Support Vector Machines and Neural Networks for the Alzheimer's Disease Diagnosis Using PCA

M. López[1], J. Ramírez[1], J.M. Górriz[1], I. Álvarez[1], D. Salas-Gonzalez[1], F. Segovia[1], and M. Gómez-Río[2]

[1] Dept. of Signal Theory, Networking and Communications
University of Granada, Spain
[2] Department of Nuclear Medicine
Hospital Universitario Virgen de las Nieves, Granada, Spain

Abstract. In the Alzheimer's Disease (AD) diagnosis process, functional brain images such as Single-Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) have been widely used to guide the clinicians. However, the current evaluation of these images entails a succession of manual reorientations and visual interpretation steps, which attach in some way subjectivity to the diagnostic. In this work, two pattern recognition methods have been applied to SPECT and PET images in order to obtain an objective classifier which is able to determine whether the patient suffers from AD or not. A common feature selection stage is first described, where Principal Component Analysis (PCA) is applied over the data to drastically reduce the dimension of the feature space, followed by the study of neural networks and support vector machines (SVM) classifiers. The achieved accuracy results reach 98.33% and 93.41% for PET and SPECT respectively, which means a significant improvement over the results obtained by the classical Voxels-As-Features (VAF) reference approach.

1 Introduction

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

Several approaches have been recently proposed in the literature aiming at providing an automatic tool that guides the clinician in the AD diagnosis process

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 142–149, 2009.

[1,2]. These approaches can be categorized into two types: univariate and multivariate approaches. The first family includes statistical parametric mapping (SPM) [3] and its numerous variants. SPM consists of doing a voxelwise statistical test, comparing the values of the image under study to the mean values of the group of normal images. Subsequently the significant voxels are inferred by using random field theory. It was not developed specifically to study a single image, but for comparing groups of images. The second family is based on the analysis of the images, feature extraction and posterior classification in different classes. Among these techniques, we can find the classical Voxels-As-Features (VAF) approach for SPECT images [1]. The main problem to be faced up by these techniques is the well-known small size sample problem, that is, the number of available samples is much lower than the number of features used in the training step.

Principal Component Analysis (PCA) corresponds to multivariate approaches and was already applied to functional brain images in [3] in a descriptive fashion, where the impossibility of using this transformation to make any statistical inference is highlighted. However, in this work, a new approach of PCA is used in combination with supervised learning methods, which in turn solves the small size sample problem since the dimension of the feature space undergoes a significant reduction. The task of the supervised learner is to predict the class of the input object after having seen a number of training examples. In this work, two of the most widely used classifiers are trained on these PCA coefficients: Support Vector Machines (SVMs) and Neural Networks (NN), and their performances in the classification task we are dealing with are compared.

2 Image Preprocessing and Feature Extraction

SPECT and PET images used in this work were taken with a PRISM 3000 machine and a SIEMENS ECAT 47 respectively. 3D brain perfusion volumes are reconstructed from projection data using the filtered backprojection (FBP) in combination with a Butterworth noise filter. All the images are spatially normalized using the SPM software [3] in order to ensure that the voxels in different images refer to the same anatomical positions in the brain [4], giving rise to volumes of size $69\times95\times79$. Finally, intensity level of the SPECT and PET images is normalized to the maximum intensity. The dimension of the volume representing each subject brain was reduced to $17\times23\times19$ by decimating the original 3D volume by a $4\times4\times4$ factor. After that, as proposed in [2], a mask is applied so that voxels whose mean intensity value averaged over all images is lower than the half of the maximum mean intensity value are rejected. This is done to reduce the dimension of the feature space and remove irrelevant information.

2.1 Principal Component Analysis and Eigenbrains

Principal Component Analysis (PCA) generates an orthonormal basis vector that maximizes the scatter of all the projected samples. After the preprocessing steps,

the N remaining voxels for each subject are rearranged in a vector form. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n]$ be the sample set of these vectors, where n is the number of patients. After normalizing the vectors to unity norm and subtracting the grand mean, a new vectors set $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_n]$ is obtained, where each $\mathbf{Y}_i$ represents a normalized vector with dimensionality N, $\mathbf{Y}_i = (y_{i1}, y_{i2}, ..., y_{iN})^t, i = 1, 2, ..., n$. The covariance matrix of the normalized vectors set is defined as

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

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

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

Note that $\mathbf{Y}\mathbf{Y}^t$ is an $N \times N$ matrix while $\mathbf{Y}^t\mathbf{Y}$ is an $n \times n$ matrix. If the sample size n is much smaller than the dimensionality N, then diagonalizing $\mathbf{Y}^t\mathbf{Y}$ instead of $\mathbf{Y}\mathbf{Y}^t$ reduces the computational complexity [5]

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

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

where $\mathbf{\Lambda}_1 = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}$ and $\mathbf{T} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_n]$. Derived from the *eigenface* concept [5], the *eigenbrains* correspond to the dominant eigenvectors of the covariance matrix. In this approach, only m leading eigenvectors are used, which define the matrix $\mathbf{P}$

$$\mathbf{P} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_m] \tag{5}$$

The criterion to choose the most discriminant eigenbrains is set by their separation ability, which is measured by the Fisher Discriminant Ratio (FDR), defined as

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{6}$$

where μ_i and σ_i denote the i-th class within class mean value and variance, respectively. For the whole database, a matrix of weights can be constructed, given by:

$$\mathbf{Z} = \mathbf{P}^t\mathbf{Y} \tag{7}$$

3 Overview of Classifiers

The goal of a binary classifier is to separate a set of binary labeled training data consisting of, in the general case, N-dimensional patterns $\mathbf{x}_i$ and class labels y_i:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l) \in (R^N \times \{\text{Normal}, \text{AD}\}), \tag{8}$$

so that a classifier is produced which maps an unknown object $\mathbf{x}_i$ to its classification label y_i.

3.1 Support Vector Machines

Support vector machines (SVM) [6] separate binary labeled training data by the hyperplane

$$g(\boldsymbol{x}) = \mathbf{w}^T \boldsymbol{x} + w_0 \tag{9}$$

where $\mathbf{w}$ is known as the weight vector and w_0 as the threshold. This hyperplane is maximally distant from the two classes (known as the maximal margin hyperplane). The objective is to build a function $f : R^N \longrightarrow \{\pm 1\}$ using training data that is, N-dimensional patterns $\mathbf{x}_i$ and class labels y_i:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_l, y_l) \in R^N \times \{\pm 1\}, \tag{10}$$

so that f will correctly classify new examples $(\mathbf{x}, y)$. When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space. If the data is mapped to some other (possibly infinite dimensional) Euclidean space using a mapping $\Phi(\mathbf{x})$, the training algorithm only depends on the data through dot products in such an Euclidean space, i.e. on functions of the form $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. If a "kernel function" K is defined such that $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, it is not necessary to know the Φ function during the training process. In the test phase, an SVM is used by computing the sign of

$$f(\mathbf{x}) = \sum_{i=1}^{N_S} \alpha_i y_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{x}) + w_0 = \sum_{i=1}^{N_S} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + w_0, \tag{11}$$

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

3.2 Neural Networks

An Artificial Neural Network (ANN) [7] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, processes information. ANNs can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges (with weights) are connections between neuron outputs and neuron inputs. Based on the connection pattern (architecture), ANNs can be grouped into two categories: *i*) feed-forward networks, in which graphs have no loops, and *ii*)recurrent (or feedback) networks, in which loops occur because of feedback connections. Different connectivities yield different network behaviors. Generally speaking, feed-forward networks are static, that is, they produce only one set of output values rather than a sequence of values from a given input. Feed-forward networks are memory-less in the sense that their response to an input is independent of the previous network state. Recurrent, or feedback, networks, on the other hand, are dynamic systems. When a new input pattern is presented, the neuron outputs are computed. Because of

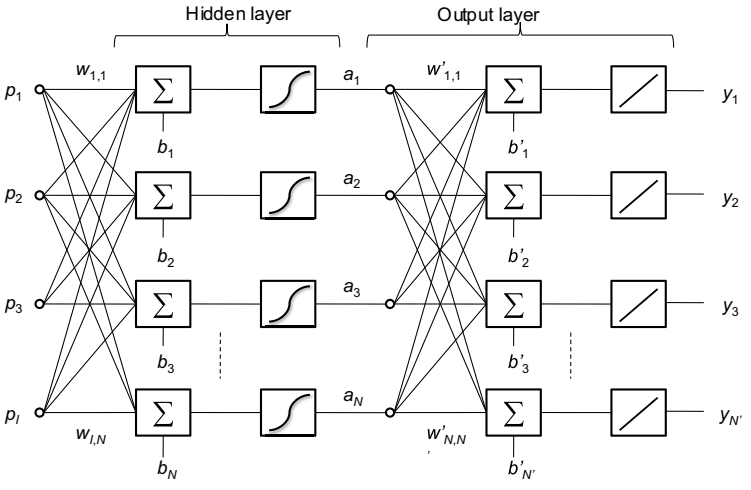


Fig. 1. Feed-forward neural network architecture with hidden layer of neurons plus linear output layerll

the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state.

Feed-forward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons as shown in Fig. 1. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.

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

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

- One hidden layer and increasing number of neurons and a linear output layer.
- Hyperbolic tangent sigmoid transfer function: $f(n) = 2/(1+exp(-2*n))-1$, for input layers.
- Linear transfer function: $f(n) = n$, for output layer.
- Weight and bias values are updated according to Levenberg-Marquardt optimization.
- Gradient descent with momentum weight and bias is used as learning function.

4 Evaluation Results

The databases used in this work consist of 91 SPECT patients (41 labeled as NORMAL and 50 labeled as AD) and 60 PET patients (18 NORMAL and 42 AD).

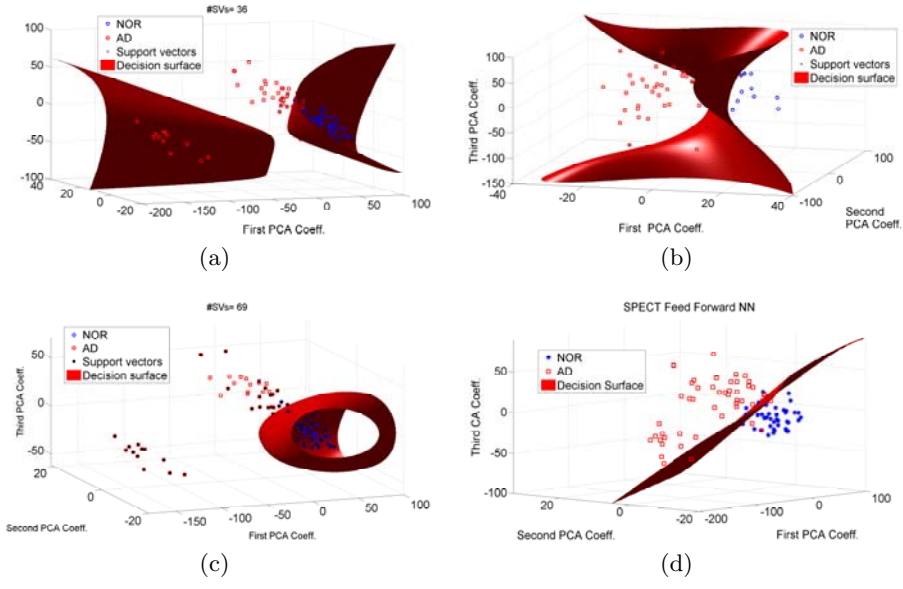


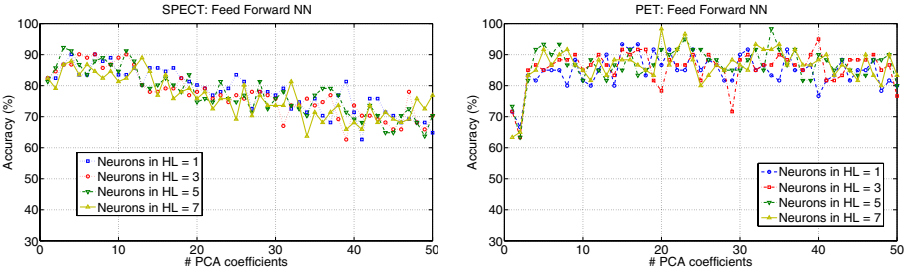
Fig. 2. Decision surfaces for SPECT images and different classifiers: (a) SVM (Quadratic), (b) SVM (Polynomial), (c) SVM (RBF), and (d) Feed-forward network

Table 1. Results obtained from the evaluation of SVM and feed-forward neural networks classifiers using PCA coefficients as features. Comparison to the VAF baseline.

	PET		SPECT	
SVM	Baseline	PCA ($m = 15$)	Baseline	PCA ($m = 3$)
Linear	96.67%	95.00%	87.71%	91.21%
Quadratic	96.67%	96.67%	82.41%	90.11%
Polynomial	30.00%	96.67%	54.95%	89.01%
RBF	70.00%	83.33%	54.95%	**93.41%**

Feed-Forward	PET ($m = 20$)	SPECT ($m = 4$)
1 Neur. in HL	86.67%	90.11%
3 Neur. in HL	78.33%	87.91%
5 Neur. in HL	91.67%	91.21%
7 Neur. in HL	**98.33%**	87.91%

The reference VAF system was compared to different classifiers using the proposed PCA features. All the classifiers were tested using the Leave-One-Out cross-validation strategy. The eigenbrain space is computed using all the patients except one. The test patient to be classified is projected into the eigenbrain space, so that projection coefficients $\mathbf{Z}$ are obtained and sorted out according to their associated FDR, as explained in Sec. 2.1. In order to determine the optimal number m of coefficients to be used for each classifier, they were all tested varying m from 1 to 50. Fig. 2 shows the 3D input space defined when three PCA coefficients ($m = 3$)



(a) (b)

Fig. 3. Accuracy for (a) SPECT and (b) PET images using feed-forward neural networks when the number m of PCA coefficients considered for the classification task increases

are used as features for SPECT images, and the ability of SVM and feed-forward neural network classifiers to separate the two classes by means of carefully trained decision surfaces. The shape of the decision rule strongly depends on the method for formulating the decision rule and its associated parameters. For feed-forward neural networks, Fig. 3 shows the accuracy values obtained when m varies from 1 to 50, for different number of neurons in the hidden layer (HL). The reference VAF approach, which directly uses the images intensity levels to train a SVM classifier, was also implemented, and results can be compared in Table 1.

5 Conclusions

A comparative between SVM and feed-forward networks for the classification of functional brain images to diagnose the AD was shown in this paper. The proposed features to be used in the training steps are the PCA image coefficients, resulting from the projection of the test patient into the eigenbrain space, which is computed over the rest of the patients. The introduction of PCA coefficients as features for the classification task clearly improves the accuracy values obtained by the VAF approach, especially when SVM with polynomial and radial basis function (RBF) kernels are used, yielding the best accuracy result for SPECT images (93.41%). For PET images instead, the use of feed-forward neural network provided the best result, reaching 98.33% accuracy.

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

References

1. Stoeckel, J., Malandain, G., Migneco, O., Koulibaly, P.M., Robert, P., Ayache, N., Darcourt, J.: Classification of SPECT images of normal subjects versus images of alzheimer's disease patients. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 666–674. Springer, Heidelberg (2001)
2. Górriz, J.M., Ramírez, J., Lassl, A., Salas-Gonzalez, D., Lang, E.W., Puntonet, C.G., Álvarez, I., López, M., Gómez-Río, M.: Automatic computer aided diagnosis tool using component-based svm. In: 2008 IEEE Nuclear Science Symposium Conference Record, pp. 4392–4395 (2008)
3. Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, London (2007)
4. Salas-González, D., Górriz, J.M., Ramírez, J., Lassl, A., Puntonet, C.G.: Improved gauss-newton optimization methods in affine registration of SPECT brain images. IET Electronics Letters 44(22), 1291–1292 (2008)
5. Turk, M., Petland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 13(1), 71–86 (1991)
6. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2(2), 121–167 (1998)
7. McCulloch, W.S., Pitts, W.: A logical calculus of ideas immanent in nervous activity. Bull. Mathematical Biophysics 5, 115–133 (1943)