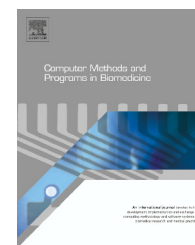




ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Functional activity maps based on significance measures and Independent Component Analysis

F.J. Martínez-Murcia^a, J.M. Górriz^{a,*}, J. Ramírez^a, C.G. Puntonet^b,
I.A. Illán^a, for the Alzheimer's Disease Neuroimaging Initiative¹

^a Department of Signal Theory, Networking and Communications, 18071 University of Granada, Spain

^b Department of Computer's Architecture and Technology, 18071 University of Granada, Spain

ARTICLE INFO

Article history:

Received 12 March 2012

Received in revised form

18 October 2012

Accepted 22 March 2013

Keywords:

Alzheimer's Disease (AD)

Computer Aided Diagnosis (CAD)

Relative Entropy

Independent Component Analysis (ICA)

Naive Bayes Classifier

Support Vector Machines (SVM)

PET and SPECT

ABSTRACT

The use of functional imaging has been proven very helpful for the process of diagnosis of neurodegenerative diseases, such as Alzheimer's Disease (AD). In many cases, the analysis of these images is performed by manual reorientation and visual interpretation. Therefore, new statistical techniques to perform a more quantitative analysis are needed. In this work, a new statistical approximation to the analysis of functional images, based on significance measures and Independent Component Analysis (ICA) is presented. After the images preprocessing, voxels that allow better separation of the two classes are extracted, using significance measures such as the Mann–Whitney–Wilcoxon U-Test (MWW) and Relative Entropy (RE). After this feature selection step, the voxels vector is modelled by means of ICA, extracting a few independent components which will be used as an input to the classifier. Naive Bayes and Support Vector Machine (SVM) classifiers are used in this work. The proposed system has been applied to two different databases. A 96-subjects Single Photon Emission Computed Tomography (SPECT) database from the “Virgen de las Nieves” Hospital in Granada, Spain, and a 196-subjects Positron Emission Tomography (PET) database from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Values of accuracy up to 96.9% and 91.3% for SPECT and PET databases are achieved by the proposed system, which has yielded many benefits over methods proposed on recent works.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Functional images, such as Positron Emitted Tomography (PET) or Single Photon Emitted Computed Tomography (SPECT), are very important in the process of diagnosis of neurodegenerative diseases. They allow clinicians to correctly interpret the presence of different functional patterns – like blood flow or glucose concentration – in several key regions of

the brain, caused by neurological diseases such as Alzheimer's Disease (AD), where early diagnosis is of particular importance in the development of new treatments. However, the evaluation of these scans often relies on manual reorientation, visual reading and semiquantitative analysis. Therefore, Computer Aided Diagnosis (CAD) tools based on medical imaging are a very valuable help for physicians in the AD detection, providing an objective, operator-independent, consistent analysis of the images.

* Corresponding author. Tel.: +34 958243271; fax: +34 958243271.

E-mail address: gorriz@ugr.es (J.M. Górriz).

¹ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.ucla.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: URL http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

0169-2607/\$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.cmpb.2013.03.015>

A CAD system is a computer method and program that assists physicians in the diagnosis task using biomedical signals. It encompasses a set of signal processing methods implemented in a high-level language, i.e. Matlab, resulting in complete systems, such as Statistical Parametric Mapping (SPM) [1]. This is the most extended tool in the neuroimaging community, and refers to the construction and assessment of spatially extended statistical processes used to test hypotheses about functional imaging data. The CAD paradigm has been successfully applied to medical imaging in many works [1–5] to perform an exhaustive quantitative and qualitative analysis of these medical images. Even in the simplest approach, which uses the raw voxels as an input to the classifier (Voxels as Features, VAF [5]), a new strategy to avoid the small sample size problem [6] is desirable. This phenomenon occurs when the number of input features to the classifier is higher than the number of samples used to train this classifier. Techniques like image segmentation and definition of Regions of Interest (ROIs) [3,7–9] allows a CAD system to consider only brain zones which are significant for AD detection. Other techniques involve statistical measures that compute significance values for each voxels. Then, voxels are ranked and the most significative ones are selected. Statistical measures can be computed by the widely used Student's t-Test [1,2,5], Mann–Whitney–Wilcoxon *U*-Test [10,11] or Relative Entropy [12,13]. These methods have been proven very useful in applications ranging from cancer diagnosis [13] to segmentation of microscope images [12] or autoradiographic images [11]. Later, feature extraction algorithms like Independent Component Analysis (ICA) [14,15], Fisher Linear Discriminant Ratio (FDR) [16], Principal Component Analysis (PCA) [17–19] or Factor Analysis [2,20] have been used for extracting common features. In [15], a method for AD early diagnosis based on ICA analysis is proposed. The method has tested in the feature extraction step with excellent results, but some improvements can be applied to it. One voxel selection step by means of statistical significance has been added in order to select the most significant voxels in the brain and use them to extract the independent components. The second improvement relies in the estimation of the mixing matrix in the ICA step. While the method in [15] relies on creating an average image of the individual samples of each class (AD and CTRL) and then compute the mixing matrix, our method uses all the sample images and their class information to compute this matrix. This is possible thanks to the feature selection step, which reduces the number of voxels from over 500,000 to less than 20,000, enabling a faster computation, and hence, allowing us to include all the samples in the calculation of the mixing matrix. This leads to a more consistent estimation of the Independent Components.

In this work, a new combination of algorithms for voxel selection and feature extraction is proposed. Firstly, the most significative voxels in the task of separating classes are selected by using the significance values, computed using Student's t-Test, Mann–Whitney–Wilcoxon *U* test and Relative Entropy. To reduce the number of features, we extract independent components from the selected voxels using ICA. Then, these features are classified using a Bayesian or Support Vector Machine (SVM) classifier.

Processing of perfusion images is performed in three stages:

- 1 **Voxel selection.** Ranking of voxels by measures of difference between two classes and posterior selection of the first N , using Student's t-Test, Mann–Whitney–Wilcoxon (MWW) and Relative Entropy (RE) techniques.
- 2 **Feature extraction.** Extraction of K common features from the voxels selected in the previous step, using an ICA technique. These K features will be used as an input vector to the classifier.
- 3 **Classification** of the features vector in two different classes: Normal Controls (CTRL) and AD. We make use of two different classifiers: Naive Bayes and SVM.

A baseline method is also defined, and will be used along with the VAF approximation for comparing purposes. This method uses the selected voxels (obtained by the algorithms used in the voxel selection stage) as input features to the classifier, without using any feature extraction technique, and so, is called Selected Voxels As Features (SVAF).

2. Materials and methods

2.1. Datasets

To evaluate the performance of the proposed method, we use two different dataset of perfusion images: a database composed by 96 SPECT brain images from “Virgen de las Nieves” Hospital in Granada (Spain) and a set of 403 PET brain images from the ADNI initiative.

2.1.1. SPECT database

The database is built up of imaging studies of subjects following the protocol of an hospital-based service. First, the neurologist evaluated the cognitive function, and those patients with findings of memory loss or dementia were referred to the nuclear medicine department in the “Virgen de las Nieves” hospital (Granada, Spain), in order to acquire complementary screening information for diagnosis.² Experienced physicians evaluated the images visually. The images were assessed using 4 different labels: Control (CTRL) for subjects without scintigraphic abnormalities and mild perfusion deficit (AD1), moderate deficit (AD2) and severe deficit (AD3), to distinguish between different levels of presence of hypo-perfusion patterns compatible with AD. In total, the database consists of $n=97$ subjects: 41 CTRL, 30 AD1, 22 AD2 and 4 AD3 (see Table 1 for demographic details). Since the patients are not pathologically confirmed, the subject's labels possesses some degree of uncertainty, as the pattern of hypo-perfusion may not reflect the underlying pathology of AD, nor the different classification of scans necessarily reflect the severity of the patients symptoms. However, when pathological information is available, visual assessments by experts have been shown to be very sensitive and specific labelling methods, in contrast to

² Clinical information is unfortunately not available for privacy reasons, but only demographic information.

Table 1 – Demographic details of the dataset.

	#Samples	Sex (M/F) (%)	μ [range/ σ]
CTRL	41	32.95/12.19	71.51 [46–85/7.99]
AD1	29	10.97/18.29	65.29 [23–81/13.36]
AD2	22	13.41/9.76	65.73 [46–86/8.25]
AD3	4	0/2.43	76 [69–83/9.90]

CTRL = Normal Controls, AD 1 = possible AD, AD 2 = probable AD, AD 3 = certain AD. μ and σ stands for population mean and standard deviation, respectively.

neuropsychological tests [21,22]. Given that this is an inherent limitation of ‘in vivo’ studies, our working-assumption is that the labels are true, considering the subject label positive when belonging to any of the AD classes, and negative otherwise.

2.1.2. ADNI PET database

Data contained in the second database, and used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (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 Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many 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. 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 up-to-date information, see www.adni-info.org.

FDG PET scans were acquired according to a standardized protocol. A 30-min dynamic emission scan, consisting of 6 5-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 (www.loni.ucla.edu/ADNI/Data/ADNIData.shtml). Following the scan, each image was reviewed for possible artefacts

at the University of Michigan and all raw and processed study data was archived.

Enrolled subjects were between 55 and 90 (inclusive) years of age. 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 and 30 (inclusive), a CDR of 0, non-depressed, nonMCI, 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 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, 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 criteria for probable AD.

For posterior analysis, the data was arranged into only one group, considering the case AD vs CTRL. This group has $n = 196$ images in total: 101 CTRL and 95 AD. We will consider that each spatially and intensity normalized image from the database corresponds to a different observation of the brain.

2.2. Preprocessing

The complexity of brain structures and the differences between brains of different subjects makes difficult to use the registered images for all type of calculations that directly compares them with each other. However, the proposed system will assume the following two conditions to be met:

- 1 the same position in the image coordinate system within different images corresponds to the same anatomical position.
- 2 the same intensity value, wherever it might be found (within an image or in other images) corresponds to the same physical value.

We met the first condition by performing a spacial normalization. This method assumes an affine generic model with 12 parameters [23] and a cost function which presents an extreme value when the image and a template that represents our ‘brain space’ correspond one with the other. The spacial normalization is achieved by optimizing the quadratic difference between the template and each image.

After this spacial normalization, the resulting images (of size $95 \times 79 \times 69$ voxels in both datasets) are filtered, yielding in a smoothing of the images in function of a parameter called “Full Width at Half Maximum” (FWHM). This filtering

increases the Signal to Noise Ratio (SNR) and guarantees that the changes present in the images will be given in an anatomically significant scale.

The second condition is met when we apply a intensity normalization on the images. This is done by normalizing the intensity value of the images to a maximum value, I_{max} , computed by averaging the 3% highest voxel intensities in the SPECT database. For the ADNI database, an average image is computed from the spatially normalized images, and then a specific mask is applied to each subject, so that the average of the voxels inside the mask is exactly one, as explained in [24].

2.3. Voxel selection

In this stage, we use the results of different statistical measures of significance, like the Student's t-Test [25], the Mann-Whitney-Wilcoxon U-Test (MWW) [10,11,25] and the Relative Entropy (RE) [12,26], to rank voxels from the most to the least significant ones. These measures of significance will be higher when the coordinate tested in both classes is more different. Then, we proceed with the selection of the first N voxels.

2.3.1. Student's t-Test

Student's t-Test (t-Test) is a widely used statistical test which quantifies the differences between two different classes. It uses a common estimation of variance for both classes. The value of statistical t can be computed [25] as:

$$t = \frac{\bar{\Omega}_1 - \bar{\Omega}_2}{\sigma_{\Omega_1\Omega_2}^2 \cdot \sqrt{(1/n_1) + (1/n_2)}} \quad (1)$$

where

$$\sigma_{\Omega_1\Omega_2}^2 = \sqrt{\frac{(n_1 - 1)\sigma_{\Omega_1}^2 + (n_2 - 1)\sigma_{\Omega_2}^2}{n_1 + n_2 - 2}} \quad (2)$$

$\sigma_{\Omega_1\Omega_2}^2$ is an estimator of common standard deviation of both samples, $\bar{\Omega}_1$ and $\bar{\Omega}_2$ are the mean of each class, n_1 is the number of samples in class ω_1 and n_2 is the number of samples in class ω_2 .

2.3.2. Mann-Whitney-Wilcoxon

Mann-Whitney-Wilcoxon U-Test (MWW) uses the absolute value of the statistical U to rank voxels. Calculation of U value is done by the following expression [25]:

$$U_i = R_i - \frac{n_i(n_i + 1)}{2} \quad (3)$$

where n_i is the sample size for sample i , and R_i is the sum of the ranks in sample i (where $i = 1, 2$). Smaller U_i value is taken as the final U value.

This statistical test measures the dissimilarity between two groups of values, and, although is similar to Student's t-Test, is less likely than it to spuriously indicate significance because of the presence of outliers. As we make use of real data, and we have no knowledge about the images' statistical distribution, MWW test could be a very good choice [25].

Relative Entropy

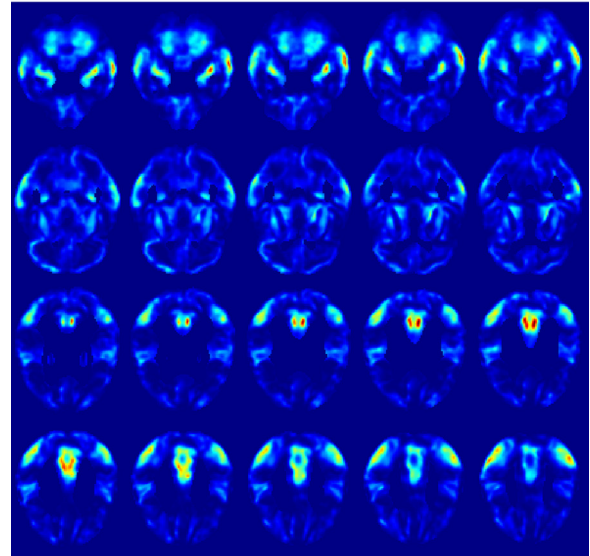


Fig. 1 – Several selected slices of the significance map obtained for each brain coordinate by applying the Relative Entropy criterium to the ADNI database.

2.3.3. Relative entropy

Relative Entropy (or Kullback-Leibler divergence) is a non-symmetric measure of the difference between two probability distributions Ω_1 and Ω_2 . Because of its non-symmetric property, we can make use of this to evaluate the difference between CTRL and AD images for each voxel. Relative Entropy can be calculated with Eq. (4) [26].

Let Ω_1 and Ω_2 be two discrete random variables. Relative Entropy is defined as:

$$KL_{\omega_1\omega_2} = \int_V \omega_1 \log \frac{\omega_1}{\omega_2} d\mu \quad (4)$$

where μ is any measure of V , the set of all voxels that are placed on a certain brain coordinate, in which $\omega_1 = \frac{d\Omega_1}{d\mu}$ and $\omega_2 = \frac{d\Omega_2}{d\mu}$ exist. Fig. 1 depicts values of Relative Entropy for each voxel.

2.4. Independent Component Analysis

Independent Component Analysis (ICA) [27], is a statistical technique that represents a multidimensional random vector as a lineal combination of non-Gaussian random variables (the so-called "independent components") to be as independent as possible, and has been used widely on segmentation and clustering of medical images [28-31]. It can be considered as a non-Gaussian version of Factor Analysis. Assume that we observe n lineal mixtures $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ of length N that can be modelled as an expression of K independent components (IC). These independent components are defined as $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K)$, where each \mathbf{s}_K vector has a length of N . So, each random vector \mathbf{x}_n can be described as a linear combination of K independent components:

$$\mathbf{x}_n = a_{1n}\mathbf{s}_1 + a_{2n}\mathbf{s}_2 + \dots + a_{Kn}\mathbf{s}_K \quad (5)$$

Without loss of generality we can assume that both the observed vectors and the independent components are zero mean. If the previous conditions are not met, the \mathbf{x} variables can be centred by subtracting the sample mean. To use a vector-matrix notation, more convenient in this case, we denote as matrix \mathbf{X} the random vector whose elements are $\mathbf{x}_1, \dots, \mathbf{x}_n$. We also denote as \mathbf{A} the matrix that contains all $a_{k,n}$ elements, the "mixing matrix" that projects each image into the space defined by the IC. Using this notation, the mixing model above remains as follows:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (6)$$

The starting point of ICA is the assumption that all components \mathbf{s}_k are statistically independent. To measure independence, we assume that all independent components have a non-Gaussian statistical distribution. It is assumed that a sum of independent signal trends to Gaussianity, so if non-Gaussianity is maximized with any independence criteria F , for instance, the kurtosis or negentropy, we obtain signals that are more independent than the previous ones [32,27]. After estimating the matrix \mathbf{A} , we can compute its inverse, \mathbf{W} and obtain the projection \mathbf{S} of the images in the dataset into the IC space with:

$$\mathbf{S} = \mathbf{W}\mathbf{X} \quad (7)$$

2.4.1. FastICA

Adaptive algorithms based on gradient descend can be problematic when they are used on an environment in which adaptation is not necessary, like this case. The convergence is often slow, and depends on the choice of convergence parameters. As a solution to this problem, block algorithms based on fixed-point iteration [33,34] can be used. In [33] a fixed-point algorithm based on kurtosis is introduced. In [34], this algorithm, known as FastICA, is generalized to general contrast functions. The single unit FastICA algorithm has the following form:

$$\mathbf{w}(k) = E\{\mathbf{x}g(\mathbf{w}(k-1)^T\mathbf{x})\} - E\{g'(\mathbf{w}(k-1)^T\mathbf{x})\}\mathbf{w}(k-1) \quad (8)$$

where the loadings vector \mathbf{w} is normalized to unit norm in each iteration, and the function $g(x)$ is a derivative of the contrast function G defined in [32]. The expected values are estimated in practice by using the mean of a significantly high number of samples of the input data. The speed of convergence of the fixed-point algorithms is clearly superior to more neural algorithms. Improvements between 10 and 100 times the speed are observed frequently [35].

2.4.2. Application to the databases

In this stage, we have a matrix \mathbf{X} that contains all \mathbf{x}_n mixture vectors, each one containing N values. Due to the preprocessing step (see Section 2.2) this vector represents the intensity values for each brain coordinate in each image. The application of the FastICA algorithm to this matrix first reduces the dimensionality by extracting the first K eigenvalues, when estimating the input matrices to the fixed-point algorithm: \mathbf{D} , the matrix of eigenvalues, of size $K \times K$ and the matrix of eigenvectors \mathbf{E} , of size $N \times K$.

Then we proceed to the whitening of the data, and the independent component is extracted using the algorithm described above. From this algorithm we obtain the mixing matrix \mathbf{A} and its inverse \mathbf{W} , of size $K \times n$, for the complete database \mathbf{X} . Later, the inverse of the mixing matrix is applied individually to each vector of selected voxels \mathbf{x}_n , to project each vector into the IC space

$$\mathbf{s}_n = \mathbf{W}\mathbf{x}'_n \quad (9)$$

The above expression shows the transformation applied to each vector of selected voxels, where \mathbf{s}_n is the projection of the data vector \mathbf{x}_n into the K -dimensional IC space, which corresponds to one image of the dataset. Finally we obtain the n representations of the images on K independent components. Each of these vectors will be used as input to the classifier. In [15], the mixing matrix \mathbf{A} is estimated from an average image of all images in each class. On the other hand, in the proposed system, \mathbf{A} is estimated from the complete database \mathbf{X} , what results in a different estimation.

Fig. 2 shows the first four independent components obtained using this system and the SPECT database. These components are extracted from the \mathbf{W} matrix, and has been spatially represented by assigning the brain coordinate of each voxel to each of the values in these IC. We can observe that each component highlights different zones that are usually related to Alzheimer's Disease, such as the posterior cingulate gyri and precunei, as well as the temporo-parietal region, both considered as typically affected by glucose hypometabolism or rCBF hypoperfusion in AD patients [36].

2.5. Classifiers

2.5.1. Naive Bayes Classifier

The Naive Bayes Classifier (from now on, Bayes classifier or simply Bayes) is based on the optimal decision rule that minimizes the error in a two-hypothesis test. Bayesian-based classifiers have been successfully used in patten recognition problems [37,17].

The Bayes classifier evaluates the *a posteriori* probability function [38]. Let $\{\omega_1, \omega_2, \dots, \omega_c\}$ denote the object classes and \mathbf{x} a feature vector obtained by some feature extraction technique. The *a posteriori* probability function of \mathbf{x} given ω_i is defined as

$$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})}, \quad i = 1, 2, \dots, c. \quad (10)$$

where $P(\omega_i)$ is a *priori* probability, $P(\omega_i|\mathbf{x})$ the conditional probability density function of \mathbf{x} given ω_i and $P(\mathbf{x})$ is the mixture density. The maximum a posteriori (MAP) decision rule for the Bayes classifier is defined as

$$P(\mathbf{x}|\omega_i)P(\omega_i) = \max_j \{P(\mathbf{x}|\omega_j)P(\omega_j)\}, \quad \mathbf{x} \in \omega_i \quad (11)$$

The test vector data \mathbf{x} is classified to ω_i of which the *a posteriori* probability given \mathbf{x} is the largest between the classes. This assumption leads to a quadratic discriminant function, which is referred to hereinafter as Quadratic Bayes Classifier (or Bayes Quadratic). If we assume that the covariances of the

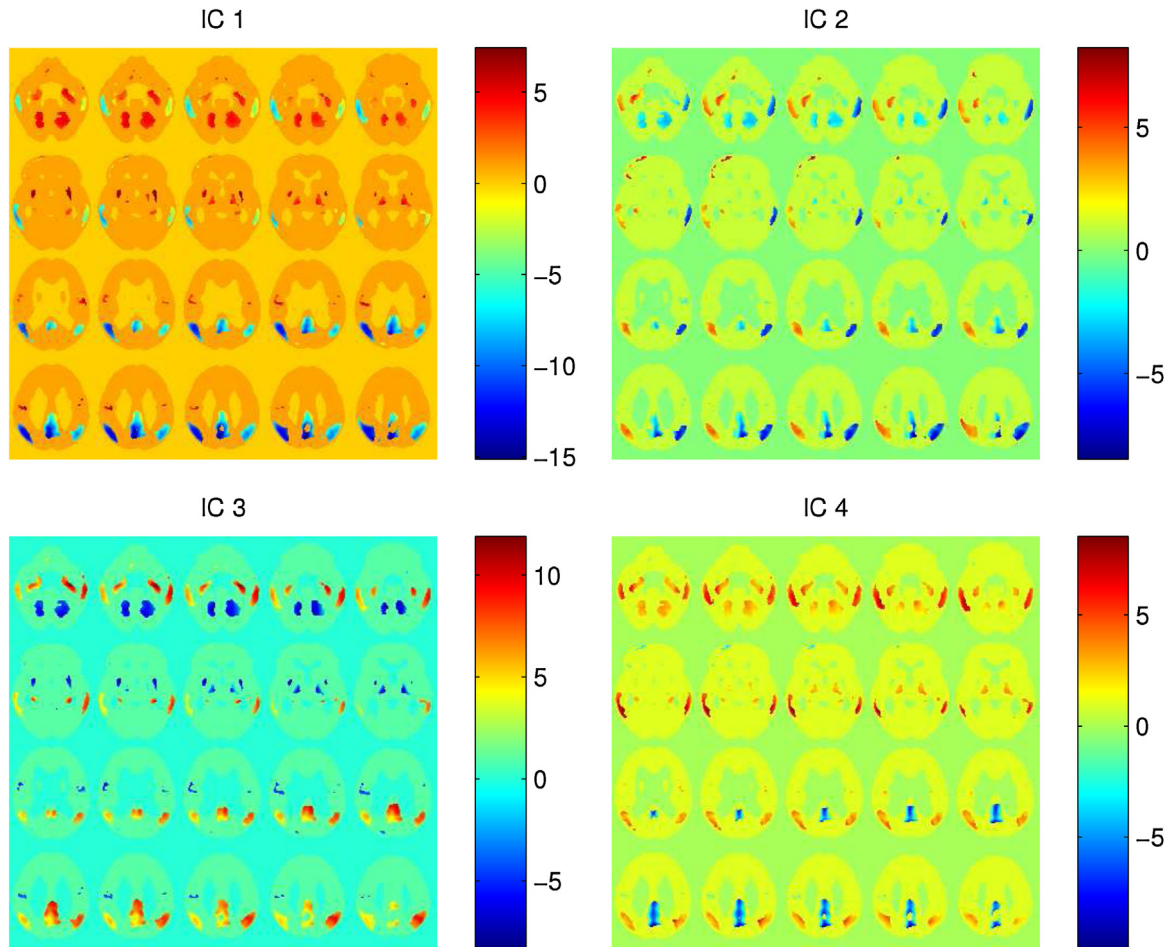


Fig. 2 – Some selected slices depicting a spatial representation of the first four components obtained by ICA in the ADNI database.

two classes ω_1 and ω_2 are identical, the quadratic terms in \mathbf{x} of the exponents of Gaussian densities can be cancelled, resulting in a linear discriminant in function of \mathbf{x} . This classifier is referred as Linear Bayes Classifier (or Bayes Linear).

For the classification problem that we are dealing with, the *a priori* probability $P(\omega_i)$ is initially set to 0.5, that is, AD and CTRL classes are equally probable. Regarding the conditional probability density function, usually there are not enough samples to estimate it for each class (within-class density). A compromise, therefore, is to make an assumption of a particular density form and convert the general density estimation problem into a parametric one. The within class densities are usually modelled as normal distributions.

2.5.2. Support Vector Machines

Support Vector Machines (SVM), introduced in the late 70s [39], are a set of related supervised learning methods widely used in pattern recognition [40], voice activity detection (VAD) [41] and classification [20]. SVM with linear discriminant functions define decision hypersurfaces or hyperplanes in a multidimensional feature space, that is:

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \omega_0 = 0, \quad (12)$$

where \mathbf{w} is known as the weight vector and ω_0 as the threshold. The weight vector \mathbf{w} is orthogonal to the decision hyperplane and the optimization task consists of finding the unknown parameters ω_i , $i=1, \dots, n$ defining the decision hyperplane.

Let \mathbf{x}_i , $i=1, 2, \dots, n$ be the feature vectors of the training set, X . These belong to either ω_1 or ω_2 , the two classes. If the classes were linearly separable, the objective would be to design a hyperplane that classifies correctly all the training vectors. Among the different design criteria, the maximal margin hyperplane is usually selected since it leaves the maximum margin of separation between the two classes. Since the distance from a point \mathbf{x} to the hyperplane is given by $z = |g(\mathbf{x})| / \|\mathbf{w}\|$, scaling w and w_0 so that the value of $g(\mathbf{x})$ is +1 for the nearest point in ω_1 and -1 for the nearest points in ω_2 , the optimization problem is reduced to minimizing a cost function $J(\omega) = 1/2 \|\omega\|^2$ subject to:

$$f_{svm}(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i \omega_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{x}) + \omega_0 \quad (13)$$

where α_i are the solution of a quadratic optimization problem that are usually determined by quadratic programming or the well-known sequential minimal optimization algorithm, and $\Phi(\mathbf{s})$ or $\Phi(\mathbf{x})$ denote the transformation of the feature

vectors into the effective feature space. This basic SVM classifier produces a linear separation hyperplane. A more general expression for SVM can be defined with the addition of kernels, that substitute the dot product of the transformed feature vectors $\Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{x}) = K(\mathbf{s}_i, \mathbf{x})$. Thus, the expression of the hyperplane in Eq. (12) can be replaced by:

$$g(\mathbf{x}) = \sum_{i=1}^{N_S} \alpha_i \omega_i K(\mathbf{s}_i, \mathbf{x}) + b \quad (14)$$

The kernel functions that we use in this work are the linear kernel $K(\mathbf{s}, \mathbf{x}) = \mathbf{s} \cdot \mathbf{x}$ (on Eq. (12)) and a Gaussian radial basis function (RBF) kernel $K(\mathbf{s}, \mathbf{x}) = \exp\{-|\mathbf{s} - \mathbf{x}|^2 / (2\sigma^2)\}$.

2.6. Evaluation

In this work, we will compare the two algorithms using accuracy, sensitivity and specificity measures, as long as values of Positive and Negative Likelihood (PL and NL) [42] for each algorithm and each dataset. These values are obtained using both k -fold and leave-one-out cross-validation methods [43]. k -Fold divides the entire dataset in k folds, then train the whole system with $k - 1$ and finally test the system with the remaining fold. This is performed k times, leaving one different fold out at a time, and then, average values of the evaluation parameters are estimated. Leave-one-out is a particular case of k -fold, in which $k = n$. In order to compare the proposed methods, we obtain values of accuracy, sensitivity, specificity, positive likelihood and negative likelihood in function number of selected voxels (N) and number of features extracted (K).

In clinical medicine, values of PL greater than 5 or NL values less than 0.2 can be applied to the pre-test probability of a patient having the disease tested for to estimate a post-test probability of the disease state existing [42]. A positive result for a test with a PL of 8 adds approximately 40% to the pre-test probability that a patient has a specific diagnosis.

3. Results

In this section the combination of different selection criteria and classifiers along with the feature extraction based on Independent Component Analysis will be analysed. The

following classifiers have been tested in order to compare the performance results: Linear and Quadratic Normal Multivariate Classifier (MV Linear and MV Quadratic) [44], Naive Bayes Classifiers with either linear and quadratic discriminant function (introduced above), a classifier that makes use of the Mahalanobis distance to compute the class membership [45] and SVM classifiers [46] with linear, quadratic, polynomial and Radial Basis Function kernels (SVM Linear, Quadratic, Polynomial and RBF).

To obtain a general view of the results that these different combinations yield, values of accuracy obtained for each combination of classifier, selection criteria and database are shown in Table 2. Peak values of accuracy could have been displayed in this table, but we have opted for averaging the 10% higher accuracy results obtained using each system, in order to provide us with an idea of both performance and stability, and thus, help us in the task of choosing the best system.

Regarding the SPECT database, best accuracy results are obtained by using Relative Entropy selection criteria, with values up to 96.9%. With this database, the best classifiers are the Multivariate and Bayesian Quadratic, and SVM-RBF. On the other hand, the ADNI database achieves its best accuracy results when using linear classifiers, along with the Wilcoxon and Relative Entropy selection criteria. Since the variance of parameters estimated using leave-one-out method might be high, a k -fold estimation with $k = 10$ may be the best option when choosing an statistical procedure [47]. Table 3 shows the average values obtained by k -fold using also the 10% higher accuracy scores for each combination.

The combinations that we had highlighted previously are confirmed by examining the above table, as the best results are obtained with the Relative Entropy criteria in both databases, and non-linear classifiers (for SPECT database) and linear classifiers (for ADNI database). Highest accuracy, sensitivity and specificity values for SPECT database are obtained with the Relative Entropy selection criteria and a Bayesian Quadratic classifier, with values of 0.958, 0.951 and 0.964, respectively. In the case of ADNI database, highest accuracy, sensitivity and specificity values are obtained with Relative Entropy and a SVM Linear classifier, yielding to values of 0.883, 0.861 and 0.905, respectively.

With all these measures, we consider the combination of entropy selection criteria and Bayesian Quadratic classifier the

Table 2 – Average values computed for the first 10% highest accuracy evaluations performed using leave-one-out cross-validation for different combinations of classifiers, selection criteria and databases.

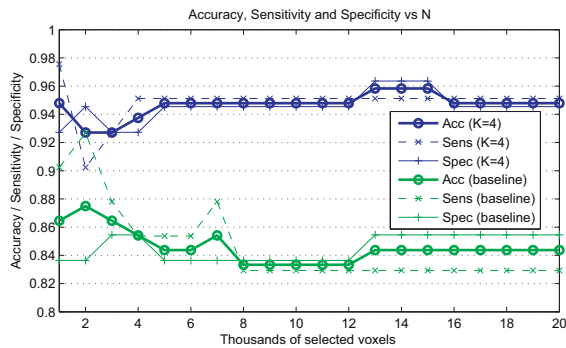
Classifier	SPECT			ADNI		
	t-Test	Entropy	Wilcoxon	t-Test	Entropy	Wilcoxon
MV Linear	0.902	0.896	0.902	0.887	0.894	0.893
Bayes Linear	0.923	0.916	0.896	0.835	0.878	0.893
MV Quadratic	0.935	0.942	0.924	0.877	0.888	0.889
Bayes Quadratic	0.934	0.951	0.924	0.816	0.884	0.873
Mahalanobis	0.895	0.892	0.896	0.857	0.842	0.851
SVM Linear	0.904	0.886	0.902	0.882	0.893	0.891
SVM Polinomial	0.888	0.899	0.885	0.852	0.865	0.865
SVM Quadratic	0.898	0.942	0.901	0.865	0.874	0.862
SVM RBF	0.928	0.953	0.916	0.866	0.879	0.859

Higher accuracy values are highlighted in bold for better understanding in the subsequent analysis.

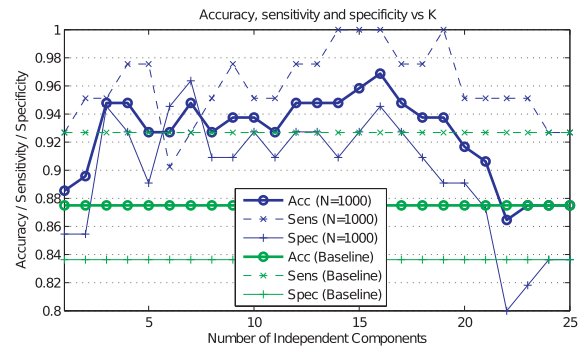
Table 3 – Average values computed for the first 10% highest accuracy evaluations performed using k-fold cross-validation for different combinations of classifiers, selection criteria and databases.

Classifier	SPECT			ADNI		
	t-Test	Entropy	Wilcoxon	t-Test	Entropy	Wilcoxon
MV Linear	0.901	0.899	0.905	0.889	0.894	0.889
Bayes Linear	0.924	0.915	0.894	0.834	0.878	0.893
MV Quadratic	0.934	0.941	0.924	0.872	0.886	0.888
Bayes Quadratic	0.933	0.950	0.924	0.817	0.886	0.873
Mahalanobis	0.899	0.890	0.900	0.857	0.841	0.852
SVM Linear	0.896	0.890	0.898	0.883	0.892	0.883
SVM Polynomial	0.888	0.900	0.882	0.849	0.864	0.866
SVM Quadratic	0.902	0.938	0.900	0.863	0.872	0.856
SVM RBF	0.929	0.951	0.917	0.863	0.878	0.857

Higher accuracy values are highlighted in bold for better understanding in the subsequent analysis.



(a) Accuracy, Sensitivity and Specificity in function of N



(b) Accuracy, Sensitivity and Specificity in function of K

Fig. 3 – Accuracy, sensitivity and specificity values for a Quadratic Bayes classifier in function of the number of selected voxels and the number of features extracted, using the SPECT database. The highest value obtained by the baseline method is presented for comparison purposes in Fig. 5(b).

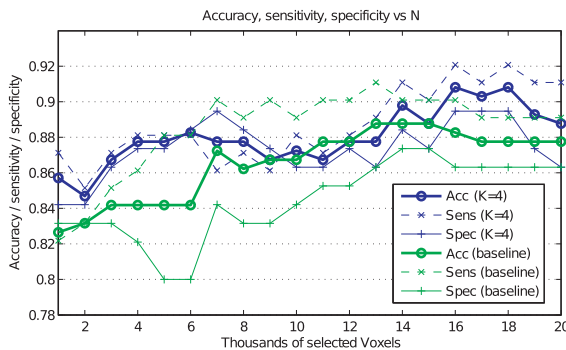
most robust system for the SPECT database. For the ADNI database, multivariate Lineal and SVM Linear classifier obtain similar results. Since the highest accuracy value is achieved with SVM linear, we finally decide on this.

3.1. SPECT

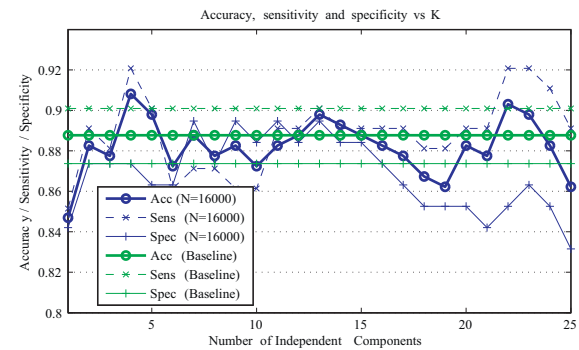
Fig. 3 depicts the accuracy, sensitivity and specificity results by means of the number of selected voxels as input vector to the feature extraction step and by means of the number

of components extracted by ICA. In these figures, the measures are presented along with the accuracy, sensitivity and specificity values obtained for the baseline method based on a Relative Entropy voxel selection.

Fig. 3(a) shows an interesting feature of this classifier. The final accuracy obtained is almost independent from the number of selected voxels. The obtained results are always better than the proposed baseline (SVAF) computed for the Relative Entropy criteria, and are always between 0.92 and 0.97. On the other hand, Fig. 3(b) depicts the behaviour of the cited



(a) Accuracy, Sensitivity and Specificity in function of N



(b) Accuracy, Sensitivity and Specificity in function of K

Fig. 4 – Accuracy, sensitivity and specificity values for a SVM linear classifier, in function of the number of selected voxels and the number of features extracted, using the ADNI database. The highest value obtained by the baseline method is presented for comparison purposes in Fig. 5(b).

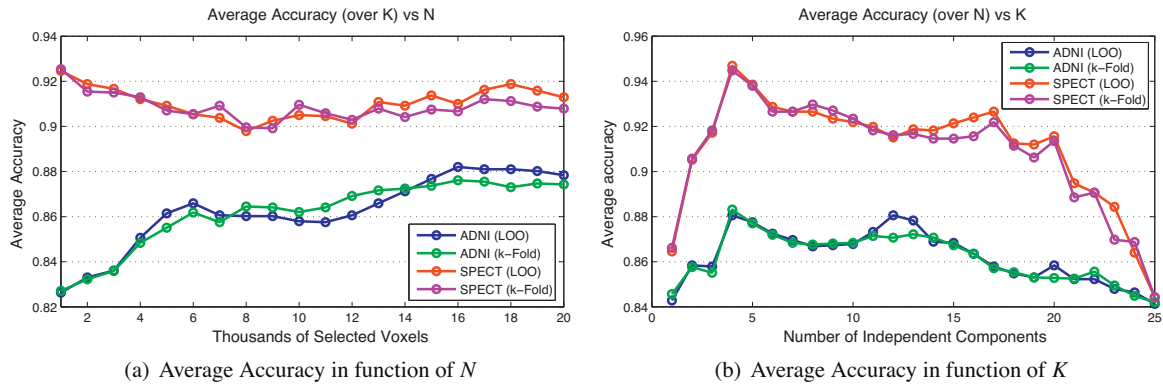


Fig. 5 – Average accuracy for each database and each evaluation method in function of the number of selected voxels (a) and the number of independent components extracted (b), computed over the number of independent components and the number of selected voxels, respectively.

parameters when varying K . The values of accuracy, sensitivity and specificity ranges from 0.9 to 0.97 in a wide range (for K between 3 and 20). The values obtained show the robustness of the proposed classifier for the SPECT database, and demonstrates its robustness in Alzheimer’s Disease patterns detection as it provides high accuracy rates with a wide range operation.

3.2. ADNI

For ADNI database, the behaviour is similar, but of lower values in evaluation parameters. Fig. 4 depicts this behaviour, and compares it to the values obtained when evaluating the Relative Entropy baseline method. The proposed method shows values that are similar to those obtained with the baseline method, and shows, in general, an ascending trend by increasing the number of selected voxels.

As commented above, accuracy and sensitivity values trends to higher values when increasing the number of voxel selected N , as shown in Fig. 4(a). Fig. 4(b) depicts these values in function of the number of features extracted K , and shows similar and occasionally lower values to those obtained for the baseline method. There are some increments on the evaluation parameters when using a K value between 12 and 15, and $K=4$, as happened with the SPECT database. Despite this, the benefits of using a feature extraction with the ADNI database are not significant, or at least not in terms of these evaluation parameters. In terms of computational efficiency, the method proposed reduces the number of parameters used as input to the classifier, although it takes a longer time to estimate the independent components, which results also in small performance improvements over the baseline method.

4. Discussion

It is interesting to note, before starting with this discussion, that the CAD systems proposed are trained to reproduce the actual medical knowledge, since they have been trained with images labelled by physicians. Images in SPECT database have been labelled by visual assessment and posterior agreement,

while images in ADNI database have been labelled using the scores obtained by patients in MMSE cognitive test, and not taking into account the visual information provided by PET images. Thus, these labels will be defined as our “Gold Standard” [48], and the statistical measures are an estimation of how a supervised learning system is able to reproduce a medical diagnosis carried out by experts. Also, the use of evaluation parameters that are sample prevalence dependent can be inappropriate. As our databases are not symmetric, other statistical values like Positive and Negative Likelihood Ratio (PL and NL) can be more reliable.

Fig. 5 depicts the average behaviour of the proposed system in function of the number of selected voxels and the number of independent component extracted. As commented above, the accuracy obtained when referring to the SPECT database is significantly constant in function of N . However, the accuracy obtained for the ADNI database trends to higher scores when increasing the number of selected voxels, which is logical, due to the higher amount of information contained in these voxels. For SPECT images, this behaviour can be caused by the higher level of noise contained [49]. As the number of selected voxels from the SPECT images increases, the level of noise included in the estimation of the independent components increases as well, so that the effect of noise counteract the theoretic benefits of increasing the number of voxels.

Concerning the value of K , Fig. 5(b) shows the values obtained for the SPECT and PET databases. This figure also shows an average common pattern that is followed by both databases: the system obtains the best values when K ranges from 4 to 15, with a significant peak in $K=4$. This peak can be caused because the meaningful estimation performed by ICA [32], as seen in Fig. 2, which show the 4 first components. These 4 components are related to typical AD patterns affecting the temporo-parietal region and the posterior cingulate gyri and precunei, considered as typically affected by glucose hypometabolism in PET images [50] or rCBF hypoperfusion in SPECT [36]. For these reasons, we will consider an operation point of $K=4$ for both databases, and $N=13$ for SPECT database and $N=18$ for ADNI database.

Table 4 compares the results obtained by the proposed system at the operation point with those obtained by the

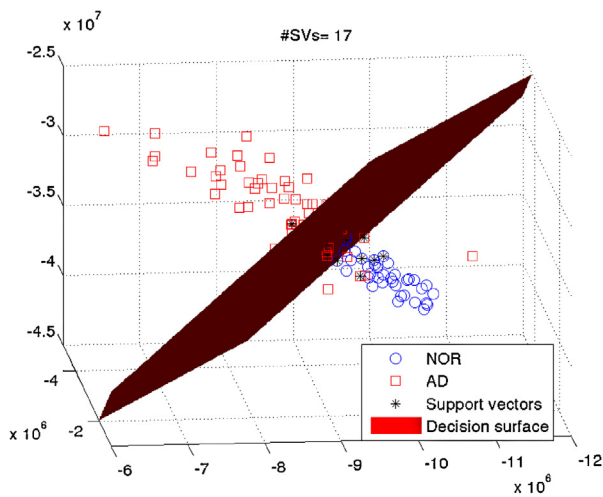
Table 4 – Results obtained for the proposed system at the operation point, compared to the baseline.

	Accuracy	Sensitivity	Specificity	PL	NL
ICA-Bayes-SPECT	0.958	0.951	0.964	26.159	0.051
ICA-SVM-SPECT	0.896	0.854	0.927	11.738	0.158
Baseline-SPECT	0.875	0.927	0.836	5.664	0.087
ICA-Bayes-ADNI	0.893	0.921	0.863	6.729	0.092
ICA-SVM-ADNI	0.908	0.921	0.895	8.748	0.089
Baseline-ADNI	0.888	0.901	0.874	7.133	0.113

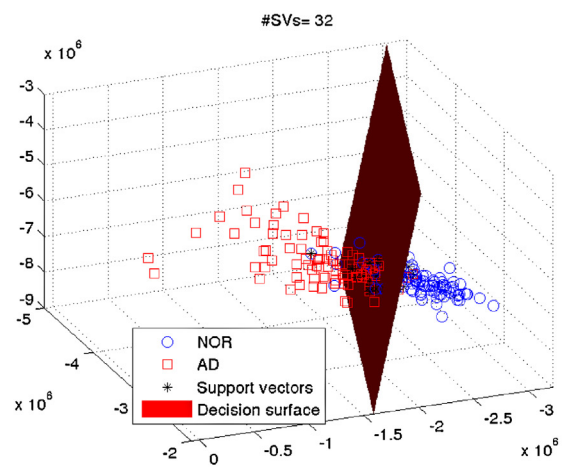
baseline approximation at its highest accuracy point for both databases. The improvements of the AD detection in SPECT images when using the Bayes classifier are patent, although the SVM linear classifier does not achieve these results. It maintains a good PL, but also the NL substantially increases, and though, the final accuracy slightly outperforms the baseline for only 0.02. Something similar holds for the ADNI database. Although high accuracy rates are obtained with the proposed system, the improvements of using a feature

extraction step over the baseline SVAF are not as significant, only a 0.02 higher, and the use of a Bayesian classifier even worsens the outcomes.

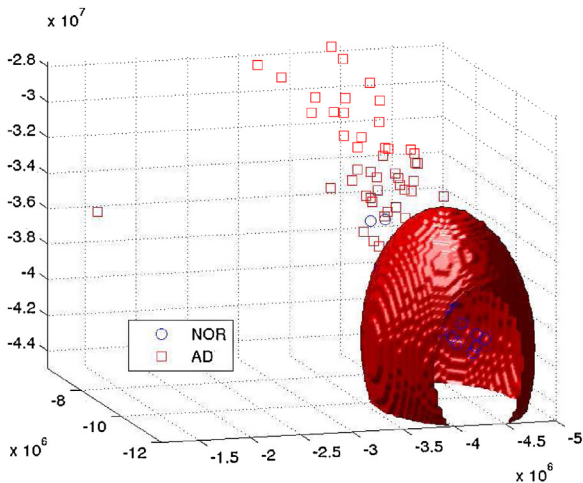
Probably, quadratic classifiers perform better in SPECT database due to the variability of classes included under the AD label (AD-1, which often corresponds to the Mild Cognitive Impairment (MCI), AD-2 and AD-3), while linear classifiers perform better in ADNI database, which only considers the Normal and AD classes, as the MCI class exposed in Section 2.1.2



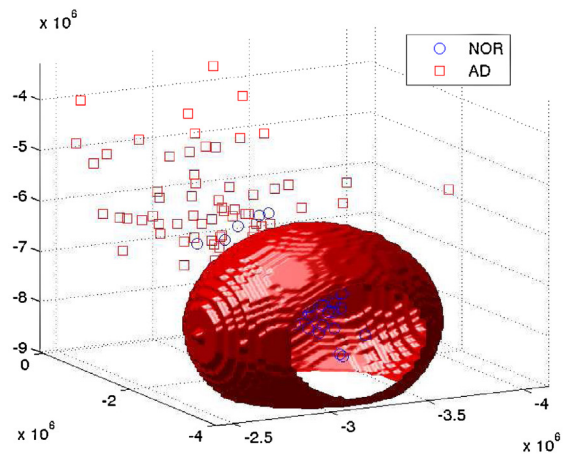
(a) Decision surface computed for VDLN and SVM-Linear



(b) Decision surface computed for ADNI and SVM-Linear



(c) Decision surface computed for VDLN and Bayes-Quadratic



(d) Decision surface computed for ADNI and Bayes-Quadratic

Fig. 6 – Value of the three first ICA projections of the images on the database, and decision surfaces designed by a SVM classifier with linear kernel ((a) for VDLN and (b) for ADNI) and a Naive Bayes Classifier with quadratic discriminant function ((c) for VDLN and (d) for ADNI). The number of voxels selected was $N = 16,000$ and the number of Independent Components was $K = 4$.

Table 5 – Results of the proposed system at the operation point compared with the SVAF baseline method, and methods proposed in the bibliography. For the ADNI database, we compare with the ICA-based method proposed in [15], Factor Analysis with SVM Linear [20], PCA with SVM-RBF [15] and VAF with SVM Linear [15]. For the SPECT database, we compare with Factor Analysis with SVM Linear [20], PCA with SVM Linear [17], GMM with SVM-RBF [52] and VAF with SVM Linear [51].

Database	Method	Accuracy	Sensitivity	Specificity	PL	NL
ADNI	ICA + Select	0.908	0.921	0.895	8.748	0.089
	ICA-SVM-RBF	0.871	0.864	0.878	6.439	0.141
	FA	0.913	0.901	0.926	12.228	0.107
	PCA-SVM-RBF	0.882	0.878	0.886	7.727	0.137
	SVAF-RE	0.888	0.901	0.874	7.133	0.113
	VAF-SVM-Lin	0.812	0.780	0.841	4.906	0.261
SPECT	ICA + Select	0.958	0.951	0.964	26.159	0.051
	FA	0.927	0.976	0.891	8.943	0.027
	PCA-SVM-Lin	0.897	0.893	0.902	9.152	0.119
	GMM-SVM-RBF	0.897	0.893	0.902	9.159	0.119
	SVAF-RE	0.875	0.927	0.836	5.664	0.087
	VAF-SVM-Lin	0.835	0.835	0.829	4.916	0.194

has not been included in the test. To illustrate this, a spatial view of the first three projections of each of the images in the space and the resulting decision surfaces of the classifiers are shown in Fig. 6. This higher separation between classes in ADNI database should lead to better results, but this does not hold. This is mainly due to our use of the “Gold Standard”, in which our final scores are computed using the labels of the databases. As the labelling in ADNI database does not make use of functional imaging, but of psychological tests, this can lead to worse results when analysing functional patterns.

Table 5 compares the proposed system with the defined baseline model and other methods proposed in the bibliography, such as the Voxels-as-Features (VAF) method, which considers all the voxels in an image as input features to the classifier [5,51], Principal Component Analysis (PCA) [17], Factor Analysis (FA) [20] or Gaussian Mixture Models (GMM) [52,51] in terms of accuracy, sensitivity, specificity, PL and NL.

VAF is considered as a baseline in many works, as different studies have concluded that this method is, at least, comparable with the visual exams performed by experts [5]. The proposed system outperforms in every evaluation parameter this baseline approximation, so improvements of our CAD system over the VAF approximation are present when applying to both databases.

When comparing to other methods in bibliography, the ICA based system is the best when applied to the SPECT database in terms of accuracy, sensitivity, specificity, PL and NL. The accuracy FA based system is 3% less than the accuracy obtained with the proposed system, which is equivalent to 3 more well-classified patients, and outperforms all the other methods proposed in the bibliography. The values of PL are more than three times the value considered to be a good class-discriminant, and the NL are far less than 0.1, considered also a good value.

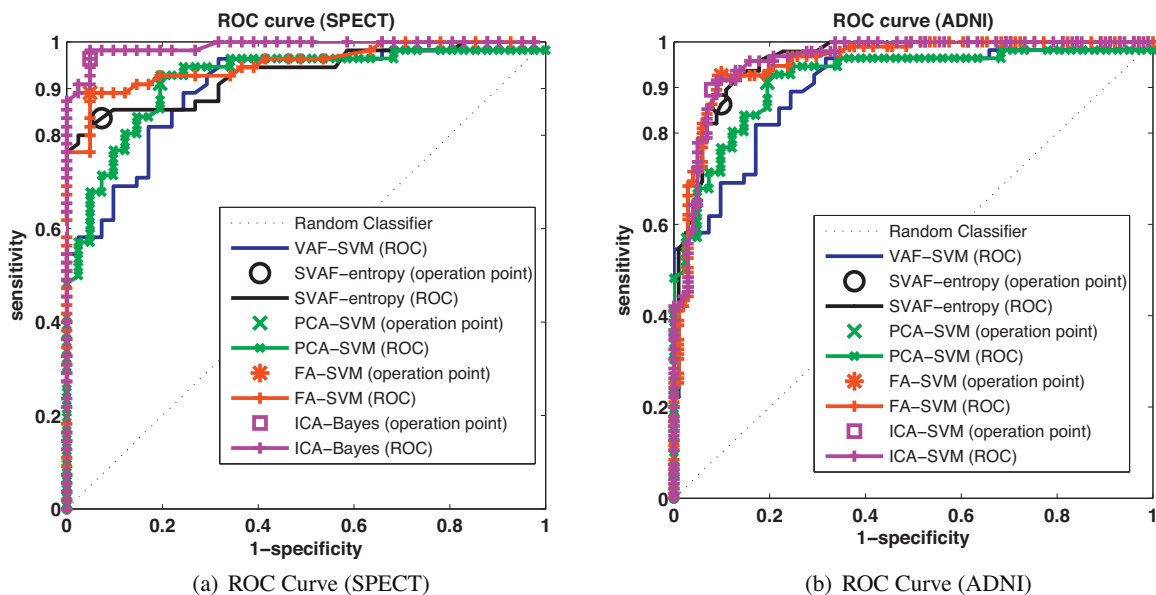


Fig. 7 – ROC curves obtained by the proposed system for SPECT (a) and ADNI (b) databases, the base systems VAF and SVAF and the PCA and FA based methods.

Regarding the ADNI database, the performance values obtained by the CAD that uses ICA outperforms most methods proposed in the bibliography, like the GMM, the ICA method in [15] and PCA. Nevertheless, the method based on Factor Analysis obtains similar values, with a slightly higher accuracy results, but a lower sensitivity in the AD pattern detection task.

Fig. 7 shows the ROC curves computed for each database when varying the distance from the points to the separation hyperplane of the classifier, and comparing the proposed methods with other methods in the bibliography, such as the FA and PCA based methods, and the VAF approximation. It compares also the proposed method with the baseline proposed. Fig. 7(a) shows the ROC curve for the SPECT database, and confirms the advantages of the proposed system over those proposed in the bibliography, yielding an Area Under the ROC Curve (AUC) of approximately 0.98. Fig. 7(b) shows that the proposed system and the methods in the bibliography, including the SVAF approximation, offer a similar performance when applied to the ADNI database. However, the system yet outperforms the VAF approximation, often considered a standard baseline for CAD systems, as indicated previously.

5. Conclusions

This work presents a new Computer Aided Diagnosis (CAD) system, which makes use of a combination of voxels selection based on statistical significance measures and feature extraction using Independent Component Analysis (ICA). The use of only a voxel selection step already outperforms the Voxels As Features (VAF) approximation, considered as a reference for many works.

The feature extraction using Independent Component Analysis allow us to extract highly representative features, which are closely related to typical Alzheimer's Disease (AD) patterns, and yielding on a high-accuracy classification. The resulting system performs specially well with SPECT database, probably due of the meaningful interpretation that ICA plays on the typical AD patterns when extracting the Independent Components, that also acts as a noise reduction system. When applied to the ADNI database, the resulting parameters were also high, but not significantly better than the proposed baseline, which uses the Selected Voxels As Features (SVAF) approximation.

The proposed system shows excellent accuracy values in the classification task with AD compared to Normal Control (CTRL) images, and displays a high robustness to variations in the input parameters, specially regarding the number of voxels selected. Even the proposed SVAF baseline method outperforms some of the methods proposed in the bibliography. In particular, the baseline Voxel As Features (VAF) approximation, took as a reference in numerous works and indicative of the effectiveness of visual examination of the images, was outperformed in every case by our baseline SVAF method and the proposed CAD system.

Conflict of interest

There is no conflict of interest.

Acknowledgements

This work was partly supported by the MICINN under the TEC2008-02113 and TEC2012-34306 projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Projects P07-TIC-02566, P09-TIC-4530 and P11-TIC-7103.

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; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514, and the Dana Foundation.

REFERENCES

- [1] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, W. Penny, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press, Amsterdam, 2007.
- [2] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, I.A. Illán, M. López, F. Segovia, R. Chaves, P. Padilla, C.G. Puntonet, Feature selection using factor analysis for Alzheimer's diagnosis using F-FDG pet images, *Medical Physics* 37 (11) (2010) 6084–6095.
- [3] A.P. Dhawan, A review on biomedical image processing and future trends, *Computer Methods and Programs in Biomedicine* 31 (3/4) (1990) 141–183, [http://dx.doi.org/10.1016/0169-2607\(90\)90001-P](http://dx.doi.org/10.1016/0169-2607(90)90001-P), URL <http://www.sciencedirect.com/science/article/pii/S016926079090001P>
- [4] P. Saxena, D.G. Pavel, J.C. Quintana, B. Horwitz, An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimer's disease, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, Vol. 1496 of *Lecture Notes in Computer Science*, Springer, Heidelberg, Germany, 1998, pp. 623–630.

- [5] 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*, Vol. 3217 of *Lecture Notes in Computer Science*, Springer, Heidelberg, Germany, 2004, pp. 654–662.
- [6] R.P.W. Duin, Classifiers in almost empty spaces, in: *Proceedings 15th International Conference on Pattern Recognition*, vol. 2, IEEE, 2000, pp. 1–7.
- [7] J.-B. Fasquel, V. Agnus, J. Moreau, L. Soler, J. Marescaux, An interactive medical image segmentation system based on the optimal management of regions of interest using topological medical knowledge, *Computer Methods and Programs in Biomedicine* 82 (3) (2006) 216–230, <http://dx.doi.org/10.1016/j.cmpb.2006.04.004>.
- [8] Z. Ji, Q. Sun, Y. Xia, Q. Chen, D. Xia, D. Feng, Generalized rough fuzzy c-means algorithm for brain MR image segmentation, *Computer Methods and Programs in Biomedicine* (2011), <http://dx.doi.org/10.1016/j.cmpb.2011.10.010>.
- [9] 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.
- [10] D. Salas-Gonzalez, J.M. Górriz, J. Ramírez, F. Segovia, R. Chaves, M. López, I.Á. Illán, P. Padilla, Selecting regions of interest in SPECT images using Wilcoxon test for the diagnosis of Alzheimer's disease, in: *HAI* (1), 2010, pp. 446–451.
- [11] W. Zhao, C. Wu, K. Yin, T.Y. Young, M.D. Ginsberg, Pixel-based statistical analysis by a 3d clustering approach: application to autoradiographic images, *Computer Methods and Programs in Biomedicine* 83 (1) (2006) 18–28, <http://dx.doi.org/10.1016/j.cmpb.2006.05.005>.
- [12] D. Shutin, O. Zlobinskaya, Application of information-theoretic measures to quantitative analysis of immunofluorescent microscope imaging, *Computer Methods and Programs in Biomedicine* 97 (2) (2010) 114–129, <http://dx.doi.org/10.1016/j.cmpb.2009.05.009>.
- [13] M.-Y. Lee, C.-S. Yang, Entropy-based feature extraction and decision tree induction for breast cancer diagnosis with standardized thermograph images, *Computer Methods and Programs in Biomedicine* 100 (3) (2010) 269–282, <http://dx.doi.org/10.1016/j.cmpb.2010.04.014>.
- [14] I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-González, M. López, F. Segovia, P. Padilla, C.G. Puntonet, Projecting independent components of SPECT images for computer aided diagnosis of Alzheimer's disease, *Pattern Recognition Letters* 31 (11) (2010) 1342–1347, <http://dx.doi.org/10.1016/j.patrec.2010.03.004>.
- [15] I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M.M. López, F. Segovia, R. Chaves, M. Gómez-Río, C.G. Puntonet, ¹⁸F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis, *Information Sciences* 181 (2011) 903–916, <http://dx.doi.org/10.1016/j.ins.2010.10.027>.
- [16] J. Ramírez, J. M. Górriz, D. Salas-González, 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*, in press. URL <http://www.sciencedirect.com/science/article/pii/S0020025509002291>
- [17] 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.
- [18] 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, URL <http://www.sciencedirect.com/science/article/B6WNP-4VCH6WN-1/2/fbde9dd344d47cc00f6923a23af3796a>
- [19] Z. Yan, Z. Wang, H. Xie, The application of mutual information-based feature selection and fuzzy LS-SVM-based classifier in motion classification, *Computer Methods and Programs in Biomedicine* 90 (3) (2008) 275–284, <http://dx.doi.org/10.1016/j.cmpb.2008.01.003>, URL <http://www.sciencedirect.com/science/article/pii/S0169260708000059>
- [20] F. Martínez-Murcia, J. Górriz, J. Ramírez, C. Puntonet, D. Salas-González, Computer aided diagnosis tool for Alzheimer's disease based on Mann–Whitney–Wilcoxon U-test, *Expert Systems with Applications* 39 (10) (2012) 9676–9685, <http://dx.doi.org/10.1016/j.eswa.2012.02.153>.
- [21] B.J.S.K.A. Jobst, L.P. Barnetson, Accurate prediction of histologically confirmed Alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria SPECT X-ray CT and APO e4 in medial temporal lobe dementias. Oxford project to investigate memory and aging, *International Psychogeriatrics* 10 (3) (1998) 271–302.
- [22] K.P.E.N.J. Dougall, S. Bruggink, Systematic review of the diagnostic accuracy of 99mTc-HMPAO-SPECT in dementia, *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry* 12 (6) (2004) 554–570.
- [23] R.P. Woods, Spatial transformation models, in: I.N. Bankman (Ed.), *Handbook of Medical Imaging*, Academic Press, San Diego, 2000, pp. 465–490, Chapter 29.
- [24] M.W. Weiner, D.P. Veitch, P.S. Aisen, L.A. Beckett, N.J. Cairns, R.C. Green, D. Harvey, C.R. Jack, W. Jagust, E. Liu, J.C. Morris, R.C. Petersen, A.J. Saykin, M.E. Schmidt, L. Shaw, J.A. Siuciak, H. Soares, A.W. Toga, J.Q. Trojanowski, The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception, *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 8 (1 Suppl.) (2012) S1–S68, <http://dx.doi.org/10.1016/j.jalz.2011.09.172>, PMID: 22047634, <http://www.ncbi.nlm.nih.gov/pubmed/22047634>
- [25] M.P. Fay, M.A. Proschan, Wilcoxon–Mann–Whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules, *Statistics Surveys* 4 (2010) 1–39.
- [26] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, Heidelberg, Germany, 2006, p. 55 (Chapter 3).
- [27] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Networks* 13 (4–5) (2000) 411–430, [http://dx.doi.org/10.1016/S0893-6080\(00\)00026-5](http://dx.doi.org/10.1016/S0893-6080(00)00026-5).
- [28] F. De Martino, F. Gentile, F. Esposito, M. Balsi, F. Di Salle, R. Goebel, E. Formisano, Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers, *NeuroImage* 34 (1) (2007) 177–194.
- [29] I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C.G. Puntonet, F. Segovia, Independent component analysis of SPECT images to assist the Alzheimer's disease diagnosis, in: *Advances in Neural Networks Research*, ISNN 2009. *Advances in Intelligent and Soft Computing*, wuhan, China, 2009.
- [30] I.A. Illán, Análisis en componentes de imágenes funcionales para la ayuda al diagnóstico de la enfermedad de Alzheimer, Ph.D. Thesis, Universidad de Granada (Junio 2009).
- [31] F.J. Theis, P. Gruber, I.R. Keck, E.W. Lang, Functional MRI analysis by a novel spatiotemporal ICA algorithm, in: *ICANN* (1), 2005, pp. 677–682.
- [32] A. Hyvärinen, Survey on independent component analysis, *Neural Computing Surveys* 2 (1999) 94–128.

- [33] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, *Neural Computation* 9 (1997) 1483–1492.
- [34] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transactions on Neural Networks* 10 (3) (1999) 626–634.
- [35] X. Giannakopoulos, J. Karhunen, E. Oja, Experimental comparison of neural ICA algorithms, in: *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, Skövde, Sweden, 1998, pp. 651–656.
- [36] 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 99m HMPAO in Alzheimer's disease. A population-based study, *Neurology* 44 (3) (1994) 454–461, <http://dx.doi.org/10.1159/000051283>.
- [37] K. Friston, C. Chu, J. Mourão-Miranda, O. Hulme, G. Rees, W. Penny, J. Ashburner, Bayesian decoding of brain images, *NeuroImage* 39 (1) (2008) 181–205, URL <http://www.sciencedirect.com/science/article/B6WNP-4PGY4T4-3/2/c373f34aec8d620248084fabca30fad0>
- [38] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1990.
- [39] V.N. Vapnik, *Estimation of Dependences Based on Empirical Data*, Springer-Verlag, New York, 1982.
- [40] S. Powell, V.A. Magnotta, H. Johnson, V.K. Jammalamadaka, R. Pierson, N.C. Andreasen, Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures, *NeuroImage* 39 (1) (2008) 238–247.
- [41] 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.
- [42] S. McGee, Simplifying likelihood ratios, *Journal of General Internal Medicine* 17 (8) (2002) 646–649, <http://dx.doi.org/10.1046/j.1525-1497.2002.10750.x>.
- [43] Y. Fan, S.M. Resnick, X. Wu, C. Davatzikos, Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study, *NeuroImage* 41 (2) (2008) 277–285.
- [44] W.J. Krzanowski (Ed.), *Principles of Multivariate Analysis: A User's Perspective*, Oxford University Press, Inc., New York, NY, USA, 1988.
- [45] G.J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley Interscience, New York, USA, 1992, ISBN 0-471-69115-1.
- [46] V.N. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, Inc., New York, 1998.
- [47] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection., in: *Proceedings of International Joint Conference on AI*, 1995, pp. 1137–1145, URL <http://citeseer.ist.psu.edu/kohavi95study.html>
- [48] D. Spiegelmann, S. Schneeweiss, A. McDermott, Measurement error correction for logistic regression models with an “alloyed gold standard”, *American Journal of Epidemiology* 145 (2) (1997) 184–196.
- [49] J. Ollinger, J. Fessler, Positron-emission tomography, *IEEE Signal Processing Magazine* 14 (1) (1997) 43–55, <http://dx.doi.org/10.1109/79.560323>.
- [50] K. Meguro, X. Blaizot, Y. Kondoh, C. Le Mestric, J.C. Baron, C. Chavoix, Neocortical and hippocampal glucose hypometabolism following neurotoxic lesions of the entorhinal and perirhinal cortices in the non-human primate as shown by pet, *Brain* 122 (8) (1999) 1519–1531, <http://dx.doi.org/10.1093/brain/122.8.1519>.
- [51] F. Segovia, J.M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, M. López, R. Chaves, P. Padilla, Classification of functional brain images using a GMM-based multi-variate approach, *Neuroscience Letters* 474 (2010) 58–72.
- [52] J.M. Górriz, F. Segovia, J. Ramírez, A. Lassl, D. Salas-Gonzalez, GMM based SPECT image classification for the diagnosis of Alzheimer's disease, *Applied Soft Computing* 11 (2) (2011) 2313–2325, <http://dx.doi.org/10.1016/j.asoc.2010.08.012>.