $K(\mathbf{x}_i,\mathbf{x}_j)=\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)$, it is not necessary to know the Φ function during the training process. The two commonly used families of kernels are polynomial kernels and radial basis functions (RBF), defined as follows:

- Polynomial:

$$K(\mathbf{x}_i,\mathbf{x}_j) = [\gamma(\mathbf{x}_i\cdot\mathbf{x}_j)+c]^d \tag{12}$$

- Radial basis functions:

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i-\mathbf{x}_j\|^2) \tag{13}$$

For polynomial kernels, the case of $d=1$ is a linear kernel, and the case of $d=2$ gives a quadratic kernel, and is very commonly used.

In the test phase, an SVM is used by computing the sign of

$$f(\mathbf{x}) = \sum_{i=1}^{N_S}\alpha_i y_i \Phi(\mathbf{s}_i)\cdot\Phi(\mathbf{x})+w_0 = \sum_{i=1}^{N_S}\alpha_i y_i K(\mathbf{s}_i,\mathbf{x})+w_0 \tag{14}$$

where N_S is the number of support vectors, $\mathbf{s}_i$ are the support vectors and y_i their associated labels.

7.2. Neural networks

An artificial neural network (ANN) [24] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, processes information. ANNs can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges (with weights) are connections between neuron outputs and neuron inputs. Based on the connection pattern (architecture), ANNs can be grouped into two categories: (i) feed-forward networks, in which graphs have no loops, and (ii) recurrent (or feedback) networks, in which loops occur because of feedback connections. Different connectivities yield different network behaviors. Generally speaking, feed-forward networks are static, that is, they produce only one set of output values rather than a sequence of values from a given input. Feed-forward networks are memory-less in the sense that their response to an input is independent of the previous network state. Recurrent, or feedback, networks, on the other hand, are dynamic systems. When a new input pattern is presented, the neuron outputs are computed. Because of the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state.

Feed-forward networks often have one or more hidden layers (HL) of sigmoid neurons followed by an output layer of linear neurons as shown in Fig. 4. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.

Learning process in the ANN context can be viewed as the problem of updating network architecture and connection weights so that a network can efficiently perform a specific task. The ability of ANNs to automatically learn from examples makes them attractive and exciting. The development of the back-propagation learning algorithm for determining weights in a multilayer perceptron has made these networks the most popular among ANN researchers.

For the experiments presented in this work a feed-forward neural network with the following configuration was used:

- One hidden layer and increasing number of neurons and a linear output layer.
- Hyperbolic tangent sigmoid transfer function: $f(n)= 2/(1+\exp(-2^*n))-1$, for input layers.
- Linear transfer function: $f(n)= n$, for output layer.
- Weight and bias values are updated according to Levenberg–Marquardt optimization.

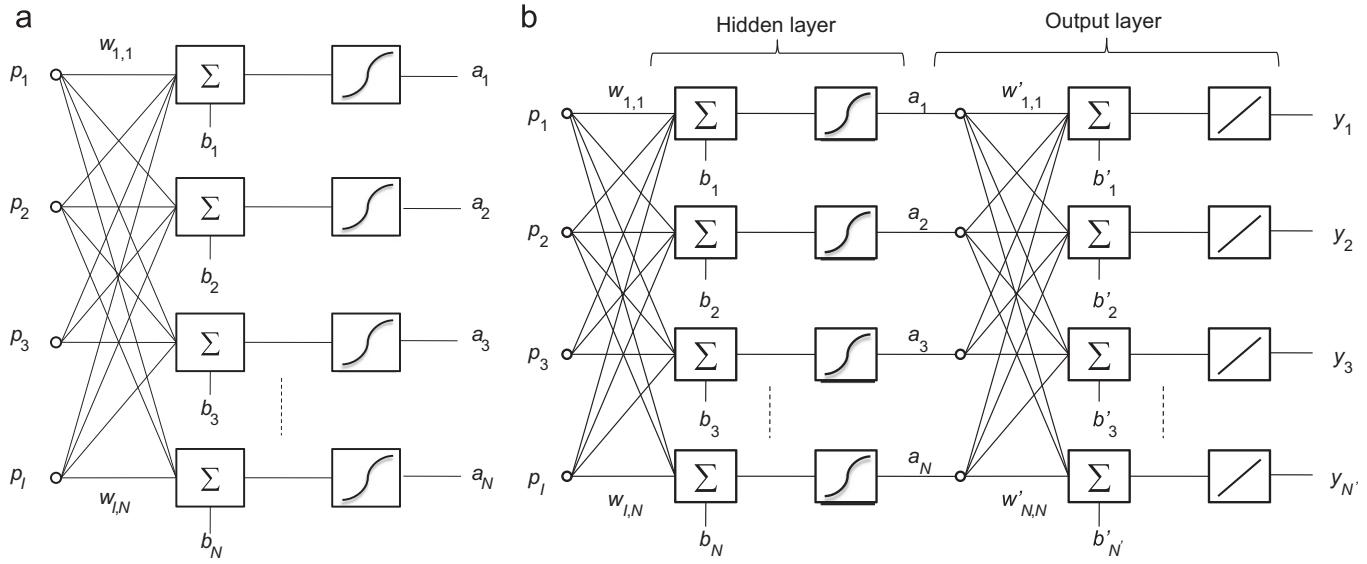


Fig. 4. Feed-forward neural network architecture with hidden layer of neurons plus linear output layer.

- Gradient descent with momentum weight and bias is used as learning function.

8. Results

In this section, different combinations of the proposed feature extraction and classification schemes are evaluated on SPECT and PET databases. The obtained results are compared with VAF method, which uses all the voxels in the brain images directly as features to train and test a linear kernel SVM [1]. All the experiments are carried out by the leave-one-out cross-validation strategy, that is, the complete classification system is trained by taking into account all the samples but one, which is used as test sample. This procedure is repeated as many times as samples in the database, leaving each sample out in each iteration. Finally, an average accuracy rate is computed. Leave-one-out has been used to assess the discriminative accuracy of different multivariate analysis methods applied to the discrimination of frontotemporal dementia from AD [25] and in classifying atrophy patterns based on magnetic resonance imaging (MRI) images [26].

Besides the accuracy rate of the classification system, sensitivity and specificity are other statistical measures of the performance of a binary classification test, and moreover, they are the most widely statistics used to describe a diagnostic test. Sensitivity measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are identified as having the condition). Specificity measures the proportion of negatives which are correctly identified (e.g. the percentage of healthy people who are identified as not having the condition), and their expressions are

$$Sensitivity = \frac{TP}{TP+FN}, \quad Specificity = \frac{TN}{TN+FP} \tag{15}$$

where TP is the number of true positives: number of AD patient volumes correctly classified; TN is the number of true negatives: number of control volumes correctly classified; FP is the number of false positives: number of control volumes classified as AD patient; FN is the number of false negatives: number of AD patient volumes classified as control. These probabilities reveal the ability to detect NOR/AD patterns thus, the best CAD system is the one that achieves the best trade-off between specificity and sensitivity. Positive (PL) and negative likelihoods (NL) are displayed as a measure of the positive and predictive value of the method, given its prevalence independence:

$$PL = Sensitivity/(1-Specificity) = \frac{TP}{FP}$$

$$NL = (1-Sensitivity)/Specificity = \frac{FN}{TN}$$

A graphical plot of sensitivity versus $(1-\text{specificity})$ is usually represented, which is called receiver operating curve (ROC). The larger the area under the ROC is, the better the trade-off between sensitivity and specificity is reached.

8.1. Results on SPECT database

Classical PCA and LDA techniques have been evaluated on the SPECT database (see Table 1 for a detailed description) as feature extraction techniques in combination with the proposed supervised classifiers SVM and NN classifiers. Results of evaluating the complete CAD system are shown in Table 2. The brain volumes are first subsampled by a factor of $2\times2\times2$ and subsequently masked as explained in Section 4, and then, the remaining voxels undergo the linear PCA and LDA transformations. The number of PCA coefficients used in the classification step has been varied from $m=1$ to $N-1$, being $N=91$, the number of samples in the database. However, as expected, the best results are achieved when only a few of them are used. Specifically, the number $m=3$ has shown to be the optimal number of PCA coefficients to be used in this classification problem for both SVM and NN classifiers. The application of LDA to these coefficients improves the accuracy rates in all cases in spite of the reduction of the feature space from $m=3$ to $l=1$, but this final feature has shown to be more useful for separating NORMAL and AD classes. The rearrangement of the PCs by the FDR criterion improves the accuracy results with respect to PCA/LDA only when NN with $k=3$ and 5 neurons in the HL are used, yielding the highest peak accuracy of 91.21%. In any case, all the experiments increased the sensitivity – and therefore, the final accuracy rate – obtained by the baseline VAF approach.

The search of SOIs in order to perform PCA only on the most interesting brain regions were also carried out on the SPECT database. In this case, we no longer apply the mask since the feature space is already reduced significantly by choosing separated slices instead of the whole volume. When applying PCA slice

Table 2
Results obtained from the evaluation of SVM and feed-forward neural networks classifiers using PCA coefficients as features.

	SPECT			
	VAF	PCA ($m=3$)	PCA/LDA	FDR-PCA/LDA
SVM				
Linear (%)	85.71	89.01	90.11	90.11
	(83.67/87.80)	(92/85.36)	(92/87.80)	(92/87.80)
Quadratic (%)	–	87.91	89.01	89.01
	–	(88/87.80)	(88/90.24)	(88/90.24)
Polynomial (%)	–	85.71	90.11	87.91
	–	(86/85.36)	(92/87.80)	(90/85.36)
RBF (%)	–	89.01	90.11	90.11
	–	(92/85.36)	(92/87.80)	(92/87.80)
Feed-forward NN				
1 Neur. in HL (%)	–	82.42	90.11	85.71
		(84/80.48)	(88/92.68)	(88/82.92)
3 Neur. in HL (%)	–	87.91	90.11	91.21
		(90/85.36)	(88/92.68)	(94/87.80)
Neur. in HL (%)	–	90.11	87.91	91.21
		(94/85.36)	(84/92.68)	(94/87.80)
7 Neur. in HL (%)	–	82.42	89.01	87.91
		(82/82.93)	(88/90.24)	(90/85.36)

Comparison to the VAF baseline.

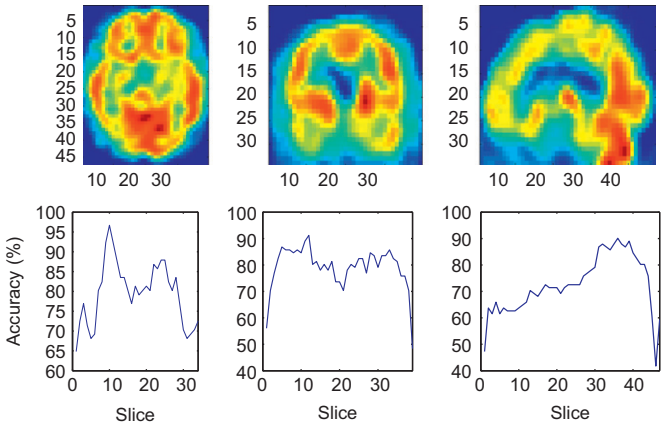


Fig. 5. Slices of interest (SOIs) found when exploring the brain volume along the three directions and accuracy rates reached by each one by means of PCA transformation and SVM classifiers.

by slice, we find that the classification results improve slightly with respect to using the whole volume. Fig. 5 shows the accuracy rates obtained by using each slice PCA coefficients as features in combination with SVM classifiers. As expected, the SOIs correspond to those regions of the brain where AD affects, i.e., the posterior cingulate gyri and precunei, as well as the temporo-parietal region. A complete set of results is shown in Table 3, where the number of PCs m used for the data projection and the SOI are specified. Again we find that no more than $m=3$ or 4 PCA coefficients are needed to classify satisfactorily when the SOI is found and selected for classification. SVM with quadratic kernel reaches 96.7% accuracy when PCA is applied on the 10th slice along the coronal axis, which was the most discriminant one. The combination of several PCA coefficients corresponding to SOIs along the three axes has been evaluated as well. However, it did not provide significant improvements.

8.2. Results on ADNI database

The variety of samples coming from different origins makes the ADNI database appealing for testing algorithms aiming at detecting

Table 3
Results obtained from the evaluation of SVM and slice-by-slice PCA application.

SPECT					
Axis direction	m/SOI	SVM Linear (%)	SVM Quadratic (%)	SVM Polynomial (%)	SVM RBF (%)
Coronal	3/10	87.91	96.7	90.11	93.41
	3/11	87.91	92.31	90.11	93.41
	4/10	86.81	90.11	85.71	92.31
	4/11	87.91	90.11	89.01	89.01
Sagittal	1/35	89.01	86.81	86.81	87.91
	1/36	87.91	89.01	90.11	90.11
Transaxial	3/11	89.01	91.21	84.62	86.81
	3/12	91.21	89.01	85.71	84.62
	4/11	89.01	87.91	81.32	85.71
	4/12	91.21	89.01	78.08	85.71

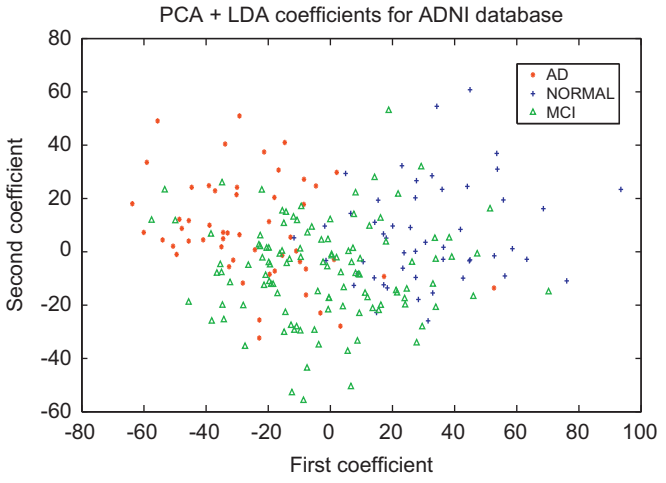


Fig. 6. Description of ADNI database by means of PCA+LDA coefficients. The original images are projected onto $m = 30$ PCs and subsequently onto $l=2$ LDA axis. MCI patterns are close to both NORMAL and AD patients, making the class separation task more difficult.

AD. A large number of MCI patients are included in the database which entails having a wide variety of perfusion patterns that range from NORMAL to AD patients. Obviously, the classification task becomes more difficult since the transition between NORMAL and AD patients is less abrupt. The 219 selected patients are represented in terms of PCA plus LDA coefficients in Fig. 6. The best description of this database in terms of PCA coefficients is found when around $m=30$ PCs are used to project the images, especially when MCI subjects are included. When only AD and NORMAL patients are considered, $m=8$ coefficients is enough to describe the existing variabilities between these two classes. Results obtained from this database are presented in the next sections depending on which classes are desired to be separated.

8.3. Group 1

In this section we present the results obtained in the classification task consisting of distinguishing between NORMAL and (AD, MCI) subjects. As shown in Fig. 6, MCI pattern of brain atrophy is complex and highly variable, and it evolves in time as the disease progresses [27,28]. Both SVM and NN classifiers reach accuracy rate peaks higher than 80%. Specifically, SVM with linear kernel and NN with $k=1$ neuron in the HL reached the highest accuracy rates of

82.19% and 81.74%, respectively. However, specificity values highlight that the classification performance carried out by the classifiers is quite poor, since it does not exceed 50% for any experiment. To obtain another sample prevalence independent measures, PL and NL are also presented. An important issue for an optimal classifier growth is the equitable proportion of positive and negative training samples, and that is not satisfied in group 1. This fact yields to deficiencies in the classifier construction, and therefore lower specificity values. All the results are shown in Table 4. In these experiments, the mask was applied and no FDR feature selection was used.

8.4. Group 2

Best results when classifying NORMAL controls versus AD patients are found when no voxel selection is performed, that is, the whole volumes are used to extract the PCA coefficients and the subsequent LDA features. An accuracy peak value is reached for both SVM and NN classifiers when $m=8$ PCA coefficients are used to project the images before applying LDA. The application of the FDR criterion on these coefficients improved the results, as shown in Fig. 7 for SVM. For this value of m, NN with $k=7$ neurons in the hidden layer reaches 88.75% accuracy rate with comparable sensitivity and specificity values (88.68/88.46%, respectively), and SVM outperforms NN where all the kernels reached the same accuracy peak of 89.52%. Best values of sensitivity and specificity obtained by SVM are represented as ROCs in Fig. 8. This figure gives us more information about the SVM classifier performances: the RBF kernel reached the best trade-off between sensitivity and specificity being the area under this curve 0.8668, followed by the quadratic kernel that covers an area of 0.8572.

8.5. Group 3

The most difficult classification task concerning ADNI database is to distinguish between NORMAL and MCI patients, due to the wide range spanned by the features extracted from MCI patients (see Fig. 6). When a conventional binary classification process is performed on these two data sets, accuracy rates do not exceed 74.1% by using PCA+LDA features (i.e., one final feature) combined with SVM with linear or quadratic kernel. Since MCI can be considered a previous stage of AD [27,29], we make profit of counting with an AD set of images to improve the classification results. Recall that LDA finds $l=c-1$ axis projection, where c is the

Table 4
Accuracy, (sensitivity/specificity) and PL/NL values obtained by evaluating SVM and NN in the classification task concerning group 1.

	Group 1				
	m	Linear	Quadratic	Polynomial	RBF
SVM	28	81.28% (91.61/48.08)% 1.764/0.174	81.28% (91.61/48.08)% 1.764/0.174	80.37% (90.42/48.08)% 1.7415/0.199	79% (89.22/46.15)% 1.657/0.234
	30	82.19% (92.21/50)% 1.844/0.156	81.74% (91.61/50)% 1.844/0.156	79.91% (89.22/50)% 1.7844/0.216	78.54% (88.02/48.08)% 1.695/0.249
		$k=1$	$k=3$	$k=5$	$k=7$
NN	30	81.28% (92.81/44.23)% 1.664/0.163	79.45% (89.22/48.08)% 1.718/0.224	79.91% (89.82/48.08)% 1.730/0.212	79.91% (89.22/50)% 1.784/0.216
	33	81.74% (92.81/46.15)% 1.723/0.156	80.82% (91.01/48.08)% 1.753/0.187	80.82% (91.02/48.08)% 1.753/0.187	80.37% (91.02/46.15)% 1.690/0.195

Data images are projected onto m PCs and the resulting coefficients onto $l=1$.

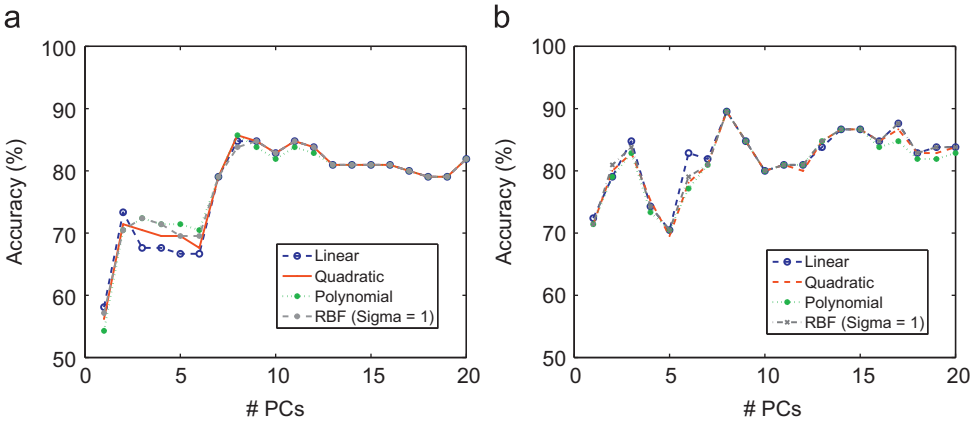


Fig. 7. Accuracy results obtained when group 2 is considered for classification using SVM with different kernels. In (a) the FDR criterion was not applied, whereas in (b) FDR was used.

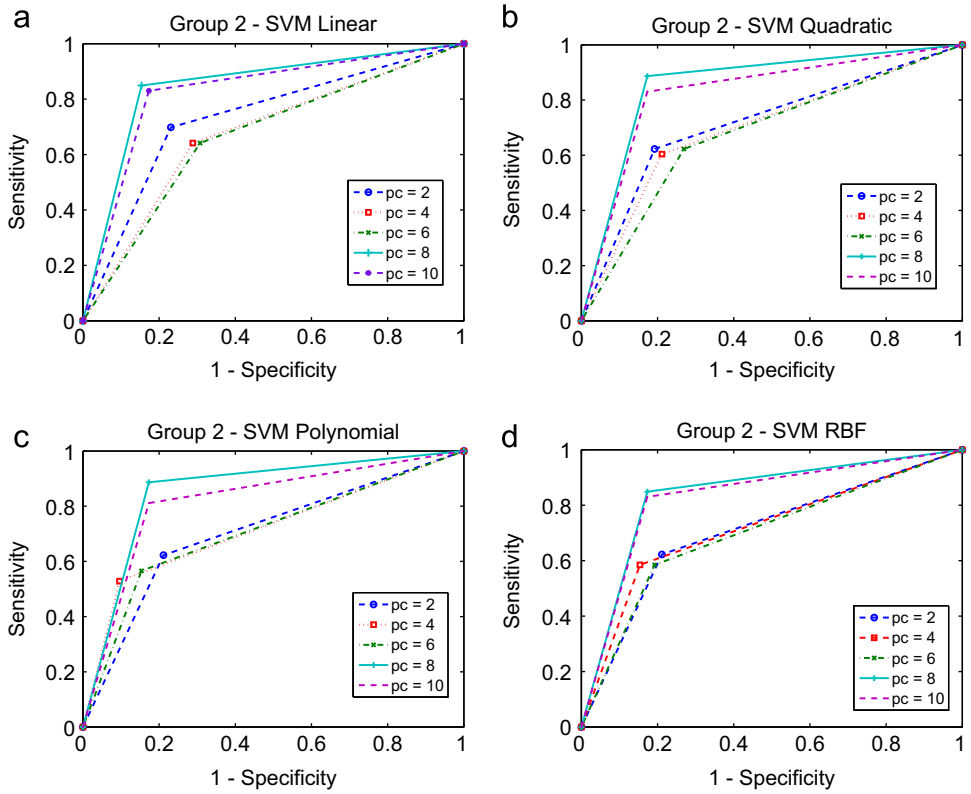


Fig. 8. ROCs for SVM and different number of PCs included in the training step when group 2 is considered for classification. For all the experiments, $m=8$ yielded the same accuracy rate, but RBF kernel reached the best trade-off between sensitivity and specificity.

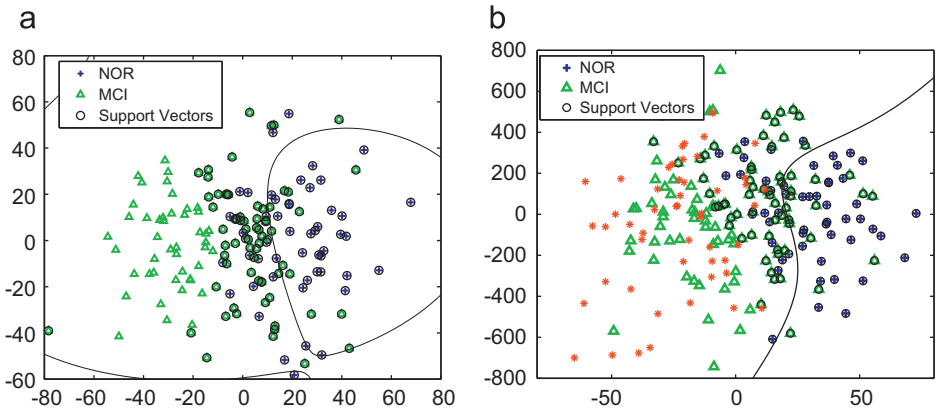


Fig. 9. NORMAL versus MCI classification task: line decision designed by an SVM classifier with polynomial kernel when (a) only MCI and NORMAL subjects are used in the training step and (b) when AD patients (depicted as red crosses) are included to design the classification rule. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

number of classes, so if we consider AD patients as a different class, we obtain $l=2$ features instead of one, as if we were dealing with a multiclass problem. Once the $l=2$ features are extracted, MCI and AD patients are labeled together as a unique class in the training step. Thus, the classifier will design a classification rule that gives strength to MCI patients, bringing them towards the AD vicinity. Once the classification rule has been established, only NORMAL and MCI samples are used to test the CAD system. Fig. 9 shows the differences on the surface decision designed by a polynomial kernel SVM classifier when solving the conventional binary classification problem and when the multiclass trick is used. This method improves classification results by increasing sensitivity values. Specificity values prove a poor performance of classifiers in recognizing NORMAL subjects, but recall group 2 is and imbalanced data set since the number of MCI samples doubles the number of NORMAL subjects. To avoid a misinterpretation of the results, PL and NL are computed as well.

Table 5 shows the results obtained by using both SVM and NN. In this classification task, best results were found when the mask was applied. In this case, the FDR criterion to rearrange the PCA coefficients did not provide better results. It can be seen that the use of the multiclass trick ($l=2$) improves the accuracy results significantly for all the experiments.

9. Conclusions

In this work, a complete CAD system for the diagnosis of the early AD has been presented. Several schemes combining different feature extraction, feature selection and classification techniques

Table 5
Accuracy, (sensitivity/specificity) and PL/NL values obtained by evaluating SVM and NN in the classification task concerning group 3.

	Group 3				
	# PCs	Linear	Quadratic	Polynomial	RBF
$l=1$	18	74.1% (86.84/46.15)% 1.613/0.285	74.1% (86.84/46.15)% 1.613/0.285	73.49% (85.96/46.15)% 1.596/0.304	73.49% (85.96/46.15)% 1.596/0.304
$l=2$	30	81.33% (97.37/46.15)% 1.808/0.057	77.71% (91.23/46.15)% 1.694/0.190	77.71% (91.23/46.15)% 1.694/0.190	77.71% (91.22/48.08)% 1.757/0.183
		$k=1$	$k=3$	$k=5$	$k=7$
$l=1$	33	71.08% (82.46/46.15)% 1.531/0.380	72.29% (83.33/48.08)% 1.605/0.347	71.08% (82.46/46.15)% 1.531/0.380	70.48% (81.58/46.15)% 1.515/0.399
$l=2$	40	79.52% (94.74/46.15)% 1.759/0.114	78.31% (93.86/44.23)% 1.683/0.139	79.52% (93.86/48.08)% 1.801/0.128	79.52% (92.10/46.15)% 1.710/0.171

have been deeply studied and tested on real SPECT and PET databases. All the proposals aim at reducing the dimension of the original images, which after the reconstruction count with more than 500 000 voxels, to a number of features comparable to the number of samples, solving that way the so-called *small sample size*. Different voxels selection methods are proposed to initially discard voxels that are not useful for the AD detection. After that, the input space is transformed by means of PCA and LDA transformations, that compress the useful information in only one or two features, holding the statistical differences between classes. By means of simple rearrangements, these features have shown to work satisfactorily for classification purposes.

The best combination of techniques that compose the complete CAD systems is not fixed, but depends on the specific database and the classification task we are dealing with. Only in the classification stage we can determine that SVM works in general better than NN, yielding in most cases higher accuracy rates when the same features are used.

PCA is a compression technique that does not need to rule out useless voxels for finding the axes that best describe the data set. However, if the data to be described by PCs concentrate more differences, the main PCA coefficients will be more useful for classification. That is the idea behind the slice-by-slice PCA application, which searches for the best slices for the classification task that concerns us. This approach reached the higher accuracy rate for SPECT image classification, yielding 96.7%.

When classes are best classified by linear surfaces or decision lines, the rearrangement of the PCA coefficients by the FDR criterion usually yields to higher accuracy rates. For PET images, FDR was useful when dealing with group 2, which was the best described group in terms of class separability by means of PCA+LDA features. However, neither for group 1 nor 3 it outperformed PCA+LDA and not more.

LDA transformation projects the PCA features onto other space more reliable for classification, but in contrast it reduces the features to only one in binary classification problems, which sometimes might be insufficient, especially with PET images when dealing with group 3. MCI class is made up of NORMAL patients that gradually move to AD class, being possible a reverse conversion to NORMAL state as well as an evolution to other different type dementia. This makes especially difficult the automatic definition of a separation rule for these two classes. To make sense to this group experiments, MCI subjects are understood as a first stage of AD. Thus, we make use of AD samples to extract not one, but two features. In the test stage though, the CAD system is evaluated only on NORMAL and MCI samples. The use of this method increases the sensitivity values noticeably.

The use of imbalanced training data sets 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 an imbalanced data set may not reflect a high false negative rate. This problem is solved by using other sample prevalence independent parameters, as PL or NL and/or by selecting balanced data sets such as group 2 (in our case this reduces the number of scans and affects the learning process, however, it is useful for comparing feature extraction methods under the same conditions). Secondly, classifiers are limited in their performance when dealing with imbalanced data sets [30,31]. However, 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.

Acknowledgements

This work was partly supported by the MICINN under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030- C02-01) and HD2008-0029 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Projects (TIC-02566 and TIC-4530).

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

We are also grateful to C. Carnero and coworkers from the "Virgen de las Nieves" hospital in Granada (Spain) for providing and classifying the SPECT images used in this work.

References

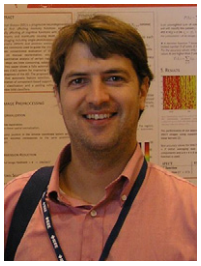
[1] J. Stoeckel, G. Malandain, O. Migneco, P.M. Koulibaly, P. Robert, N. Ayache, J. Darcourt, Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients, Medical Image Computing and Computer-Assisted Intervention—MICCAI, Lecture Notes in Computer Science, vol. 2208, Springer2001, pp. 666–674 . doi:10.1007/3-540-45468-3-80.
[2] 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: IEEE Nuclear Science Symposium Conference Record, Medical Imaging Conference, 2008, pp. 4392–4395, dresden (Germany).
[3] J. Ramírez, J. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, M. Gómez-Río, Computer-aided diagnosis of alzheimer's type dementia combining support vector machines and discriminant set of features, Information Sciences, accepted for publication, doi:10.1016/j.ins.2009.05.012.
[4] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, M. López, I.A. Illán, C.G. Puntonet, M. Gómez-Río, Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines, Neuroscience Letters 461 (1) (2009) 60–64.
[5] J. Ramírez, J.G.R. Chaves, M. López, D. Salas-Gonzalez, I. Álvarez, F. Segovia, SPECT image classification using random forests, Electronics Letters (2009) 604–605.
[6] K.J. Friston, J. Ashburner, S.J. Kiebel, T.E. Nichols, W.D. Penny, Statistical Parametric Mapping: The Analysis of Functional Brain Images, Academic Press, 2007.
[7] E. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E. Stadlan, 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 (1984) 939–944.
[8] 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.
[9] Available at: 〈http://www.fil.ion.ucl.ac.uk/spm/software/spm5〉.
[10] J. Ashburner, K.J. Friston, Nonlinear spatial normalization using basis functions, Human Brain Mapping 7 (4) (1999) 254–266.
[11] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, A. Lassl, C.G. Puntonet, Improved Gauss–Newton optimization methods in affine registration of SPECT brain images, IET Electronics Letters 44 (22) (2008) 1291–1292. doi:10.1049/el:20081838.
[12] R. Chaves, J. Ramírez, J.M. Górriz, 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 (3) (2009) 293–297.
[13] I. Jolliffe, Principal Component Analysis, Springer Verlag, New York, 1986.
[14] P.G. Spetsieris, Y. Ma, V. Dhawan, D. Eidelberg, Differential diagnosis of parkinsonian syndromes using functional PCA-based imaging features, NeuroImage 45 (4) (2009) 1241–1252.
[15] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[16] M. López, J. Ramírez, J.M. Górriz, D. Salas-Gonzalez, I. Álvarez, 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.
[17] 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 (4) (2009) 233–238.
[18] I. Álvarez, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C.G. Puntonet, F. Segovia, Alzheimer's diagnosis using eigenbrains and support vector machines, IET Electronics Letters 45 (7) (2009) 342–343.
[19] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (1936) 179–188.
[20] R. Duda, P. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[21] P.N. Belhumeour, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[22] 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 (2009) 472–485.
[23] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998) 121–167.
[24] W.S. McCulloch, W. Pitts, A logical calculus of ideas immanent in nervous activity, Bulletin of Mathematical Biophysics 5 (1943) 115–133.
[25] 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. doi:10.1002/sim.1719.
[26] Y. Fan, N. Batmanghelich, C. Clark, C. Davatzikos, Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline, NeuroImage 39 (2008) 1731–1743.
[27] S. Minoshima, B. Goirdani, S. Berent, K. Frey, N. Foster, D. Khul, Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease, Annals of Neurology 42 (1) (1997) 85–94. doi:10.1002/ana.410420114.
[28] 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.
[29] 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. doi:10.1001/jama.286.17.2120 〈http://jama.ama-assn.org/cgi/content/full/286/17/2120〉.
[30] N. Japkowicz, The class imbalance problem: significance and strategies, in: Proceedings of the 2000 international conference on artificial intelligence, ICAI 2000, vol. 1, 2000, pp. 111–117.
[31] F. Vilario, P.S.J. Vitriá, P. Radeva, Experiments with SVM and stratified sampling with an imbalanced problem: detection of intestinal contractions, in: Lecture Notes in Computer Science, vol. 1, 2005, pp. 783–791.

M. López received the B.Sc. degree in Telecommunications Engineering from the University of Seville (Spain) in 2007 and the Master degree in Multimedia Technologies from the University of Granada (Spain) in 2008. She is currently a Ph.D. Student in the Department of Signal Theory, Networking and Communications at the University of Granada. Her research interests lie in the field of signal processing, in particular image processing and classification for biomedical applications.

J. Ramírez received the M.A.Sc. degree in Electronic Engineering in 1998, and the Ph.D degree in Electronic Engineering in 2001, all from the University of Granada. Since 2001, he is an Associate professor at the Department of Signal Theory Networking and Communications of the University of Granada (Spain). His research interest includes signal processing and biomedical applications including brain image processing, robust speech recognition, speech enhancement, voice activity detection, seismic signal processing and implementation of high performance digital signal processing systems. He has coauthored more than 150 technical journal and conference papers in these areas. He has served as reviewer for several international journals and conferences.

J.M. Górriz received the B.Sc. degree in physics and electronic engineering from the University of Granada, Granada, Spain, and the Ph.D. degree from the Universities of Cádiz and Granada, Spain, in 2000, 2001, 2003, and 2006, respectively. He is currently an Associate Professor with the Department of Signal Theory, Networking, and Communications at the University of Granada. He has coauthored more than 170 technical journals and conference papers in these areas and has served as Editor in Chief for the Open Acoustics Journal, Bentham, since 2007. His present interests lie in the field of statistical signal processing and its application to speech and image processing.

Ignacio Alvarez received the B.Sc. Degree in Physics from the Complutense of Madrid University, Spain and the Ph.D from the University of Granada, Spain in 2004 and 2009. Actually, he is working under a local fellowship in the Department of Signal Theory, Networking and Communications at the University of Granada. His present research interests include supervised learning, signal processing, independent component analysis and biomedical applications.



D. Salas-Gonzalez was born in Málaga, Spain, in 1980. He obtained his M.Sc. and B.Phil. degrees in physics from the University of Granada in 2003 and 2005, respectively. He was a granted national researcher from the Ministry of Education and Science of Spain from 2004 to 2008. His research interests are in statistical signal processing, Bayesian inference and applications in bioinformatics and image processing.

F. Segovia received the B.Sc. Degree in Computer Science in 2006, and the Master degree in Computer Engineering and Networks in 2009, both from the University of Granada, Spain. He is currently a Ph.D. student of the Department of Signal Theory, Networking and Communications at the University of Granada. His main research interests include Image Processing and Biomedical Applications.

R. Chaves received the B.Sc. degree in Telecommunications Engineering and the Master degree in Multimedia Technologies from the University of Granada, Spain in 2008 and 2009 respectively. She is a Ph.D. student in Classification Techniques of neurologic alterations at the University of Granada (SIPBA research group, Department Signal Theory, Networking and Communications). Her research interests at present lie in the field of Signal Processing and Biomedical Applications.

P. Padilla received the Telecommunication Engineer degree from Technical University of Madrid (UPM), Spain, in 2005 and Ph.D. in 2009. From 2005 to 2009, he was with the Radiation Group of the Signal, Systems and Radiocommunications Department of UPM. In 2007, he was with the Laboratory of Electromagnetics and Acoustics at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, as invited Ph.D. Student and in 2009, he carried out a 4 month postdoc at Helsinki University of Technology (TKK). Since September 2009, he is Assistant Professor at the University of Granada. His research interests include signal processing and its applications to image processing.

M. Gómez-Río received the B.Sc. and Ph.D. degrees in Medicine (1980–1986) from the University of Granada, Spain. He is currently a Specialist in Nuclear Medicine (1991) with the Nuclear Medicine Department of the "Virgen de las Nieves" Hospital, Spain.