

# NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease

P. Padilla\*, M. López, J. M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, and  
The Alzheimer's Disease Neuroimaging Initiative

**Abstract**—This paper presents a novel computer-aided diagnosis (CAD) technique for the early diagnosis of the Alzheimer's disease (AD) based on nonnegative matrix factorization (NMF) and support vector machines (SVM) with bounds of confidence. The CAD tool is designed for the study and classification of functional brain images. For this purpose, two different brain image databases are selected: a single photon emission computed tomography (SPECT) database and positron emission tomography (PET) images, both of them containing data for both Alzheimer's disease (AD) patients and healthy controls as a reference. These databases are analyzed by applying the Fisher discriminant ratio (FDR) and nonnegative matrix factorization (NMF) for feature selection and extraction of the most relevant features. The resulting NMF-transformed sets of data, which contain a reduced number of features, are classified by means of a SVM-based classifier with bounds of confidence for decision. The proposed NMF-SVM method yields up to 91% classification accuracy with high sensitivity and specificity rates (upper than 90%). This NMF-SVM CAD tool becomes an accurate method for SPECT and PET AD image classification.

**Index Terms**—Alzheimer's disease, nonnegative matrix factorization, positron emission tomography (PET), single photon emission computed tomography (SPECT), support vector machines (SVMs).

## I. INTRODUCTION

ALZHEIMER'S disease (AD) is the most common cause of dementia in aged people and affects more than 30 million individuals worldwide. The particular evolution of AD patients and their increasing dependence on the close affective environment provokes an important social repercussion, as the cognitive functions of the patient gradually disappear and his individual essence blurs [1]. The effects of this disease are of great importance not only in terms of familiar dependence and affliction but also economic: with the growth of the older population in developed nations, the prevalence of AD is expected to triple over the next 50 years. Functional imaging modalities are often used with the aim of achieving early diagnosis, although this early diagnosis remains as a demanding task, usually based on the information provided by a careful clinical examination carried out by experts.

Emission computed tomography images have been widely employed in biomedical research and clinical medicine during the last decade. These emission-based functional images reproduce a map of physiological functions along with anatomical structure images, providing information about physiological phenomena and their location in the body. In this work, two different modalities are used for brain image acquisition: positron emission tomography (PET) and single photon emission computed tomography (SPECT). Both techniques are noninvasive, nuclear medicine imaging techniques which produce a three-dimensional image of functional processes in the body, such as blood perfusion or glucose metabolism, by means of emitting radionuclides (tracers) [2], [3]. In both techniques, PET and SPECT, all these detected emissions are processed and a three-dimensional image of the region under study is obtained, in this case the brain, by means of subsequent computer analysis and back-projection Fourier algorithms [2].

### A. Computer Aided Diagnosis (CAD) Techniques

For the past several decades, researchers in the medical imaging field have focused on bringing new imaging modalities to clinicians while improving the performance of existing systems [4], [5]. Nowadays, signal processing engineers are

Manuscript received June 22, 2011; accepted August 29, 2011. Date of publication September 12, 2011; date of current version February 03, 2012. This work was supported in part by the Spanish Government under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030-C02-01) projects and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Projects TIC-02566 and TIC-4530. The SPECT database was provided by "Virgen de las Nieves" hospital in Granada (Spain). 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, Inogenetics, 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 nonprofit 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 work was also supported by the National Institutes of Health (NIH) under Grant P30 AG010129, Grant K01 AG030514, and the Dana Foundation. *Asterisk indicates corresponding author.*

\*P. Padilla is with the Department of Signal Theory, Telematics and Communications, University of Granada, 18071 Granada, Spain, and also with the Radiation Group, Technical University of Madrid, 28040 Madrid, Spain (e-mail: [ppadilla00@gr.ssr.upm.es](mailto:ppadilla00@gr.ssr.upm.es)).

M. López, J. M. Górriz, J. Ramírez, D. Salas-González, and I. Álvarez are with the Department of Signal Theory, Telematics and Communications, SIPBA Group, University of Granada, 18071 Granada, Spain.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2011.2167628

beginning to take the next step by introducing software improvements, enabling computers to help clinicians to make sense of such an amount of noninvasive medical information. Computer aided diagnosis (CAD) is a general term used for a variety of techniques applied to whatever kind of medical data, such as to medical images, to assist physicians in their diagnosis work. CAD systems help physicians by either identifying patterns that might have been overlooked or by providing a road map of suspicious areas, making their efforts more efficient.

Several approaches for designing CAD systems for the Alzheimer's disease diagnosis can be found in the literature [6]–[9]. The first approach is the statistical parametric mapping (SPM) [10] tool, widely used in neuroscience, and its numerous variants. It was not developed specifically to study a single image, but for comparing groups of images. SPM is designed as a univariate approach since the classical multivariate techniques require the number of available observations (i.e., images) to be greater than the number of components (i.e., voxels) of the multivariate observation. The second approach is based on the analysis of the images, the analysis of regions of interest (ROI), feature extraction and posterior classification in different classes by means of some discriminative functions. This multivariate approach faces up the well-known small sample size problem, that is, the number of available samples is much lower than the number of features used in the training step. In this case, dimensionality reduction and feature selection is an issue of great concern. As referred in [11] and [12], statistical learning classification methods and CAD tools have not been explored in depth for dementia, quite possibly due to the fact that images represent large amounts of data and most imaging studies have relatively few subjects (typically <100) [11], [13], [14].

A supervised learning-based CAD system applied to functional imaging consists of several important stages:

- 1) functional image acquisition and normalization;
- 2) feature selection and extraction;
- 3) classification, with a train and test strategy.

The design and proper validation of a CAD tool for Alzheimer's disease diagnosis, along with the adequate description of its forming techniques for feature selection and extraction and for posterior classification, is the main issue in this work.

Sections II and III provide the main techniques applied in the CAD tool provided in this work. The first one is devoted to the reduction and rearrangement of the data by means of nonnegative matrix factorization (NMF), meanwhile the second is devoted to the proper definition of a support vector machine (SVM) based classifier with bounds of confidence. Section IV provides the Experimental setup of the CAD tool, along with the input data description. Section V provides the evaluation results of the CAD tool. Finally, conclusions are drawn in Section VI.

## II. FEATURE SELECTION AND REDUCTION

Each voxel of a brain functional 3-D image contains information of the corresponding brain point. However, not all the voxels have the same level of relevance in terms of discrimination between groups of subjects. In this case, two groups of subjects are defined: Alzheimer's disease patients, labeled as AD, and subjects not affected by this disease, labeled as NOR. Thus,

an initial feature selection based on discrimination capability is typically selected [15]–[17], obtaining a vector of discriminant voxels for each participant. In addition, the selected discriminant voxel vectors can be projected onto a different subspace. This subspace is chosen so that only a few variables represent the most discriminant features of each patient images in each database. These steps are conveniently described below.

### A. Intensity Normalization

Previous to any kind of feature selection, the data sets have to be normalized in intensity in order to be able to compare images according to their voxel normalized intensity levels. Regarding the intensity normalization, the normalization to the maximum intensity level may introduce problems in some images that can have peak intensity values due to noise. Thus, these images are badly normalized as the normalization is based on wrong noisy voxels. In this work, in order to avoid these possible normalization errors, it is applied intensity normalization based on the mean value of a group of voxels with the highest intensity values. According to [18], the mean value of the 0.1% voxels with the highest intensity levels is selected for the intensity normalization.

### B. Fisher Discriminant Ratio for Feature Selection

The Fisher discriminant ratio (FDR) criterion is characterized by its separation ability as shown in [19]. For the two-class case, it may be defined as follows:

$$\text{FDR} = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (1)$$

where  $\mu_i$  and  $\sigma_i^2$  denote the  $i$ th class mean value and variance for each input variable, respectively. For a given variable, the ratio value grows as the difference of the mean values of each two classes increases or the cumulative scattering in each class decreases, thus being useful to reveal discriminant variables. In the case of the functional images, the voxels that satisfy a particular FDR threshold level are selected as the most discriminative variables [11], [17], [20]. In addition, the selection of voxels that pass the FDR threshold lets reduce the dimensionality of the problem, which means a lower number of variables in each observation.

### C. Nonnegative Matrix Factorization for Feature Reduction

Nonnegative matrix factorization (NMF) is a technique for finding parts-based, linear representations of nonnegative data [21], [22], being a useful decomposition tool for multivariate data. This technique is especially suitable for nonnegative data sets such as functional images in general, and for the PET and SPECT brain images of this work in particular, where all the variables consist of positive values. Given a nonnegative data matrix  $\mathbf{A}$ , NMF finds an approximate factorization  $\mathbf{A} \approx \mathbf{WH}$  into nonnegative matrices  $\mathbf{W}$  and  $\mathbf{H}$ . The nonnegativity condition forces the representation to be purely additive, in contrast to other existing representations such as principal component analysis (PCA) [23], kernel PCA [24], etc. PCA or NMF can all be seen as matrix factorization, with different choices of objective function and/or constraints [25].

NMF has been widely applied in the field of image processing. In the literature, it can be found a variety of works based on NMF, for image processing in general (i.e., face recognition [26], [27]) and in medical image processing and analysis in particular: brain image analysis [28], dynamic myocardial image analysis [29], etc. Many works are available related to NMF image transformation for image analysis [27], [30], NMF algorithms [21], [31], kernel techniques applied to NMF [31], [32], etc.

Formally, nonnegative matrix factorization is a linear, non-negative approximate data representation where the original database  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M]$  ( $N$  by  $M$  elements), which consists of  $M$  measurements (profiles) of  $N$  nonnegative scalar variables, is approximated by a nonnegative matrix product, as given in

$$\mathbf{A} \approx \mathbf{W}\mathbf{H} \quad (2)$$

where the matrix  $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_K]$  has dimension  $N \times K$ , and the matrix  $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_M]$  has dimension  $K \times M$ . Thus, each element of matrix  $\mathbf{A}$  is decomposed, as shown in

$$\mathbf{A}_{nm} = \sum_{k=1}^K \mathbf{W}_{nk} \mathbf{H}_{km}. \quad (3)$$

An appropriate decision on the value of  $K$  is critical in practice, but the choice of  $K$  is very often problem dependent. In most cases, however,  $K$  is usually chosen such that  $K \ll \min(M, N)$  in which case  $\mathbf{W}\mathbf{H}$  can be seen as a compressed form of the data in  $\mathbf{A}$  [33]. This property yields a reduced-variable matrix  $\mathbf{H}$  that represents  $\mathbf{A}$  in terms of the NMF basis  $\mathbf{W}$ . After NMF factorization, the data contained in  $\mathbf{H}$  ( $K$  by  $M$  elements) can be considered a transformed database with lower rank ( $K$ ), than the original database  $\mathbf{A}$ . Thus, a few variables are representing the data of each profile in the new representation. The relative error (%) of the factorization can be computed by means of the comparison of matrix  $\mathbf{A}$  and the approximation  $\mathbf{W}\mathbf{H}$ . The minimum number of vectors ( $K$ ) in the NMF basis is selected so that a predefined level of relative error is not exceeded.

1) *Factorization Rule*: Given the data matrix  $\mathbf{A}$ , the optimal choice of matrices  $\mathbf{W}$  and  $\mathbf{H}$  are defined to be those nonnegative matrices that minimize the reconstruction error between  $\mathbf{A}$  and  $\mathbf{W}\mathbf{H}$ . A variety of error functions ( $Err$ ) have been proposed [21], [22], some of the most useful are given below, in (4) and (5), and applied in this work [21]

$$\begin{aligned} Err_1 &= \frac{1}{NM} \|\mathbf{A} - \mathbf{W}\mathbf{H}\|^2 \\ &= \frac{1}{NM} \sum_{nm} (\mathbf{A}_{nm} - (\mathbf{W}\mathbf{H})_{nm})^2 \end{aligned} \quad (4)$$

$$\begin{aligned} Err_2 &= D(\mathbf{A} \parallel \mathbf{W}\mathbf{H}) \\ &= \sum_{nm} (\mathbf{A}_{nm} \log \frac{\mathbf{A}_{nm}}{(\mathbf{W}\mathbf{H})_{nm}} - \mathbf{A}_{nm} + (\mathbf{W}\mathbf{H})_{nm}) \end{aligned} \quad (5)$$

where (4) is known as the Frobenius norm (reduction of the Euclidean distance), and (5) as the Kullback–Leibler divergence,

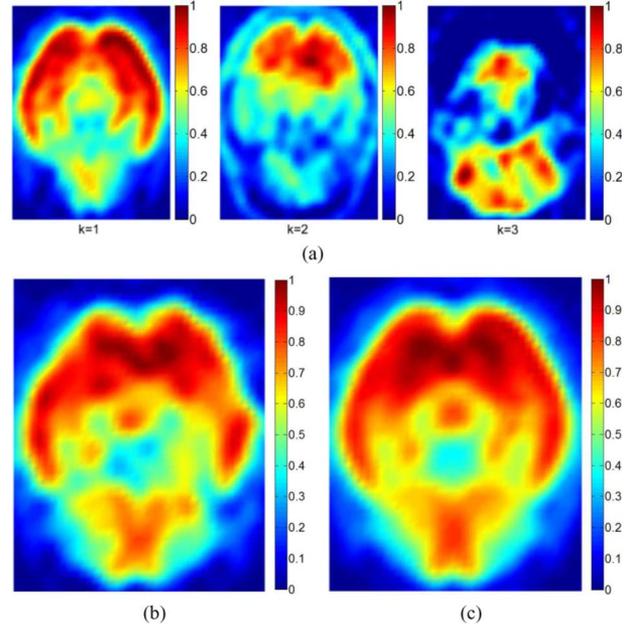


Fig. 1. NMF projection for the same transaxial slice of all the SPECT database subjects. (a) NMF eigenvectors for  $k = 1, 2, 3$ . (b) One example subject slice. (c) Its NMF reconstruction. Note the special relevance of some brain regions in the eigenvectors.

among others. The NMF process is, thus, translated into an optimization problem, subject to minimization of  $Err$ , according to the one chosen.

Some NMF algorithms are proposed in [33]. There are different approaches for these algorithms [21]: with multiplicative update rule, with additive update rule or alternating least squares algorithms (ALS). Due to their fast convergence and lower iteration requirements, the last one is selected for NMF in this work.

Although there are a variety of nonlinear techniques for feature reduction, in this work only the linear case is considered and NMF is selected, due to the simplicity of the proposed factorization and the preservation of a linear relation between the original space of features and the new one.

For the sake of clarity, Fig. 1 provides the first three vectors of the NMF basis  $\mathbf{W}$ , in the form of 2-D images, derived from one of the data sets used in this work (SPECT database, see Section IV below), along with one particular transaxial slice and the one obtained from the NMF projection. The transaxial slices provided in Fig. 1 are oriented from posterior (top) to anterior (bottom). In the rest of the document the same orientation criterion is followed. For all the data sets used in this work, a relative error lower than 5% is guaranteed in all the NMF transformations.

### III. SUPPORT VECTOR MACHINE (SVM) BASED CLASSIFIER WITH BOUNDS OF CONFIDENCE

#### A. SVM Background

Support vector machine (SVM) is a widely used technique for pattern recognition and classification in a variety of applications for its ability for detecting patterns in experimental databases [34]–[36]. SVM has become an essential machine-learning method for the detection and classification of particular patterns

in medical images. In the literature, it can be found several fields in which SVM are applied: cancer, tumor, or nodule detection [37]–[42], vascular analysis [43]–[46], dementia detection [11], [14], [16], [47], [48], etc. Regarding image modalities, SVM has been applied to a variety of image types: magnetic resonance images (MRI) [42], [49], [50], SPECT or PET [14], [48], ultrasound images [37], [39], [51], etc.

SVM techniques consist of two separate steps: first of all a given set of binary labeled training data is used for training; then new unlabeled data can be classified according to the learned behavior. SVM separates a given set of binary labeled training data by means of a hyperplane that is maximally distant from the two possible classes (in our particular case, NOR an AD classes). The objective is to build a function  $F$  with the training data, as expressed in (6), able to properly classify new unclassified data

$$F : R^N \rightarrow \{\pm 1\}. \quad (6)$$

The training data are formed by  $p$  different profiles (set of images of each subject), each one containing  $N$  variables, together with their proper label (NOR or AD). Thus, the training database can be expressed as

$$S_p = [(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), y]_p = [\mathbf{x}, y]_p \quad (7)$$

where  $\mathbf{x}_n$  are the variables of the profile  $p$  and  $y$  the corresponding label.

Linear discriminant functions define decision hyperplanes in the  $N$ -dimensional feature space

$$g(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + \omega_0 \quad (8)$$

where  $\boldsymbol{\omega}$  is the weight vector that is orthogonal to the decision hyperplane and  $\omega_0$  is the threshold. The optimization task consists of finding the unknown parameters  $\omega_n$ , and  $\omega_0$  that define the decision hyperplane. The hyperplane is not unique and the selection process focuses on maximizing the generalization performance of the classifier, that is, the ability of the classifier, designed using the training set, to operate satisfactorily with new data. The vectors that define the separation hyperplane are called support vectors (SVs). Among the different design criteria, the maximal margin hyperplane is usually selected since it leaves the maximum margin of separation between the two classes [36].

When no linear separation of the training data is possible, SVM can work in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a nonlinear decision boundary [52]. However, in this work, only the linear case is treated.

### B. Bounds of Confidence

As referred, SVM is based in the definition of a decision hyperplane in the  $N$ -dimensional space. The SVs are a subset of the training dataset, chosen so that the decision hyperplane that is defined by means of them is the best one in terms of separability between two classes. The classification of one sample is directly related to the subspace in which it is located, regarding the decision hyperplane. The particular distance-like decisions

in SVM make this machine learning approach especially suitable for the selection of a security region around the decision hyperplane in which decisions may not be adopted. This security region is defined by means of the selection the bounds of confidence for the decision.

When the conventional SVM learning approach is performed and the resulting SVM classifier is trained with a set of training observations, the separation hyperplane of these data is defined. Ideally all the training data should be properly classified when they are re-entered in the SVM for test and classification [36]. However this is not typically the situation as the variables of the different observations spanned in a particular space of interest (i.e., NMF) may not be separable in this space. What is more, the hyperplane defined by all the SVs may not be able to properly define the class of one particular support vector. That means that although the SVs help to define the best hyperplane in terms of separability, some of them may not be in the right subspace defined by the hyperplane. In most of the cases, the two classes are not completely separable and there is some kind of overlapping between them in the space of interest. In these cases, as referred in [36] and [53], although the defined hyperplane is the best one in terms of separation, some SVs may be wrongly located regarding the hyperplane, thus being wrongly labeled if re-entered in the SVM classifier for testing. As an example, Fig. 9, provided in Section V-B, shows clear examples of the overlapping among classes and SVs located in the wrong subspace, in three-dimensional NMF subspaces.

In this particular scenario, it is useful to compute the classification of all the SVs and to derive the error probability for the proper classification of the SVs of class +1 (SVPs) and SVs of class -1 (SVNs). If we denote SVP\_N as a support vector of class +1 wrongly labeled as -1 and SVN\_P on the contrary, the error probabilities for SVPs ( $P_{\text{errPos}}$ ) and SVNs ( $P_{\text{errNeg}}$ ) can be defined as follows:

$$P_{\text{errPos}} = \frac{N_{\text{SVP\_N}}}{N_{\text{SVPs}}} \quad P_{\text{errNeg}} = \frac{N_{\text{SVN\_P}}}{N_{\text{SVNs}}} \quad (9)$$

where  $N_{\text{SVN\_P}}$ ,  $N_{\text{SVP\_N}}$ ,  $N_{\text{SVPs}}$ , and  $N_{\text{SVNs}}$  denote the number of vectors of each group. Although other error probabilities may be defined, these ones will be useful for our purpose. This error probability can be reduced if SVs which are nearer to the hyperplane (hopefully wrong ones) are not considered. The consequence of not considering these SVs in the SVM with bounds of confidence is that all the observations located between the discarded SV which is furthest to the hyperplane and the hyperplane itself are not classified because they are in the *security zone*. Fig. 2 reveals clearly this fact. Notice that the hyperplane is still the same which was defined with all the SVs.

Consider a train data set  $\{x_{\text{train}}, y_{\text{train}}\}$  and the corresponding SVM classifier trained with this data set, with  $\{\text{SVP}_1, \dots, \text{SVP}_{n1}\}$  as the SVs of the positive region (+1) and  $\{\text{SVN}_1, \dots, \text{SVN}_{n2}\}$  as the SVs of the negative region (-1), and some constraints in terms of maximum accepted value for  $P_{\text{errPos}}$  and  $P_{\text{errNeg}}$ . Then, if the SVs are entered in the classifier, the SVP\_N and SVN\_P can be computed, ordered from the closest to the largest distance to the decision hyperplane:  $\{\text{SVP\_N}_1, \dots, \text{SVP\_N}_{k1}\}$  and  $\{\text{SVN\_P}_1, \dots, \text{SVN\_P}_{k2}\}$ .

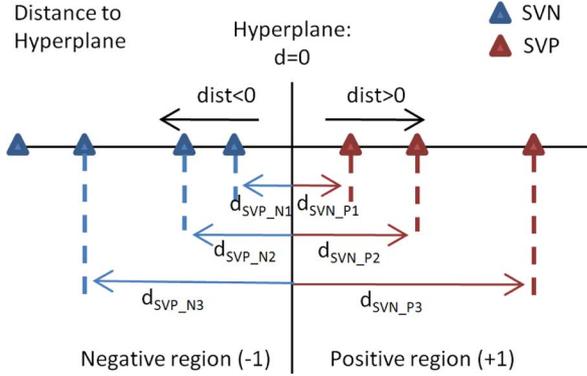


Fig. 2. Example of SVP<sub>N</sub> and SVN<sub>P</sub> distribution in terms of distance to hyperplane.

Fig. 2 shows an example of SVP<sub>N</sub> and SVN<sub>P</sub> distribution in terms of distance to the hyperplane. Notice that the classifier places them in the wrong subspace. According to these wrongly classified SVs, it can be defined the security region where decisions are risky and no classification is performed [53]. In this case, new values of  $P_{\text{errNeg}}$  and  $P_{\text{errPos}}$  are obtained. For instance, imagine that SVP<sub>N</sub><sub>1</sub> is extracted from the classifier and no decision is made for observations whose classifier output is between  $d_{\text{SVP}_N1}$  and the separation plane: in this case, the  $P_{\text{errPos}}$  is reduced because at least one SVP<sub>N</sub> (SVP<sub>N</sub><sub>1</sub>) is not considered in (9). For new unseen data the classifier will decrease the error rate, with the drawback of not classifying some observations whose SVM output is inside the security region and the decision is considered as too risky. In general, the new SVM classifier with bounds of confidence can be defined as follows.

- 1) Select the desired value of  $P_{\text{errNeg}}$  and  $P_{\text{errPos}}$ :  $P_{\text{errNegDesired}}$ , and  $P_{\text{errPosDesired}}$ , respectively.
- 2) Take all the SVs and obtain  $\{\text{SVP}_N1, \dots, \text{SVP}_Nk1\}$  and  $\{\text{SVN}_P1, \dots, \text{SVN}_Pk2\}$ .
- 3) Define  $i_a$  for indexing the SVP<sub>N</sub> and  $i_b$  for indexing SVN<sub>P</sub>. Set them initially to 1.
- 4) Compute the values of  $P_{\text{errNeg}}$  and  $P_{\text{errPos}}$  with the desired ones in step 1.
  - a) If  $P_{\text{errPos}} \geq P_{\text{errPosDesired}}$ , extract the SVP<sub>N</sub>( $i_a$ ) :  $[d_{\text{SVP}_N}(i_a), 0]$  is the security region of the negative region (-1). Increase  $i_a$  and compute again step 4a.
  - b) If  $P_{\text{errNeg}} \geq P_{\text{errNegDesired}}$ , extract SVN<sub>P</sub>( $i_b$ ) :  $[0, d_{\text{SVN}_P}(i_b)]$  is the security region of positive region (+1). Increase  $i_b$  and compute again step 4b.
  - c) The SVM-based classifier is ready, with  $\{+1, -1\}$  for properly classified data and  $\{-0.5, +0.5\}$  for risky decisions.

The use of bounds of confidence and the so-called “security zone” provokes that the new classifier turns the old binary output of the classifier (which was  $\{+1, -1\}$ ) into in a new four-cases output  $\{+1, +0.5, -0.5, -1\}$ , where  $\{+1, -1\}$  have the same meaning than in the old classifier and  $\{+0.5, -0.5\}$  are related to observations remaining in the +1 or -1 side of the hyperplane, respectively,

but too risky to be defined as class +1 or -1. It is clear that, as the values of  $P_{\text{errPos}}$  and  $P_{\text{errNeg}}$  are defined when considering the SVs as the evaluation set, these values are guaranteed *a priori* when considering an evaluation set of data different from the one in the training step. However, although the values of  $P_{\text{errPos}}$  and  $P_{\text{errNeg}}$  are not guaranteed, the existence of the security zone increases the success rate of the classifier, at the expense of having some unclassified outputs.

### C. Modified SVM-Based Classifier With Bounds of Confidence

Consider a SVM with the four output values:  $\{+1, +0.5, -0.5, -1\}$ , where the central ones make reference to risky +1 decision and risky -1 decision, respectively. Although these central outputs are not valid for decision, they may define a tendency. Consider a new classifier that combines the output of the SVM classifier for different NMF projections, that is to say, for different  $K_j$  ( $j = 1, \dots, T$ ) values, as defined in next equation

$$F_{\text{NewClassifier}}(\mathbf{x}) = \frac{1}{T} \sum_{j=1}^T F_{K_j}(\mathbf{x}). \quad (10)$$

In this case, if  $T$  is an odd number, the output of the new classifier function is either positive or negative. Thus, a new binary decision is set  $\{+1, -1\}$ .

$$F_{\text{NewClassifier}}(\mathbf{x}) > 0 \rightarrow \text{class}(+1) \quad (11)$$

$$F_{\text{NewClassifier}}(\mathbf{x}) < 0 \rightarrow \text{class}(-1). \quad (12)$$

Other strategies based on the combination of classifiers may be considered but the one provided above is taken as the reference in this work.

## IV. CAD TOOL EXPERIMENTAL SETUP

### A. Functional Brain Image Data Sets

In order to validate the performance and outcomes of the designed NMF-SVM based CAD tool for Alzheimer’s disease detection, two different databases are used. The first one involves SPECT brain images, whereas the second one consists of PET brain images. These two databases, described below, contain spatially normalized functional brain images of different subjects. This normalization step ensures that a given voxel in one patient refers to the same brain position than the same voxel in another patient. Then, the intensities of the functional images are normalized to the maximum intensity [18]. This normalization is computed for each image individually by referring each voxel to the average value of the %0.1 highest intensity voxels, in order to allow statistical comparison among different subjects.

1) *SPECT Database*: This baseline SPECT data set from 97 participants was collected from the “Virgen de las Nieves” hospital, Granada, Spain. Each patient was injected with a gammaemitting technetium-99m labeled ethyl cysteinate dimer (99mTc-ECD) radiopharmaceutical and the SPECT scan was acquired by means of a three-head gamma camera Picker Prism

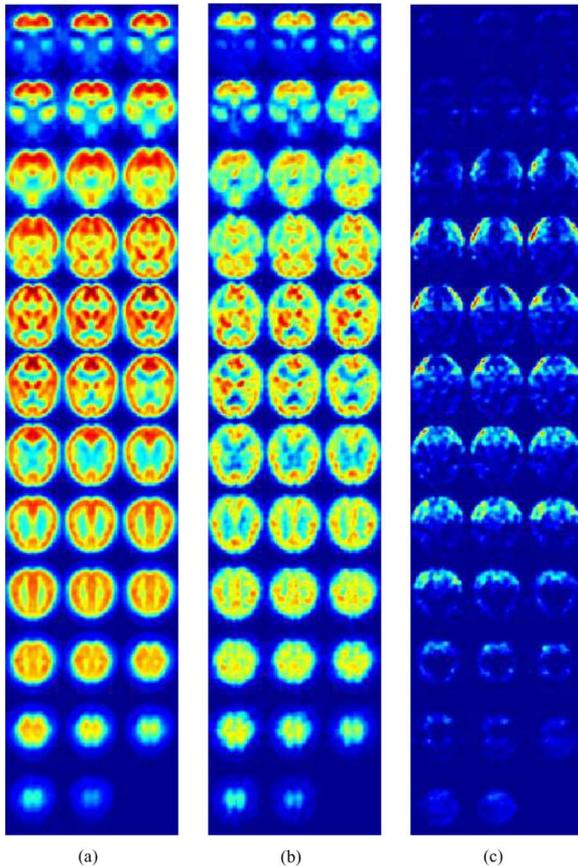


Fig. 3. SPECT transaxial brain slices, oriented from posterior (top) to anterior (bottom) in each slice, and from ventral (top left) to dorsal (bottom right) in the complete set. (a) Man mask for NOR subjects. (b) Example AD patient. (c) Between-subjects variability between AD and NOR groups.

3000. Brain perfusion images were reconstructed from projection data by filtered backprojection (FBP) in combination with a Butterworth noise filter [54]. The SPECT images were labeled by experts of the “Virgen de las Nieves” hospital using two different labels: NOR for subjects without any symptom, and AD for Alzheimer’s patients. The complete SPECT database consists of 97 patients: 41 NOR and 56 AD. Fig. 3 shows a set of transaxial brain slices of the SPECT database, for one AD patient and one NOR subject.

2) *PET Database*: This PET data set selected for the validation of the CAD tool was obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database.<sup>1</sup> Launched in 2003 by NIA (National Institute on Aging), NIBIB (National Institute of Biomedical Imaging and Bioengineering), and FDA (Food and Drug Administration), the ADNI main purpose was focused on the measurement of the progression of Alzheimer’s disease (AD) in its initial stages. The ADNI, coordinated by M. W. Weiner (VA Medical Center and University of California, San Francisco), is the product of the effort of a variety of researchers from a wide number of academic institutions and private corporations. 800 participants from USA and Canada were recruited: approximately 200 participants without symptoms as a reference, about 400 MCI subjects, along with their temporal

<sup>1</sup><http://www.loni.ucla.edu/ADNI>

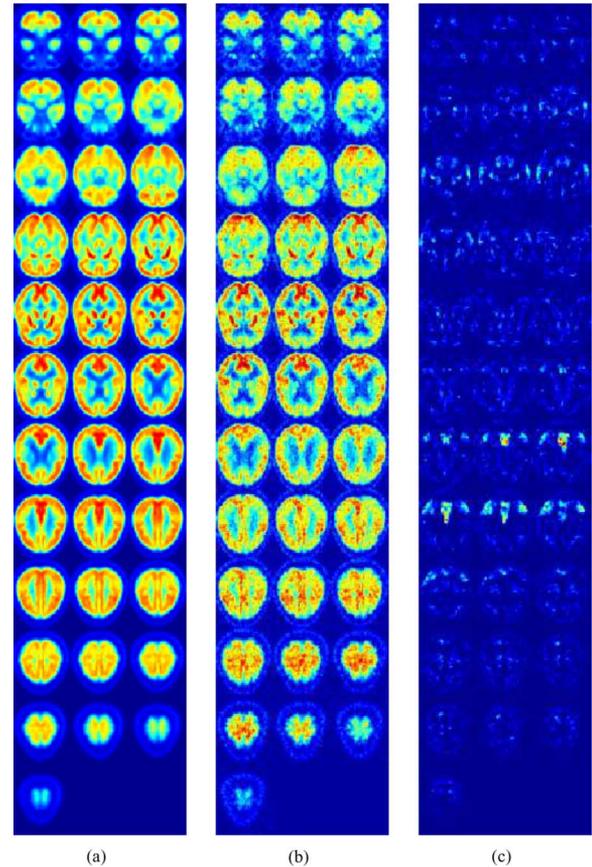


Fig. 4. PET transaxial brain slices, oriented from posterior (top) to anterior (bottom) in each slice, and from ventral (top left) to dorsal (bottom right) in the complete set. (a) Mean mask for NOR subjects. (b) Example AD patient. (c) Between-subjects variability between AD and NOR groups.

evolution over three years and 200 patients with early AD symptoms, with their progression in two years.

Among all the data available from ADNI, baseline Fludeoxyglucose (18F-FDG) PET data from 219 ADNI participants was selected. Participants in this database are classified into three different groups, according to their Mini-Mental State Exam (MMSE) or Clinical Dementia Rating (CDR):

- 114 mild cognitive impairment subjects (MCI): patients with memory loss clinically measured, but absence of significant levels of impairment in other cognitive domains.
- 53 AD patients (AD): patients meeting the NINCDS/ADRDA criteria [55] for probable AD.
- 52 normal control subjects (NOR): non-MCI, and nondemented subjects.

Fig. 4 shows a set of transaxial plane brain slices of the PEC database, for one AD patient and one NOR subject.

In Fig. 3(c) and Fig. 4(c), the regions with high discrepancies between NOR and AD subjects can be noticed. In AD patients, characteristic brain regions show decreased glucose metabolism or blood hypo-perfusion regions, 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. The slices in Figs. 3 and 4 are oriented from posterior (top) to anterior (bottom) in each one, and from ventral (top left) to dorsal (bottom right), in each set of slices.

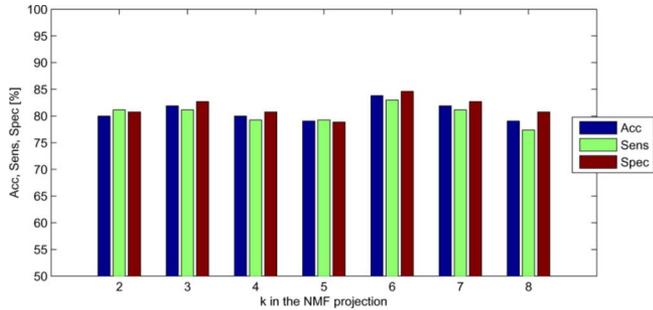


Fig. 5. Performance of the basic NMF-SVM CAD system with the PET database, for different  $K$  values in NMF.

### B. CAD Tool Evaluation

In order to evaluate the developed NMF-SVM CAD tool in all its variations, the success rate (Acc), sensitivity (Sens), and specificity (Spec) are obtained, these last two defined as in

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{Spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (13)$$

where TP is the number of true positives (AD patients correctly classified); TN is the number of true negatives (NOR patients correctly classified); FP is the number of false positives (NOR classified as AD); FN is the number of false negatives (AD classified as NOR).

These statistics are estimated by means of a leave-one-out cross-validation strategy.

- 1) Patient  $n$  (from 1 to  $N$ ) is extracted and the resulting  $n - 1$  subject data set is used for feature selection, NMF projection and for the classifier training.
- 2) data from Patient  $n$  are projected in the previously generated NMF space and then entered into the classifier for testing.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

This section provides the experimental results of the evaluation of the CAD tool developed in this work, along with its variants. First of all a NMF-SVM based CAD tool is developed, with linear SVM as classifier, according to Section III-A. The proposed method is later enhanced with the addition of bounds of confidence for the classification decision, according to Section III-B. Finally a new SVM-based classifier, according to Section III-C is evaluated. Both databases are applied as incoming data for these CAD tools, in order to extract their outcomes for validation. In the PET database, only NOR and AD groups (105 subjects) are considered for the validation of the CAD system, in order to avoid errors due to the MCI patients and the uncertainty of their actual state as AD or NOR. In the SPECT database, all the 97 patients are considered.

### A. Basic NMF-SVM CAD Tool

Figs. 5 and 6 provide the experimental results of the basic NMF-SVM system, for a variety of  $K$  values in the NMF projection. As it is seen, levels in the range of 80% to 90% are achieved, for both databases. These results are considered as a reference for the modifications of the CAD tool. The lower performance for the PET data set is justified by the particular nature

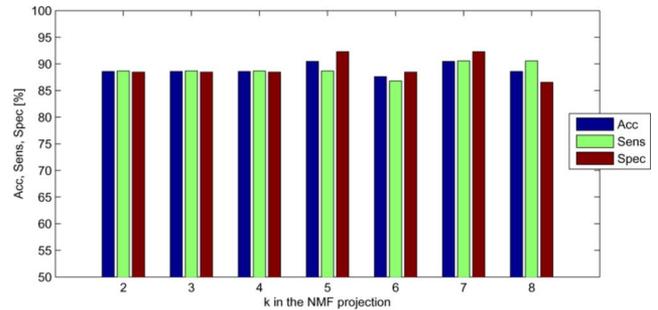


Fig. 6. Performance of the basic NMF-SVM CAD system with the SPECT database, for different  $K$  values in NMF.

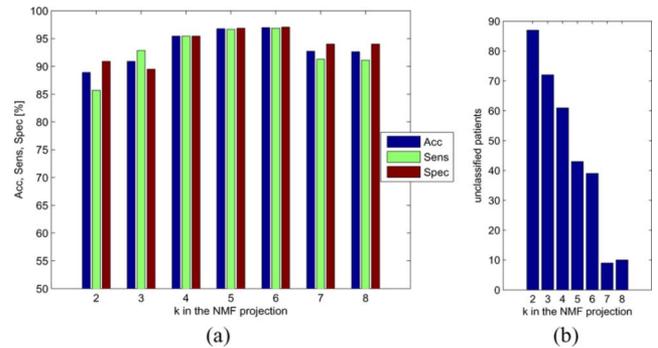


Fig. 7. (a) Performance of the basic NMF-SVM CAD system with bounds of confidence in the PET database, for different  $K$  values in NMF, (b) number of subjects without classification in the NMF-SVM CAD system with bounds of confidence.

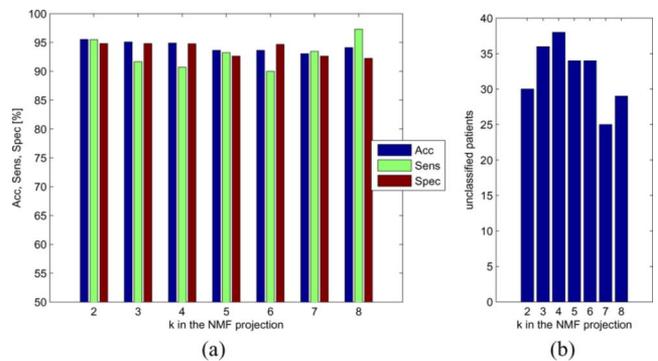


Fig. 8. (a) Performance of the basic NMF-SVM CAD system with bounds of confidence in the SPECT database, for different  $K$  values in NMF, (b) number of subjects without classification in the NMF-SVM CAD system with bounds of confidence.

of the data set: the images are labeled according to a mental state evaluation of the subjects and not with the expertise inspection of the images, as the case of the SPECT data set. This fact makes the initial PET image labeling available less reliable.

### B. Basic NMF-SVM Tool With Bounds of Confidence

The addition of bounds of confidence to the SVM classifier permits the improvement of the results, at the expense of having some unlabeled outputs. Although in this case some subjects are not classified, this approach is similar to real-life cases: sometimes some patients are difficult to be classified with reliability even by experts. Figs. 7 and 8 show the diagnosis results of the

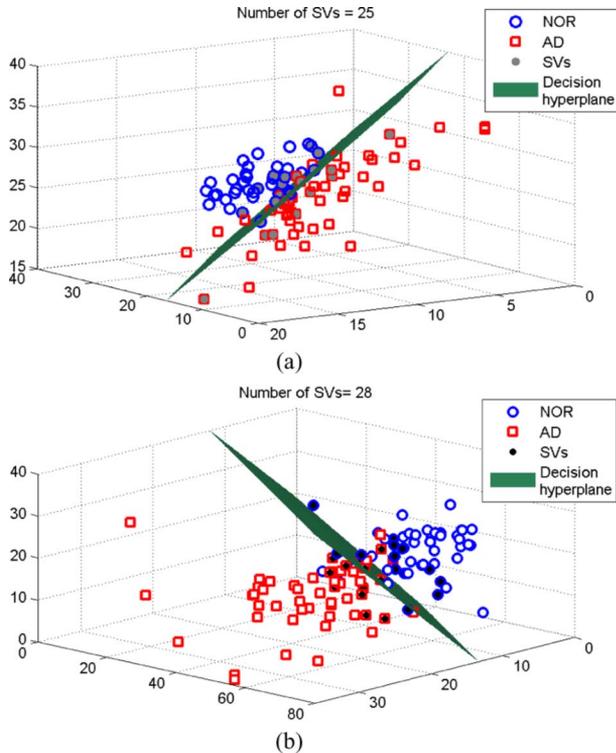


Fig. 9. SVM Separation hyperplane, along with the SVs, of the NMF factorization for  $K = 3$ , prior to the definition of the security region, (a) for the SPECT data, (b) for the PET data. Note the SVs wrongly placed regarding the decision hyperplane.

basic NMF-SVM tool with bounds of confidence, along with the number of unlabeled patients due to the existence of a security region where classification decisions are not allowed.

In Figs. 7 and 8, the presence of a high level of unclassified subjects is directly related to the value of  $K$ . The NMF factorization with a lower  $K$  provides a reduced database in terms of new features representing the data (matrix  $\mathbf{H}$ ), related to the NMF basis (matrix  $\mathbf{W}$ ), where a nonnegligible amount of information is lost regarding the original data set. In these circumstances (e.g.,  $K = 3$ ), the available data for each patient are not relevant enough to establish a clear difference between classes, in terms of linear separability between classes. For the sake of completeness, Fig. 9 provides the SVM separation hyperplane of the NMF factorization of both data sets, for  $K = 3$ , prior to the definition of the bounds of confidence. The linear SVM classifier with bounds of confidence determines the decision region according to the available data of the calibration data set, defining a security region where decisions are risky to be stated. In the case of low values of  $K$  in the NMF factorization, many profiles in the evaluation set are located in the security region defined by the classifier, not being classified as a consequence. Due to the existence of some profiles without classification, a modification of the classifier is proposed (Section III-C).

### C. Modified NMF-SVM CAD Tool With Bounds of Confidence

The approach of Section III-C lets obtain higher levels of Acc, Sens, and Spec without having unclassified subjects, as mentioned before. In this case, the new tool combining the output

TABLE I  
PERFORMANCE [%] OF BASIC NMF-SVM TOOL AND MODIFIED NMF-SVM TOOL WITH BOUNDS OF CONFIDENCE

$K$	Basic NMF-SVM tool							Modified NMF-SVM tool
	2	3	4	5	6	7	8	$K_j = 2:8$
<b>SPECT</b>								
Acc	88.57	88.57	88.57	90.47	87.61	90.47	88.57	<b>91.42</b>
Sens	88.67	88.67	88.67	88.67	86.79	90.56	90.56	<b>90.56</b>
Spec	88.46	88.46	88.46	92.30	88.46	92.30	86.53	<b>92.30</b>
<b>PET</b>								
Acc	80.00	81.90	80.00	79.04	83.80	81.90	79.04	<b>86.59</b>
Sens	81.13	81.13	79.24	79.24	83.01	81.13	77.35	<b>87.50</b>
Spec	80.76	82.69	80.76	78.84	84.61	82.69	80.76	<b>85.36</b>

TABLE II  
PERFORMANCE [%] OF VAF-SVM AS REFERENCE, BEST BASIC NMF-SVM TOOL, AND MODIFIED NMF-SVM TOOL WITH BOUNDS OF CONFIDENCE

	VAF-SVM	Best basic NMF-SVM	Modified NMF-SVM
<b>SPECT</b>			
Acc	81.90	90.47	<b>91.42</b>
Sens	77.35	90.56	<b>90.56</b>
Spec	86.53	92.30	<b>92.30</b>
<b>PET</b>			
Acc	69.54	83.80	<b>86.59</b>
Sens	71.13	83.01	<b>87.50</b>
Spec	67.76	84.61	<b>85.36</b>

of the SVM classifier for different NMF projections provides higher levels of Acc, Sens, and Spec without unclassified subjects. Table I provides the results of this new classifier along with the ones of Section V-A for comparison.

As it is seen, this CAD tool provides the best results in terms of tool performance, without unclassified subjects.

### D. Discussion and Comparison With Other Methods

Some other approaches are available in the literature. Among these techniques, the classical voxels-as-features (VAF) [8] approach for functional images is considered the baseline reference in terms of comparison between results. The VAF approach considers that all the available voxels for each set of subject images provide information and all of them are selected as input features for the train and test step in the CAD system. Thus, in this subsection, both databases are analyzed with a VAF strategy. The classifier selected for this reference case is a linear SVM. Table II provides the results of the VAF-SVM method, for both databases. In addition, the reference results of other techniques are provided, for comparison. Table III provide results obtained with PET and SPECT databases, for a variety of techniques: kernel techniques [24], linear discriminant analysis (LDA) [14], [24], principal component analysis (PCA) [48], [56], independent component analysis [48], [56], etc., for feature selection and reduction, and neural networks (NN) [14], SVM [48], [24], etc., for classification.

In this work, both techniques, NMF and SVM, are combined and applied as constituting parts of a robust CAD system for the early detection of Alzheimer's patterns. Although these techniques are already well known and widely applied, the proper

TABLE III  
RESULTS [%] IN THE LITERATURE WITH OTHER METHODS, FOR COMPARISON

	Acc	Sens	Spec
<b>SPECT</b>			
PCA+Linear_SVM	89.01	92	80.48
ICA+Linear_SVM	84.81	81.58	87.80
VAF+Linear_SVM	72.15	65.79	78.05
FDR+PCA+LDA+Neural_Networks	82.42	84	80.48
PCA+Neural_Networks	85.71	88	82.95
PCA+LDA+Linear_SVM	89.01	88	90.24
quadratic_kernel_PCA+LDA+Linear_SVM	86.81	88	85.36
Polynomial_kernel_PCA+LDA+Linear_SVM	86.81	86	87.80
<b>Modified NMF-SVM</b>	<b>91.42</b>	<b>90.56</b>	<b>92.30</b>
<b>PET</b>			
PCA+Linear_SVM	82.35	70.45	95.12
ICA+Linear_SVM	75.29	59.09	92.68
PCA+LDA+Linear_SVM	81.28	91.61	48.08
PCA+LDA+Neural_Networks	81.28	92.81	44.23
<b>Modified NMF-SVM</b>	<b>86.59</b>	<b>87.50</b>	<b>85.36</b>

combination is an issue of interest in this work and is considered one of the main contributions in this document. In the case of SVM, another important contribution is also stated regarding the use of bounds of confidence and the definition of the security region: it is provided an algorithm for the computation of the SVM classifier with bounds of confidence and the definition of the security region considering the support vectors. In addition, also a new classifier (Section III-C) is proposed, based on the combination of different SVMs with bounds of confidence, considering different  $K$  NMF dimensions in each one. Other contributions are related to the proper results obtained with the application of this CAD tool for the proper identification of the Alzheimer's disease, compared with other strategies.

## VI. CONCLUSION

This paper presents a NMF-SVM based technique for computed aided diagnosis of Alzheimer's disease. The proposed technique is based on the combination of nonnegative matrix factorization (NMF) for feature selection and reduction and SVM with bounds of confidence for classification. The feature reduction step provides a reduced set of variables representing the original data. This feature reduction is especially suitable for machine learning techniques such as SVM. Three different approaches for the classifier are provided and detailed, two of them including bounds of confidence and taking advantage of the definition of a "security region" in the SVM hyperplane, where no decision is assumed. The NMF-SVM CAD tool, along with its variations, is validated with two brain functional image databases: a SPECT data set which provides information about the blood perfusion in the brain and a PET data set which yields information about the glucose metabolism. The validation results of the proposed NMF-SVM method yields up to 91% classification accuracy with high sensitivity and specificity values (upper than 85%) for both data sets.

## ACKNOWLEDGMENT

The authors would like to thank the reviewers for their helpful advices and valuable suggestions to improve the quality of the work provided in this document. The corresponding author is also especially grateful to M. Sierra-Castañer for his entire disposal to facilitate the collaboration of the corresponding author

with SIPBA group in this research during the main period of this work, in 2009. Part of the data used in the preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.<sup>1</sup> 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 manuscript. ADNI investigators include (complete listing available at [http://www.loni.ucla.edu/ADNI/Collaboration/ADNI\\_Manuscript\\_Citations.pdf](http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf)).

## REFERENCES

- [1] J. R. Petrella, R. Coleman, and P. Doraiswamy, "Neuroimaging and early diagnosis of Alzheimer disease: A look to the future," *Radiology*, vol. 226, pp. 315–336, 2003.
- [2] M. N. Wernick and J. N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT*. New York: Elsevier, 2004.
- [3] D. L. Bailey, *Positron Emission Tomography*. New York: Springer Verlag, 2005.
- [4] M. Mignotte and J. Meunier, "Three-dimensional blind deconvolution of SPECT images," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 2, pp. 274–280, Feb. 2000.
- [5] D. Pareto, P. Aguiar, J. Pavia, J. Gispert, A. Cot, C. Falcon, A. Benabarre, F. Lomena, E. Vieta, and D. Ros, "Assessment of SPM in perfusion brain SPECT studies. A numerical simulation study using bootstrap resampling methods," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 7, pp. 1849–1853, Jul. 2008.
- [6] S. Minoshima, K. Frey, and R. Koeppe, "A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of Fluorine-18-FDG PET," *J. Nucl. Med.*, vol. 36, no. 7, pp. 1238–1248, Jul. 1995.
- [7] K. Ishii, A. Kono, H. Sasaki, N. Miyamoto, T. Fukuda, S. Sakamoto, and E. Mori, "Fully automatic diagnostic system for early- and late-onset mild Alzheimer's disease using FDG PET and 3d-SSP," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 33, pp. 575–583, 2006.
- [8] G. Fung and J. Stoeckel, "SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information," *Knowledge Inf. Syst.*, vol. 11, pp. 243–258, 2007.
- [9] E. Westman, A. Simmons, Y. Zhang, J.-S. Muehlboeck, C. Tunnard, Y. Liu, L. Collins, A. Evans, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, and L.-O. Wahlund, "Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls," *NeuroImage*, vol. 54, no. 2, pp. 1178–1187, 2011.
- [10] K. J. Friston, J. Ashburner, S. J. Kiebel, T. E. Nichols, and W. D. Penny, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. New York: Academic, 2007.
- [11] J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Río, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," *Inf. Sci.*, May 2009.
- [12] J. M. Górriz, A. Lassl, J. Ramírez, D. Salas-Gonzalez, C. Puntonet, and E. Lang, "Automatic selection of ROIs in functional imaging using Gaussian mixture models," *Neurosci. Lett.*, vol. 460, no. 2, pp. 108–111, 2009.
- [13] J. Stoeckel, G. Malandain, O. Migneco, P. Koulibaly, P. Robert, N. Ayache, and J. Darcourt, W. Niessen and M. Vieregger, Eds., "Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients," in *Med. Image Computing Computer-Assisted Intervention (MICCAI 2001)*, Berlin, Germany, 2010, vol. 2208, pp. 666–674.
- [14] M. López, J. Ramírez, J. M. Górriz, I. Alvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, and M. Gómez-Río, "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," in *Neurocomputing*, 2011, vol. 74, pp. 1260–1271.
- [15] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, and R. Chaves, "Automatic tool for the Alzheimer's disease diagnosis using PCA and Bayesian classification rules," *IET Electron. Lett.*, vol. 45, no. 7, pp. 342–343, 2009.
- [16] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. A. Illán, C. Puntonet, and M. Gómez-Río, "Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines," *Neurosci. Lett.*, vol. 461, no. 1, pp. 60–64, 2009.

- [17] R. Chaves, J. Ramírez, J. M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, and F. Segovia, "SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test NMSE feature selection with feature correlation weighting," *Neurosci. Lett.*, vol. 461, pp. 293–297, 2009.
- [18] I. Illán, J. M. Górriz, J. Ramírez, D. Salas-González, M. López, F. Segovia, R. Chaves, M. Gómez-Río, and C. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Inf. Sci.*, vol. 181, no. 4, pp. 903–916, 2011.
- [19] A. R. Webb, Ed., *Statistical Pattern Recognition*, 2nd ed. New York: Wiley, 2003.
- [20] I. Álvarez, J. M. Górriz, J. Ramírez, D. Salas, M. López, C. G. Puntonet, and F. Segovia, "Alzheimer's diagnosis using eigenbrains and support vector machines," *IET Electron. Lett.*, vol. 45, no. 1, pp. 165–167, Feb. 2009.
- [21] D. D. Lee and S. Seung, "Algorithms for non-negative matrix factorization," *Adv. Neural Inf. Process. Syst.*, vol. 13, pp. 556–562, 2001.
- [22] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111–126, 1994.
- [23] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
- [24] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-González, F. Segovia, and R. Chaves, "SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA," *Neurosci. Lett.*, vol. 464, no. 3, pp. 233–238, 2009.
- [25] E. Lang, R. Schachtner, D. Lutter, D. Herold, A. Kodewitz, F. Blochl, F. J. Theis, I. R. Keck, J. M. Górriz, P. G. Vilda, and A. M. Tomé, *Recent Advances in Biomedical Signal Processing*. Oak Park, IL: Bentham, 2010.
- [26] X. Liu, S. Yan, and H. Jin, "Projective nonnegative graph embedding," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1126–1137, May 2010.
- [27] I. Buciu and I. Pitas, "Application of non-negative and local non negative matrix factorization to facial expression recognition," in *Int. Conf. Pattern Recognit.*, 2004, vol. 1, pp. 288–291.
- [28] P. Sajda, S. Du, T. Brown, R. Stoyanova, D. Shungu, X. Mao, and L. Parra, "Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain," *IEEE Tran. Med. Imag.*, vol. 23, no. 12, pp. 1453–1465, Dec. 2004.
- [29] J. S. Lee, D. Lee, S. Choi, K. S. Park, and D. S. Lee, "Non-negative matrix factorization of dynamic images in nuclear medicine," in *IEEE Nucl. Sci. Sym. Conf. Rec.*, 2001, vol. 4, pp. 2027–2030.
- [30] W. Liu and N. Zheng, "Non-negative matrix factorization based methods for object recognition," *Pattern Recognit. Lett.*, vol. 25, no. 8, pp. 893–897, 2004.
- [31] S. Zafeiriou and M. Petrou, "Nonlinear non-negative component analysis algorithms," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1050–1066, Apr. 2010.
- [32] C. Ding, T. Li, and M. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.
- [33] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computat. Stat. Data Anal.*, vol. 52, no. 1, pp. 155–173, 2007.
- [34] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
- [35] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 2000.
- [36] C. Burges, *Data Mining and Knowledge Discovery*. Norwell, MA: Kluwer, 1998.
- [37] A. Takemura, A. Shimizu, and K. Hamamoto, "Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the adaBoost algorithm with feature selection," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 598–609, Mar. 2010.
- [38] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, and G. Beddoe, "Shape-based computer-aided detection of lung nodules in thoracic CT images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1810–1820, Jul. 2009.
- [39] M. Moradi, P. Mousavi, A. Boag, E. Sauerbrei, D. Siemens, and P. Abolmaesumi, "Augmenting detection of prostate cancer in transrectal ultrasound images using SVM and RF time series," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 9, pp. 2214–2224, Sep. 2009.
- [40] L. Wei, Y. Yang, R. Nishikawa, M. Wernick, and A. Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications," *IEEE Trans. Med. Imag.*, vol. 24, no. 10, pp. 1278–1285, Oct. 2005.
- [41] I. El-Naqa, Y. Yang, M. Wernick, N. Galatsanos, and R. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1552–1563, Dec. 2002.
- [42] Y. Artan, M. Haider, D. Langer, T. van de Kwast, A. Evans, Y. Yang, M. Wernick, J. Trachtenberg, and I. Yetik, "Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2444–2455, Sep. 2010.
- [43] K. Goatman, A. Fleming, S. Philip, G. Williams, J. Olson, and P. Sharp, "Detection of new vessels on the optic disc using retinal photographs," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 972–979, Apr. 2011.
- [44] H. Narasimha-Iyer, J. Beach, B. Khoobehi, and B. Roysam, "Automatic identification of retinal arteries and veins from dual-wavelength images using structural and functional features," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, pp. 1427–1435, Aug. 2007.
- [45] J. Rojo-Alvarez, J. Bermejo, V. Juarez-Caballero, R. Yotti, C. Cortina, M. Garcia-Fernandez, and J. Antoranz, "Support vector analysis of color-Doppler images: A new approach for estimating indices of left ventricular function," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 1037–1043, Aug. 2006.
- [46] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1357–1365, Oct. 2007.
- [47] J. Morra, Z. Tu, L. Apostolova, A. Green, A. Toga, and P. Thompson, "Comparison of adaboost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 30–43, Jan. 2010.
- [48] I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Río, and C. G. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Inf. Sci.*, vol. 181, pp. 903–916, Feb. 2011.
- [49] J. Levman, T. Leung, P. Causer, D. Plewes, and A. Martel, "Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 688–696, May 2008.
- [50] L. Kuncheva, J. Rodriguez, C. Plumpton, D. Linden, and S. Johnston, "Random subspace ensembles for fMRI classification," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 531–542, Feb. 2010.
- [51] Y. Zhan and D. Shen, "Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method," *IEEE Trans. Med. Imag.*, vol. 25, no. 3, pp. 256–272, Mar. 2006.
- [52] B. Schölkopf, C. J. Burges, and A. J. Smola, *Advances in Kernel Methods—Support Vector Learning*. Cambridge, MA: MIT Press, 1998.
- [53] M. Li and I. Sethi, "Confidence-based active learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, Aug. 2006.
- [54] J. Ramírez, J. M. Górriz, M. Gómez-Río, A. Romero, A. Lassi, A. Rodríguez, C. G. Puntonet, F. Theis, and E. Lang, "Effective emission tomography image reconstruction algorithms for SPECT data," in *Proc. 8th Int. Conf. Computat. Sci.*, 2008, vol. 5101, Lecture Notes in Computer Science, pp. 741–748.
- [55] E. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease," *Neurology*, vol. 34, 1984.
- [56] I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, P. Padilla, and C. Puntonet, "Projecting independent components of SPECT images for computer aided diagnosis of Alzheimer's disease," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1342–1347, 2010.