Expert Systems with Applications 39 (2012) 11766–11774

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Association rule-based feature selection method for Alzheimer's disease diagnosis ☆

R. Chaves [a,*], J. Ramírez [a], J.M. Górriz [a], C.G. Puntonet [b] for the Alzheimer's Disease Neuroimaging Initiative

[a] Dept. Signal Theory, Networking and Communication, ETSIIT, 18071 Granada, University of Granada, Spain
[b] Dept. Computer's Architecture and Technology, ETSIIT, 18071 Granada, University of Granada, Spain

ARTICLE INFO

Keywords:
Association rules
Feature selection and extraction
Alzheimer's disease
Computer aided diagnosis

ABSTRACT

A fundamental challenge that remains unsolved in the neuroimage field is the small sample size problem. Feature selection and extraction, which are based on a limited training set, are likely to display poor generalization performance on new datasets. To address this challenge, a novel voxel selection method based on association rule (AR) mining is proposed for designing a computer aided diagnosis (CAD) system. The proposed method is tested as a tool for the early diagnosis of Alzheimer's disease (AD). Discriminant brain areas are selected from a single photon emission computed tomography (SPECT) or positron emission tomography (PET) databases by means of an AR mining process. Simultaneously activated brain regions in control subjects that consist of the set of voxels defining the antecedents and consequents of the ARs are selected as input voxels for posterior dimensionality reduction. Feature extraction is defined by a subsequent reduction of the selected voxels using principal component analysis (PCA) or partial least squares (PLS) techniques while classification is performed by a support vector machine (SVM). The proposed method yields an accuracy up to 91.75% (with 89.29% sensitivity and 95.12% specificity) for SPECT and 90% (with 89.33% sensitivity and 90.67% specificity) for PET, thus improving recently developed methods for early diagnosis of AD.

1. Introduction

Because of the increasing life expectancy, the incidence of dementia is expected to double during the next 20 years with important and dramatic health as well as socio-economic implications particularly in the United States, Europe, and Japan (Muellera et al., 2005). The essential treatment for Alzheimer's disease (AD) has not been established yet. However, it is possible to delay progression of symptoms by effective enforcement of medicine and nonmedicine treatment based on early diagnosis of AD (Ito, 2006). At the research level, structural and functional neuroimaging studies are the unique methodologies allowing the *in vivo* study of brain pathology at macro and micromolecular level. Functional brain imaging techniques, such as single photon emission computed tomography (SPECT), provide functional information, i.e. regional cerebral blood flow (rCBF) or metabolic activity, and enable identifying pathologic anomalies in internal tissues or organs, even before anatomical and structural alterations are observable (Ramírez et al., in press). On the other hand, positron emission tomography (PET) functional image modality measures the rate of glucose metabolism with the tracer [^{18}F] Fluorodeoxyglucose. In AD, characteristic brain regions show decreased glucose metabolism, specifically bilaterally regions in the temporal and parietal lobes, posterior cingulate gyri and precunei, as well as frontal cortex and whole brain in more severely affected patients (Illán et al., 2011). These anomalies are frequently minimal changes in late-onset AD that make visual diagnosis a difficult task even for experienced explorers. Even with this problem still unsolved, the potential of data mining technologies for computer aided diagnosis (CAD) has not been explored in depth (Chaves et al., 2009). Several approaches for designing AD CAD systems can be found in the literature (López et al., 2009; Ramírez et al., 2009). The first approach (univariate) is the statistical parametric mapping (SPM) (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007), which treats each voxel more or less independently, and hence cannot define the functional connectivity in the brain directly. Related to this work, the second approach, multivariate pattern analysis (MVPA), is based on the analysis of regions of interest (ROIs) by means of some discriminant functions. There is increasing interest in applying multivariate approaches to study the patterns of neurodegenerative diseases and mental disorders (Chu, Hsu, Chou, Bandettini, & Lin, 2012), which had been applied to Voxel Based Morphometry (VBM), especially studies on

☆ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (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: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

* Corresponding author.
E-mail address: rosach@ugr.es (R. Chaves).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.04.075

classification of AD have grown significantly in recent years. The key of these methods is their potential to detect global, complex and distributed patterns of abnormalities that cannot be efficiently identified by means of univariate voxel-based methods (Duchesnay et al., 2011).

The high dimensionality of the input feature space in comparison with the relatively small number of subjects (curse of dimensionality) is a widespread concern, so some form of feature selection (FS) is often applied. Given an input set of features, dimensionality reduction can be achieved in two different ways. The first is to select the hopefully best subset of features of the input feature set, also called FS. As an example, association rules (ARs) were proposed in Karabatak and Cevdet Ince (2009) as an FS methodology which takes the associations and/or relationships in erythemato-squamous disease data eliminating in this way unnecessary inputs. The second approach or feature extraction (FE) creates new features based on transformation (linear or non-linear) from the original features to a lower dimensional space. Some of the most popular FE techniques for dealing with this problem are principal component analysis (PCA) method or partial least squares (PLS) among others. PCA generates an orthonormal basis vector that maximizes the scatter of all the projected samples, which is equivalent to finding the eigenvalues of the covariance matrix (López et al., 2011). PLS represents a wide class of methods for modeling relations between sets of observed variables by means of latent variables (Ramírez et al., 2010).

In this study, an AR-based FS method (similar to the one described in Karabatak & Cevdet Ince (2009)) in combination with FE techniques such as PCA (López et al., 2009) or PLS (Ramírez et al., 2010) is proposed as dimensionality reduction method previous to a support vector machine (SVM) (Chaves et al., 2009) classifier for AD early diagnosis (Sigut, Pineiro, Gonzalez, & Torres, 2007). The paper is organised as follows. Section 2 shows an introduction to feature selection methods. Section 3 shows the basis of AR mining. The SPECT and PET image databases used to evaluate the proposed method are described in Section 4. Section 5 shows the voxel selection method based on activation estimation, AR mining concepts, PCA/PLS for FE and kernel SVM classification. Finally, the evaluation experiments are shown and discussed in Section 6 and conclusions are drawn in Section 7.

2. Background on feature selection

The high dimensionality of the input feature space in comparison with the relatively small number of subjects (Costafreda, Chu, Ashburner, & Fu, 2009) is a widely known problem called *curse of dimensionality*, so some form of feature selection is often applied in order to solve it. In addition, FS has benefits as it speeds up the testing process and increases classification accuracy (Duchesnay et al., 2011; Zhao, Fu, Ji, Tang, & Zhou, 2011). The development of feature selection has two major directions: subset selection and feature ranking. In the feature subset selection problem, a learning algorithm is faced with the problem of selecting a relevant subset of features upon which to focus its attention, while ignoring the rest (Kohavi & John, 1997). Subset selection typically falls into three categories: filters, wrappers and embedded (Maldonado, Weber, & Basak, 2011). Filter method uses statistical properties of the features in order to filter out poorly informative ones (Liu et al., 2005). Wrapper explores the whole set of variables to score feature subsets according to their predictive power, optimizing a performance criterion of the subsequent algorithm that uses the respective subset for classification (Kohavi & John, 1997). The last approach (embedded methods) performs FS in the process of model construction. For example, an embedded method for feature selection using SVMs is proposed in Maldonado et al. (2011).

As feature ranking methodology is concerned, the method makes use of ranking information, instead of simply viewing the ranks as flat categories. The final objective is to find a set of features with maximum importance and minimum similarity, that is, it tries to avoid selecting redundant features providing a greedy search algorithm to solve the optimization problem (Geng, Liu, Qin, & Li, 2007).

3. Background on association rules

ARs find interesting associations and/or relationships among input items of any system in large databases and later eliminate some unnecessary input (Chaves et al., 2011). ARs have drawn researcher's attention in the past (Agrawal & Srikant, 1994) and are typically used in market basket analysis, cross-marketing, catalog design, loss-leader analysis, store layout and customer buying pattern. In addition, ARs are a promising alternative in medical image classification and deserves further attention. For example, for diagnosis of cancerous cell in CT-Scan Brain images, FP-tree method has been used to find the frequent pattern for building the ARs (Rajendran & Madheswaran, 2010). In Zaiane, Antonie, and Coman (2002), when a mammogram has to be classified, the categorization system returns the ARs that applies to that image. The concepts of AR using object, object-based attribute and relationship predicate, and image-based attribute in medical images are extended in Pan, Li, and Wei (2005). Ribeiro, Marques, Traina, and Traina (2006) aims at improving the precision of medical retrieval by content, introducing an approach that combines techniques of statistical AR mining and feedback. However, AR application to AD is still a challenge. In AR-learning context, m activated regions called items $I = \{i_1, i_2, \ldots, i_m\}$ are extracted for defining implications of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$. Apriori is a state-of-the-art algorithm since most of the AR algorithms are variations of this (Agrawal & Srikant, 1994). Apriori works iteratively in order to identify the frequent itemsets and prepare for generating strong association rules. To perform a search, the user has to specify min support (*minsupp*) and min confidence (*minconf*) for frequent itemsets (He, Xiong, Yang, & Park, 2011). Each rule has support s if s% of the transactions contain $X \cup Y$ and a confidence c, if c% of the transactions that contain X also contain Y, that is (Chaves et al., 2011):

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} \quad (1)$$

There are different measures of significance for AR selection (Tan, Kumar, & Srivastava, 2002). Note that, while the support is a measure of the frequency of a rule, the confidence is a measure of the strength of the relation between sets of items (Dasseni, Verykios, Elmagarmid, & Bertino, 2001).

4. Databases

4.1. SPECT database

SPECT images used in this work were acquired by means of a PRISM 3000 gammacamera after injecting a gamma emitting technetium-99 m labeled ethyl cysteinate dimer (^{99m}Tc-ECD) to each subject. These images were reconstructed from projection data by filtered backprojection (FBP) in combination with a Butterworth noise filter. Then, SPECT images were spatially normalized (Salas-Gonzalez, Górriz, Ramírez, Lassl, & Puntonet, 2008) in order to ensure that a given voxel in different images refer to the same anatomical position. This process was done by using statistical parametric mapping (SPM) (Friston et al., 2007) yielding $69 \times 95 \times 79$ normalized SPECT images.

A direct comparison of the voxel intensities of the images of different subject is not possible without normalization of the intensities. Intensity level of the images is normalized to the maximum intensity, which is computed for each image individually by averaging over the 0.1% of the highest voxel intensities as in López et al. (2009).

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.[1] 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 1a for demographic details). Since the patients are not pathologically confirmed, the subject's labels possess 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 labeling methods, in contrast to neuropsychological tests (Dubois et al., 2007; Jobst, Barnetson, & Shepstone, 1998). 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. This work does not imply any experimental intervention and has been performed under the approval and supervision of the Clinical and Investigation Ethical Commission of the University Hospital Virgen de las Nieves (CEIC).

4.1.1. PET database

PET data was obtained from the ADNI Laboratory on NeuroImaging (LONI, University of California, Los Angeles) website (http://www.loni.ucla.edu/ADNI/). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million, 5-year public- private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many 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 to 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 http://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. Following the scan, each image was reviewed for possible artifacts at the University of Michigan and all raw and processed study data was archived. Subsequently, the images were normalized through a general affine model, with 12 parameters (Salas-Gonzalez et al., 2008) using the SPM5 software. 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 lowest-frequency components of the three-dimensional cosine transform bases (Ashburner & Friston, 1999). A small-deformation approach was used, and regularization was by the bending energy of the displacement field, ensuring that the voxels in different FDG-PET images refer to the same anatomical positions in the brains. After spatial normalization, an intensity normalization was required in order to perform direct images comparisons between different subjects. The intensity of the images was normalized to a value I_{max}, obtained averaging the 0.1% of the highest voxel intensities exceeding a threshold. The threshold was fixed to the 10th bin intensity value of a 50-bins intensity histogram, for discarding most low intensity records from outside-brain regions, and preventing image saturation. Participant's enrolment was conditioned to some eligibility criteria. General inclusion–exclusion criteria were as follows:

- Normal control subjects: Mini Mental State Examination (MMSE) scores between 24–30 (inclusive), a Clinical Dementia Ratio (CDR) of 0, non depressed, non MCI, and non demented. The age range of normal subjects will be roughly matched to that of MCI and AD subjects. Therefore, there should be minimal enrolment of normals under the age of 70.
- Mild AD: MMSE scores between 20–26 (inclusive), CDR of 0.5 or 1.0, and meets NINCDS/ ADRDA criteria for probable AD.

The PET database collected from ADNI consists of 150 labeled PET images: 75 control subjects and 75 AD patients (see Table 1b

Table 1a
Demographic details of the SPECT dataset.

	Demographic details		
	Num. of samples	Sex (M/F) (%)	Age μ [range/σ]
CTRL	41	32.95/12.19	71.51 [46–85/7.99]
AD 1	30	10.97/18.29	65.20 [23–81/13.36]
AD 2	22	13.41/9.76	65.73 [46–86/8.25]
AD 3	4	0/2.43	76 [69–83/9.90]

AD 1 = mild perfusion deficit, AD 2 = moderate deficit, AD 3 = severe deficit. μ and σ stands for population mean and standard deviation, respectively.

Table 1b
Demographic details of the PET dataset.

	Demographic details		
	Num. of samples	Sex (M/F) (%)	Age μ [range/σ]
CTRL	75	58.67/41.33	75.97 [62–86/4.91]
AD	75	62.67/37.33	75.72 [55–88/7.40]

μ and σ stand for population mean and standard deviation, respectively.

[1] Clinical information is unfortunately not available for privacy reasons, but only demographic information.

for demographic details). ADNI patient diagnostics are not pathologically confirmed, introducing some uncertainly on the subject's labels. Using these labels, allows to test the robustness of the classifier. This should be also considered when comparing to other methods tested on autopsy confirmed AD patients, on which every classifier is expected to improve its performance (Illán et al., 2011).

5. Description and methodology of the CAD system

As pattern recognition techniques were originally not designed to cope with large amounts of irrelevant features (Saeys, Inza, & Larranaga, 2007), ARs in combination with PCA or PLS are proposed in this work as an effective method to obtain a reduced dimension feature vector addressing the sample size problem. This problem arises whenever the number of samples is smaller than the dimensionality of the feature space (Chen, Liao, Ko, Lin, & Yu, 2000).

Fig. 1 shows a block diagram of the proposed system that consists of four stages: (i) preprocessing, (ii) voxel selection, (iii) feature extraction, and (iv) SVM classification. The preprocessing stage performs spatial and intensity normalization of SPECT or PET data. Voxel selection is enabled by activation estimation (AE) and AR mining techniques. Then PCA or PLS are used to further reduce the dimension of the input feature vector to a kernel SVM. The proposed method is described as follows.

5.1. Activation estimation

In the first step (preprocessing) and after functional images are normalized, a 3D mask is computed by averaging all control images from the database. An activation threshold intensity a_T is fixed by 50% of the maximum intensity of this mean image, defining a 3D mask. Voxels outside this mask are discarded from the database images. In this way, voxels outside the brain and poorly activated regions are excluded from this analysis. Then, an image voxel is considered as activated when its intensity value I is above the activation threshold, that is, when $I > a_T$.

Voxel selection by means of AR mining is addressed by first dividing each image into 3D $v \times v \times v$ cubic blocks. The number of blocks in which an image is divided depends on the size of the blocks (v) and the overlap between adjacent blocks, g, defined in terms of number of voxels. In this way, the centers of the blocks considered define a 3D $g \times g \times g$ grid, considering the possibility of some overlap between adjacent blocks. AE parameterizes a block by the proportion of activated voxels t. It is considered that a *block is activated* if the ratio of activated voxels inside of it is greater than a given threshold in a method similar to the one described in Turkeltaub, Eden, and Zeffiro (2002). It has been considered that a block is activated only if the ratio of activated voxels inside of it is greater than a given threshold, i.e. $t > 90\%$. This provides a good trade-off between computational cost and accuracy and allows the inclusion of all relevant brain regions in the detection of AD.

5.2. Voxel selection by AR mining and combination

Since the functional image pattern of control subjects shows reduced variability when compared to AD patterns, the Apriori algorithm is used to mine ARs identifying relevant relationships among simultaneously activated brain regions from control subjects. AR implications are established in terms of antecedents and consequents (see Section 3 for generic information of AR) between the previously obtained 3D activated blocks. The relevance of an AR is defined in terms of its support and confidence values which are always above a selected minimum confidence (*minconf*) and minimum support (*minsupp*). In addition, ARs are mined by means of leave-one-out (loo) cross-validation strategy (Gidskehaug, Anderssen, & Alsberg, 2008), that is, if a given control subject is being evaluated, it is not taken into account in the AR mining process in order to avoid overtraining.

The voxel selection algorithm considers just the voxels that are involved as antecedents or consequents of the mined ARs in the previous step. Voxels are selected by AR combination stage, that is, the antecedents and consequents of previously obtained ARs are merged without repetition and the output of the AR-enabled voxel selection method is the activation level of these ROIs.

The effect of varying *minsupp*, *minconf* and the activation threshold is shown in Fig. 2 which represents the voxels selected for three axial, sagittal and coronal slices. Note that, reducing the activation threshold t, *minsupp* and *minconf* increases the number of voxels selected by the proposed method. These parameters need to be selected in order to attain the best trade-off between retaining the most discriminant regions characterizing the disease and reducing the computational cost of the system. This effect is shown in the first and second rows of Fig. 2 that corresponds to maximum AR supports and confidences and varying activation thresholds. Note that, these ROIs correspond to those regions commonly affected by AD, i.e., the posterior cingulate gyri and precunei, as well

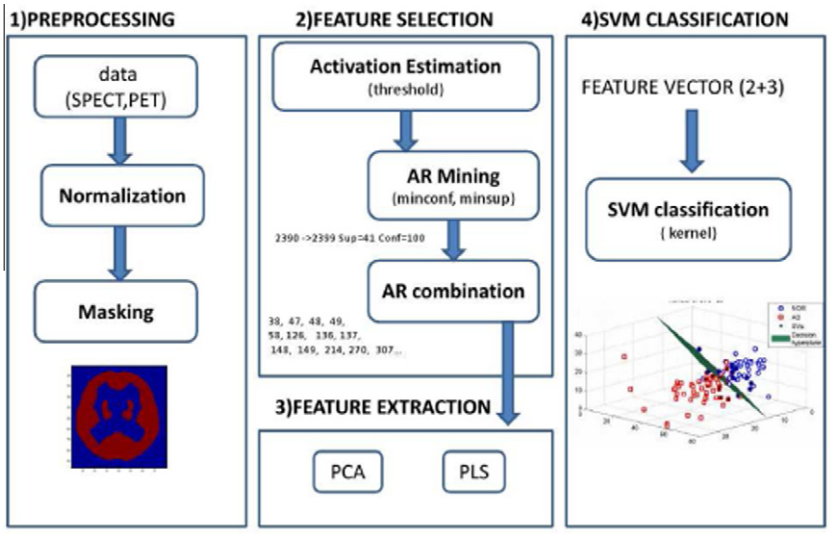

Fig. 1. Pipeline with the preprocessing, feature selection, feature extraction and classification stages of the proposed CAD system.

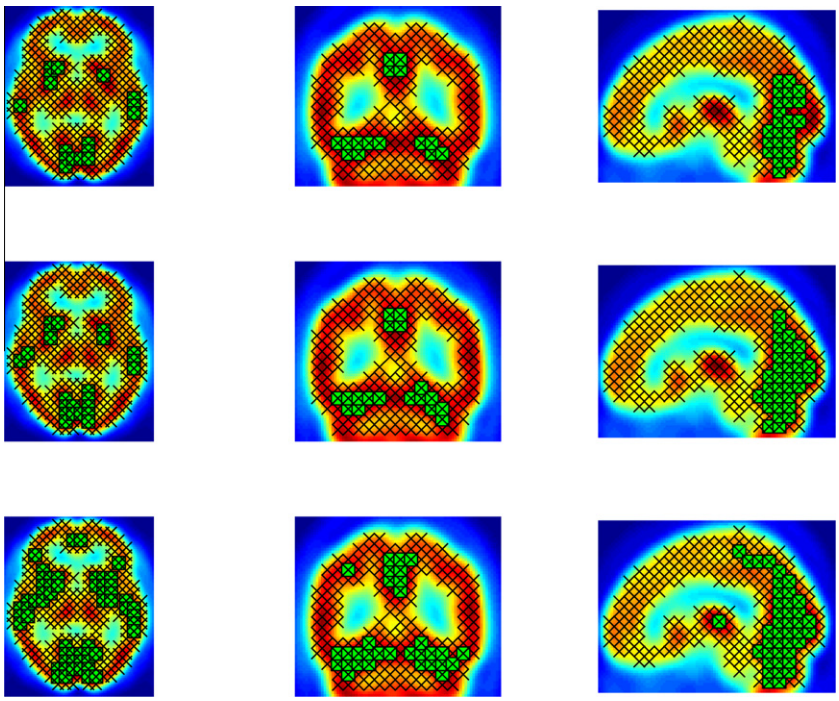

Fig. 2. Voxels selected by the proposed AR-mining selection method in axial, sagittal and coronal slices for SPECT database: (a) first row: *minsupp* = 100%, *minconf* = 100%, *t* = 92%, (*b*) second row: *minsupp* = 100%, *minconf* = 100%, *t* = 90%, and (*c*) *minsupp* = 90%, *minconf* = 90%, *t* = 92%.

as the temporo-parietal region (López et al., 2011). Nonetheless, as the *minsupp* and *minconf* parameters are reduced, more rules are obtained and consequently more voxels are selected. However, most of the additional ARs represents brain areas that are not discriminant of the disease and could slow down the speed of the CAD system working without increasing its accuracy rate (see experimental section for further details).

5.3. Feature extraction and classification

Dimensionality reduction is performed by means of PCA (López et al., 2011) or PLS (Ramírez et al., 2010). The resulting set of features may provide a better discriminant ability than the subset of given features selected by ARs.

Finally, classification is performed by a SVM classifier (Zhao et al., 2011) which has been successfully used in a number of applications (Chaves et al., 2009; Ramírez et al., in press). Moreover, recent developments in defining and training statistical classifiers make it possible to build reliable classifiers in very small sample size problems and even may find non-linear decision boundaries for small training sets.

6. Experimental results

6.1. Experiments on SPECT database

Several experiments were conducted in order to test the reliability of the proposed CAD system whose explanatory diagram is outlined in Fig. 1. After spatial and intensity normalization of SPECT data (see Section 3 for further details), 3D-Blocks of $9 \times 9 \times 9$-voxel located inside the corresponding mask were obtained. The center coordinates of the 3D blocks are restricted to be in $4 \times 4 \times 4$ 3D grid. At the activation estimation stage, a 3D block is defined as activated if more than a fraction (650 voxels) of the total (729 voxels) are activated. This threshold *t* represents a percentage above 90% of activated voxels, since it allows only brain volumes with a high activation to be considered as input for the AR-mining.

In the AR-mining stage, different *minconf* and *minsupp* values (from 100% to 90%) were considered as input for the Apriori algorithm. The most discriminant rules are obtained when selecting the highest minsupp and minconf values. In addition, the computational cost is improved by reducing the number of selected features.

Fig. 3 shows the fraction (%) of AR-based selected ROIs from the total of 2445 ROIs when percentages of 90% and 92% of activation are used. The number of rules is progressively reduced in a similar way to a negative exponential distribution from 90% to 100% of minsupp and minconf. While at 90% of minsupp and minconf, the fraction (%) of AR combination selected from the total of ROIs reaches 18.85% and 14.88% (for activation of 90% and 92%), at 100% this number appears reduced to 10.22% or 8.05%,

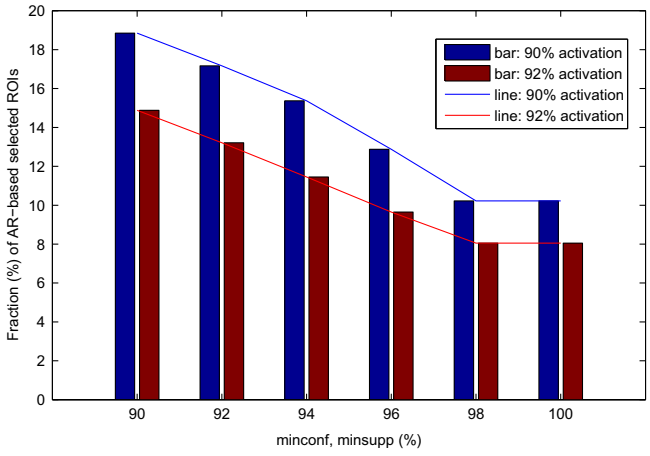

Fig. 3. Fraction (%) of AR-based selected ROIs versus minsupp and minconf (%) for different parameters: activation threshold and minconf/minsup. SPECT database.

Table 2
Comparison of Acc/Sens/Spe rate and AUC for different FE techniques with final reduction of 6 features (SPECT database).

	PCA: Acc/Sen/Spe (%)	PLS: Acc/Sen/Spe (%)	AUC: PCA/PLS
AR (90% AE, 100% minconf, minsup)	89.69/87.5/92.68	90.72/89.29/92.68	0.9549/0.9259
AR (90% AE, 95% minconf, minsup)	87.62/85.71/90.24	88.66/91.07/85.37	0.9017/0.9011
AR (90% AE, 90% minconf, minsup)	84.53/83.93/85.37	86.60/87.50/85.37	0.8915/0.8924
AR (92% AE, 100% minconf, minsup)	85.56/82.14/90.24	87.63/83.93/92.68	0.9277/0.9150
t-test	81.44/82.14/80.49	85.57/87.50/82.93	0.9317/0.9029
Wilcoxon	79.38/80.36/78.05	89.69/91.07/87.80	0.9205/0.9230
Energy	53.61/92.86/0	60.82/92.86/17.07	0/0

respectively. Despite decreased the number of merged ARs as the activation level increases above 90%, this provokes the losing of some discriminant rules for the AD early diagnosis in the light of the results (see Table 2 or Figs. 4 and 6 for further details) when 92% of activation is used. Although this reduction is considerable, a subsequent PCA or PLS feature extraction is necessary to improve the computational time and accuracy rate of the system.

The accuracy (Acc), sensitivity (Sen) and specificity (Spe) are estimated by means of *k*-fold cross-validation with a linear SVM classifier. Note that, sensitivity of a classifier measures the proportion of true-positive correctly detected, while specificity is defined as the proportion of true negatives detected as such. Fig. 4a and 4b considered an increasing number of PCA or PLS features, respectively yielding maximum Acc, Sen and Spe rates of 90.72%, 89.29%, 92.68% at 3 PCA features or 91.75%, 89.29%, 95.12% at 12 PLS features respectively outperforming other recently reported methods such as t-student (Chaves et al., 2009), wilcoxon (Di Bucchianico, 1999) or entropy tests (Martínez et al., 2011). The first (t-student) is a two-sample t-test with pooled variance estimate with feature correlation weighting. Mann–Whitney-Wilcoxon test (Martínez-Murcia, Górriz, Ramírez, Puntonet, & Salas-Gonzalez, 2012) and Relative Entropy gives us information about voxel class separability but with different ranking criteria.

FE dimensionality reduction requires an effective strategy for finding the optimal subset of features or the optimal transformation to make a reliable comparison with other baseline methods. Fig. 4c shows the percentage of variance explained for extracted features drawn as bars and a line representing the cumulative variance explained. Variance explained accounts for the variation of a feature subset when PCA or PLS strategies are applied and in this case it does not change significantly above 6 components.

Table 2 shows the Acc, Sen and Spe of the proposed AR-based FS method and comparison with other FS baseline techniques at 6 features. Acc, Spe, Sen values converge to 89.69%, 87.5% and 92.68% for PCA and 90.7%, 92.68% and 89.29% for PLS when a minsupp and minconf of 100% and an AE threshold of 90% are used, which outperform the other baseline FS-based methods.

6.2. Experiments on PET database

While voxel intensity is related to blood flow in SPECT imaging, for PET image modality the activation level is related to glucose metabolism. Generally speaking, PET images are experimentally found to be less activated than SPECT images in each ROI. For this reason 3D-Blocks are selected in these experiments with a smaller size, i.e. $5 \times 5 \times 5$-voxels per block, with center coordinates restricted to be in a $2 \times 2 \times 2$ grid. A voxel activation above 40% allows including relevant ROIs in the AR-based feature selection. In a similar way to SPECT modality, the number of AR-based selected ROIs decreases more at 100% of minconf and minsupp from the total of 3776. In particular, the fraction of AR-based selected ROIs is 4.87% or 6.09% when activation is of 44% or 40%, respectively. Although at 44% of activation the reduction is higher, the system discards some relevant ROIs as shown in accuracy results of

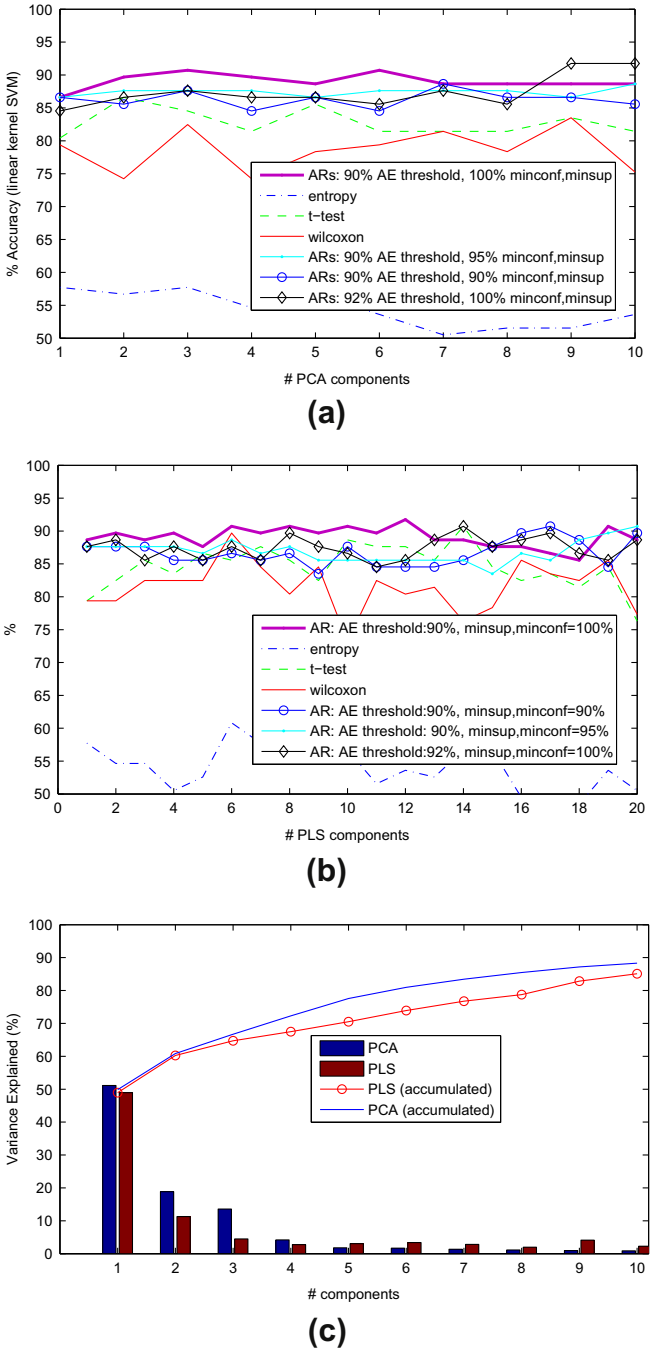

Fig. 4. Accuracy values as a function of the number of components (SPECT database): (a) PCA, and (b) PLS. (c) Variance explained (%) as a function of the number of PCA and PLS components in a bar diagram (lines represent the accumulated variance explained.)

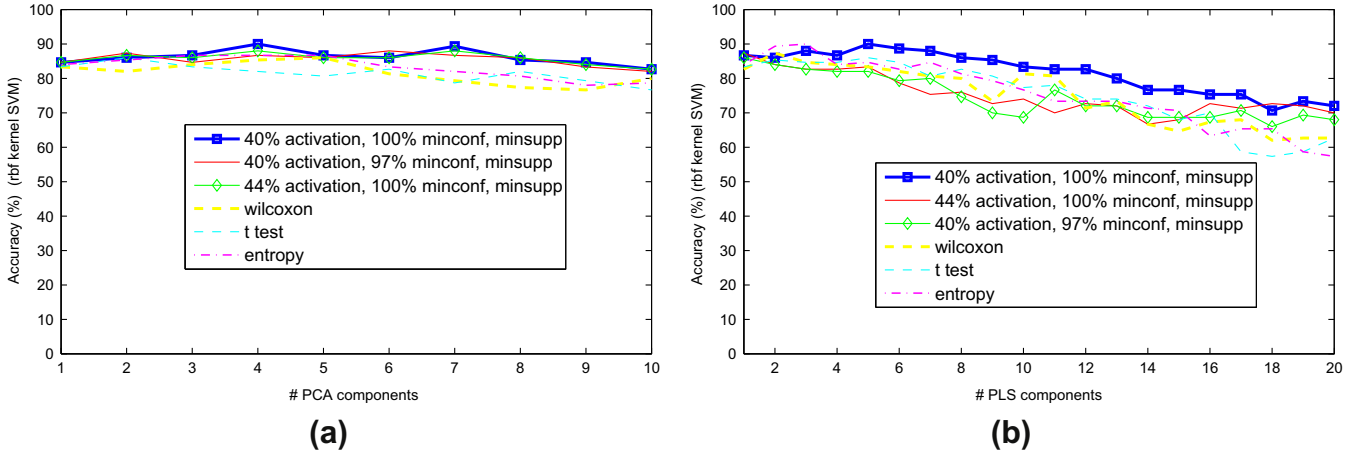

Fig. 5. Accuracy values as a function of the number of components for PET database: (a) PCA, and (b) PLS.

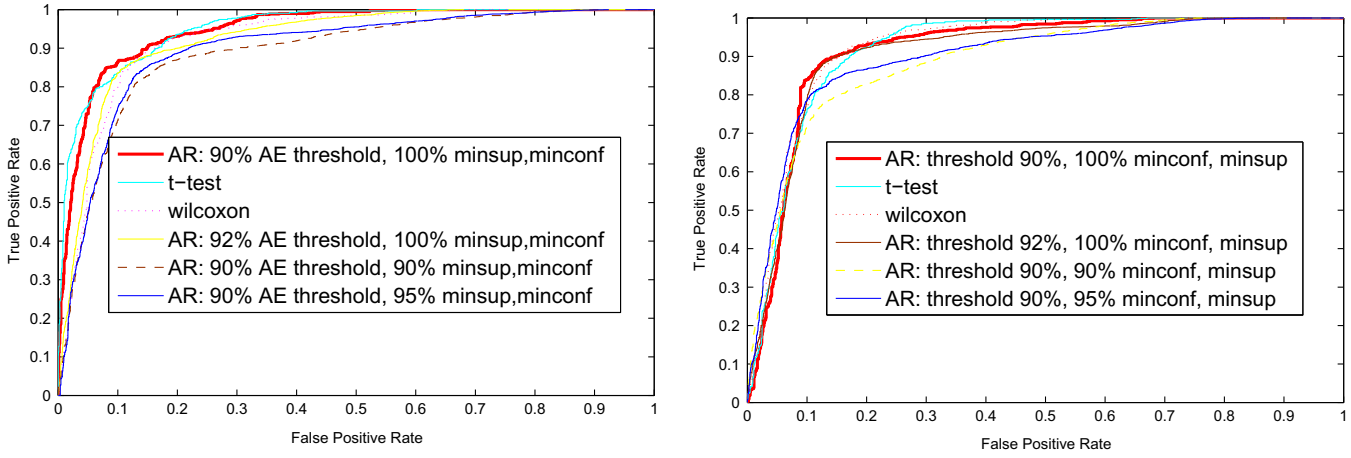

Fig. 6. ROC curve for SPECT database (a) PCA and (b) PLS. Comparison to other recently reported methods.

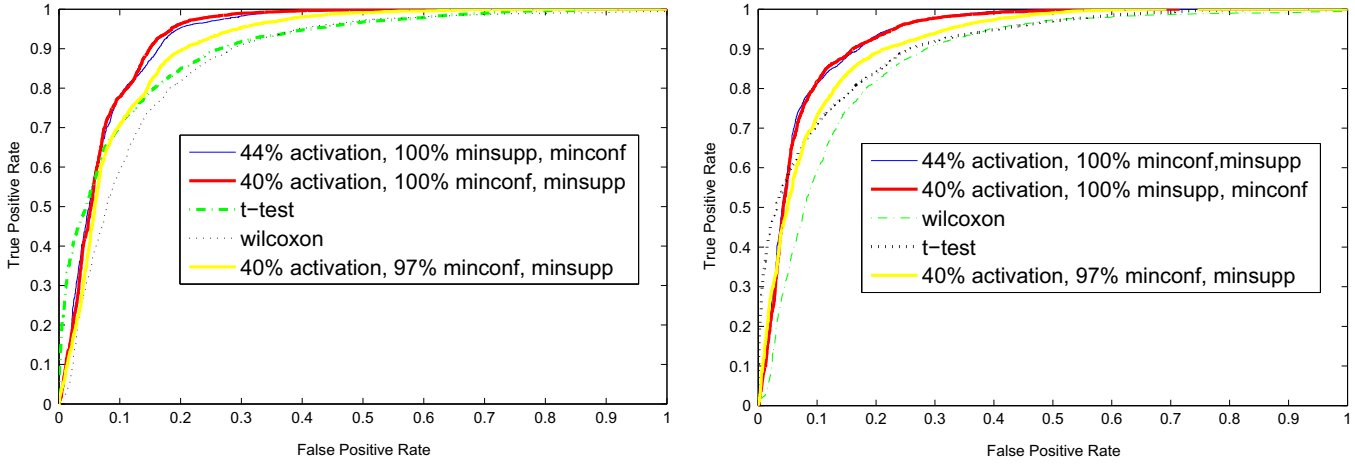

Fig. 7. ROC curve for PET database (a) PCA and (b) PLS. Comparison to other recently reported methods. The AUC obtained for each AR-FS-based ROC (PCA/PLS) is: 100% minconf and minsupp with 40% activation (0.93441/0.9391), 44% activation (0.9107/0.9181) and 97% minsupp and minconf with 40% activation (0.9107/0.9181). AUC (PCA/PLS) for other reported methods is: wilcoxon (0.8662/0.8662) or t test (0.8968/0.9027).

Fig. 5. The Acc, Sen and Spe are estimated by means of *k*-fold cross-validation with a rbf SVM classifier as it performs better for PET modality. Fig. 5a and b considered an increasing number of PCA or PLS features respectively yielding the maximum rates Acc, Sen and Spe of 90%, 85.33%, 94.67% at 4 PCA features or 90%, 89.33%, 90.67% at 10 PLS features respectively that outperfom other reported techniques as t-student, Mann–Whitney-Wilcoxon test and Relative Entropy-based.

6.3. ROC analysis

The receiver operating characteristic (ROC) curves have shown to be very effective for the evaluation of CAD systems. These plots show the trade-off between the specificity and sensitivity of the CAD system as the detection threshold varies.

Figs. 6 and 7 show the ROC curves of the proposed AR-FS-based CAD system for respectively the SPECT and PET modalities considering the PCA (Figs. 6a and 7a) and PLS (Figs. 6b and 7b) FE techniques.

The proposed AR FS-based method outperforms the baseline techniques as well since its ROC is shifted up to the left in the ROC space. This improvement of the proposed method is also demonstrated by the area under curve (AUC), which is detailed in Table 1 for SPECT yielding values of 0.9549 and 0.9259 for PCA and PLS modalities, respectively at 100% minconf and minsupp and 90% AE threshold. As PET database is concerned, the AUC shows a trend (detailed in caption of Fig. 7) similar to SPECT, that is, it yields the highest results of 0.9344 and 0.9391 for PCA and PLS modalities at 100% minconf and minsupp and 40% AE threshold.

7. Conclusion

AR-based FS technique was proposed as an effective way to solve the sample size problem of a functional database (SPECT or PET) for the early diagnosis of AD in order to design a CAD system. Firstly, ARs were obtained from activated blocks of controls at different minsupp and minconf values. When the maxima values are tunned, that is 100%, the most important voxels are selected and the rest discarded. Moreover, this CAD system was defined as an offline method, since the rules were obtained by means of the Apriori algorithm from a set of control subjects and every test subject was computed by means of the same set of rules. The voxels selected are formulated in terms of the activation of antecedents and consequents of ARs merged without repetition. The FS dimension is in this case reduced from 2445 to 250 (for SPECT) and from 3776 to 230 (for PET) features. In the following, a subsequent reduction by means of PCA and PLS techniques is necessary to set up the main feature vectors used as input for a kernel SVM classifier (linear for SPECT and rbf for PET modality). Several experiments were conducted in order to test the reliability of the proposed AR FS-based method in terms of Acc, Spe, Sen converging to an optimum operating point of 91.75%, 95.12%, 89.29% respectively for 12 PLS features (SPECT) and 90%, 90.67%, 89.33% (PET), outperforming other recently reported techniques.

Acknowledgments

This work was partly supported by the MICINN of Spain under the TEC2008-02113 project and the Consejeria de Innovacion, Ciencia y Empresa (Junta de Andalucia, Spain) under the Excellence Projects P07-TIC-02566, P09-TIC- 4530 and P11-TIC-7103. We are grateful to M. Gómez-Río and coworkers from the "Virgen de las Nieves" hospital in Granada (Spain) for providing and labeling the SPECT images used in this work. The PET 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., 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 (http://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. The authors are grateful to the reviewers for their helpful advice and valuable suggestions to improve the quality of the work provided in this document.

References

Agrawal, R. & Srikant, R. (1994). Fast algorithms for mining association rules. In *International conference on VLDB, Santiago de Chile, Chile* (pp. 487–499).

Ashburner, J., & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping, 7*(4), 254–266.

Chaves, R., Górriz, J. M., Ramírez, J., Illán, I. A., Salas-González, D., & Gómez-Río, M. (2011). Efficient mining of association rules for the early diagnosis of Alzheimer's disease. *Physics in Medicine and Biology, 56*(18), 6047–6063.

Chaves, R., Ramírez, J., Górriz, J. M., López, M., Salas-Gonzalez, D., Álvarez, I., et al. (2009). SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting. *Neuroscience Letters, 461*(3), 293–297.

Chen, L.-F., Liao, H.-Y. M., Ko, M.-T., Lin, J.-C., & Yu, G.-J. (2000). A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition, 33*, 1713–1726.

Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., & Lin, C. P.Alzheimer's Disease Neuroimaging Initiative. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImage, 60*, 59–70.

Costafreda, S. G., Chu, C., Ashburner, J., & Fu, C. H. (2009). Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS One, 4*(7), e6353.

Dasseni, E., Verykios, V. S., Elmagarmid, A. K., & Bertino, E. (2001). Hiding association rules by using confidence and support. *Lecture Notes in Computer Science, 2137*, 369–383.

Di Bucchianico, A. (1999). Combinatorics, computer algebra and the Wilcoxon–Mann–Whitney test. *Journal of Statistical Planning and Inference, 79*(2), 349–364.

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS ADRDA criteria. *The Lancet Neurology, 6*(8), 734–746.

Duchesnay, E., Cachia, A., Boddaert, N., Chabane, N., Mangin, J. F., Martinot, J. L., et al. (2011). Feature selection and classification of imbalanced datasets Application to PET images of children with autistic spectrum disorders. *NeuroImage, 57*(3), 1003–1014.

Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (2007). *Statistical parametric mapping: the analysis of functional brain images*. Academic Press.

Geng, X., Liu, T.-Y., Qin, T., & Li, H. (2007). Feature selection for ranking. *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 407–414).

Gidskehaug, L., Anderssen, E., & Alsberg, B. K. (2008). Cross model validation and optimisation of bilinear regression models. *Chemometrics and Intelligent Laboratory Systems* (93), 1–10.

He, R., Xiong, N., Yang, L. T., & Park, J. H. (2011). Using multi-modal semantic association rules to fuse keywords and visual features automatically for web image retrieval. *Informative Fusion, 12*, 223–230.

Illán, I. A., Górriz, J. M., López, M. M., Ramírez, J., Gonzalez, D. S., Segovia, F., et al. (2011). Computer aided diagnosis of Alzheimer's disease using component based SVM. *Applied Soft Computing, 11*, 2376–2382.

Ito, K. (2006). PET/SPECT for dementia early diagnosis of Alzheimer's disease. *International congress series* (Vol. 1290, pp. 123–127).

Jobst, K. A., Barnetson, L. P., & Shepstone, B. J. (1998). 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), 271–302.

Karabatak, M., & Cevdet Ince, M. (2009). A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications, 36*(10), 12500–12505.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273–324.

Liu, H., Dougherty, E. R., Dy, J. G., Torkkola, K., Tuv, E., Peng, H., et al. (2005). Evolving feature selection Intelligent Systems. *IEEE, 20*(6), 64–76.

López, M., Ramírez, J., Górriz, J. M., Álvarez, I., Salas-Gonzalez, D., Segovia, F., et al.Alzheimer's Disease Neuroimaging Initiative. (2011). Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. *Neurocomputing, 74*(8), 1260–1271.

López, M., Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., et al. (2009). Automatic tool for the Alzheimer's disease diagnosis using PCA and Bayesian classification rules. *IET Electronics Letters, 45*(8), 389–391.

Maldonado, S., Weber, R., & Basak, J. (2011). Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences, 181*, 115128.

Martínez, F. J., Salas-González, D., Górriz, J. M., Ramírez, J., Puntonet, C. G., & Gómez-Río, M. (2011). Analysis of spect brain images using Wilcoxon and relative entropy criteria and quadratic multivariate classifiers for the diagnosis of Alzheimer's disease, IWINAC 2011, Part II. *LNCS* (Vol. 6687, pp. 41–48).

Martínez-Murcia, F. J., Górriz, J. M., Ramírez, J., Puntonet, C. G., & Salas-Gonzalez, D.for the Alzheimer's Disease Neuroimaging Initiative. (2012). Computer aided diagnosis tool for Alzheimer's disease based on Mann–Whitney–Wilcoxon U-test. *Expert Systems with Applications*. http://dx.doi.org/10.1016/j.eswa.2012.02.153.

Muellera, S. G., Weinera, M. W., Thald, L. J., Petersene, R. C., Jackf, C. R., Jagustg, W., Trojanowskih, J. Q., Togai, A. W., & Beckettj, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's disease neuroimaging initiative (ADNI). *Alzheimer's and Dementia, 1*, 55–66.

Pan, H., Li, J., Wei, Z. (2005). Mining interesting association rules in medical images, ADMA. *LNAI* (Vol. 3584, pp. 598–609).

Rajendran, P., & Madheswaran, M. (2010). Hybrid medical image classification using association rule mining with decision tree algorithm. *Journal of Computing, 2*(1), 127–136.

Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., Romero, A., López, M., Illán, I. A., Gómez-Río, M. (in press). Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features. *Information Science*. http://dx.doi.org/10.1016/j.ins.2009.05.012.

Ramírez, J., Górriz, J. M., Chaves, R., Salas-Gonzalez, D., López, M., Salas-Gonzalez, D., et al. (2009). SPECT image classification using random forests. *IET Electronics Letters, 45*(12), 604–605.

Ramírez, J., Górriz, J. M., Segovia, F., Chaves, R., Salas-Gonzalez, D., López, M., et al. (2010). Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. *Neuroscience Letters, 472*, 99–103.

Ribeiro, M. X., Marques, J., Traina, A. J. M., Traina, C. (2006). Statistical association rules and relevance feedback: powerful allies to improve the retrieval of medical images. In *Proceedings of the 19th IEEE symposium on computer-based medical systems (CBMS'06)* (pp. 887–892).

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507–2517.

Salas-Gonzalez, D., Górriz, J. M., Ramírez, J., Lassl, A., & Puntonet, C. G. (2008). Improved Gauss–Newton optimization methods in affine registration of SPECT brain images. *IET Electronics Letters, 44*(22), 1291–1292.

Sigut, J., Pineiro, J., Gonzalez, E., & Torres, J. (2007). An expert system for supervised classifier design: Application to Alzheimer diagnosis. *Expert Systems with Applications, 32*, 927–938.

Tan, P., Kumar, V., Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. In *8th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 32–41).

Turkeltaub, P. E., Eden, K. J., & Zeffiro, T. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage, 16*, 765–780.

Zaiane, O. R., Antonie, M. L., & Coman, A. (2002). Mammography classification by an association rule-based classifier. In *International workshop on multimedia data mining (ACM SIGKDD)* (pp. 62–69).

Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications, 38*, 5197–5204.