Automatic System for Alzheimer's Disease Diagnosis Using Eigenbrains and Bayesian Classification Rules

M. López[1,⋆], J. Ramírez[1], J.M. Górriz[1], I. Álvarez[1], D. Salas-Gonzalez[1], F. Segovia[1], and C.G. Puntonet[2]

[1] Dept. of Signal Theory, Networking and Communications, University of Granada, Spain
[2] Dept. of Architecture and Computer Technology, University of Granada, Spain

Abstract. Alzheimer's Disease (AD) is a progressive neurologic disease of the brain that leads to the irreversible loss of neurons and dementia. The new brain imaging techniques PET (Positron Emission Tomography) and SPECT (Single Photon Emission Computed Tomography) provide functional information about the brain activity and have been widely used in the AD diagnosis process. However, the diagnosis currently relies on manual image reorientations, visual evaluation and other subjective, time consuming steps. In this work, a complete computer aided diagnosis (CAD) system is developed to assist the clinicians in the AD diagnosis process. It is based on bayesian classifiers, made up from features previously extracted. The small size sample problem, consisting of having a number of available samples much lower than the dimension of the feature space, is faced up by applying Principal Component Analysis (PCA) to the features. This approach provides higher accuracy values than other previous approaches do, yielding 91.21% and 98.33% accuracy values for SPECT and PET images, respectively.

Keywords: SPECT, PET, Alzheimer Type Dementia, Principal Component Analysis, Bayesian Classification.

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. The current method to diagnose the AD is based on a whole study of the patient, including personal interviews with the patient's family members, physical and neurological exams and brain imaging.

⋆ This work was partly supported 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 Project (TIC-02566).

J. Cabestany et al. (Eds.): IWANN 2009, Part I, LNCS 5517, pp. 949–956, 2009.

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. Two types of non-invasive (i. e., no surgery is required) tests have been widely used in the AD diagnosis. Positron Emission Tomography (PET) scan is an imaging scan that measures the level of functional activity of the brain by measuring its use of glucose. Single Photon Emission Computed Tomography (SPECT) scan is a procedure that measures blood flow in different areas of the brain. For both tests PET and SPECT, a small amount of radioactive material is injected into the patient and emission detectors are placed on the brain, providing functional information about the brain activity. Nowadays, the evaluation of the brain images is carried out by an expert clinician who manually reorients and visually examines the images, giving a diagnostic about the patient's state.

Several approaches have been recently proposed in the literature aiming at providing an automatic tool that guides the clinician in the AD diagnosis process [1,2]. These approaches can be categorized into two types: univariate and multivariate approaches. The first family is 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, depending on the patient's state. The main problem to be faced up by these techniques is the well-known *curse of the dimensionality*, that is, the number of available samples is much lower than the number of features used in the training step. Among these techniques, Voxels-As-Features (VAF) for SPECT images [1] and others based on the study of regions of interest (ROIs) for SPECT and PET images [4,5] have been developed.

In this work, a fully computer aided diagnosis (CAD) tool is shown. This approach belongs to the multivariate family, and faces up the small size sample problem by applying Principal Component Analysis (PCA) to the feature vector, which allows us to reduce drastically the dimension and makes it comparable to the number of training samples. PCA was already applied in [3] in a qualitative way, but never used the coefficients as features for classification. In this approach, after the PCA transformation, the resultant feature vectors are used to made up a bayesian classifier, which uses the *a posteriori* information to classify new coming images.

2 Image Preprocessing

In order to make possible a direct comparison of the voxel intensities between SPECT or PET images, a previous normalization of the images is needed. For all the experiments, we normalize the images by applying an affine transformation to the intensities as suggested in [3]. All the images of the database are transformed

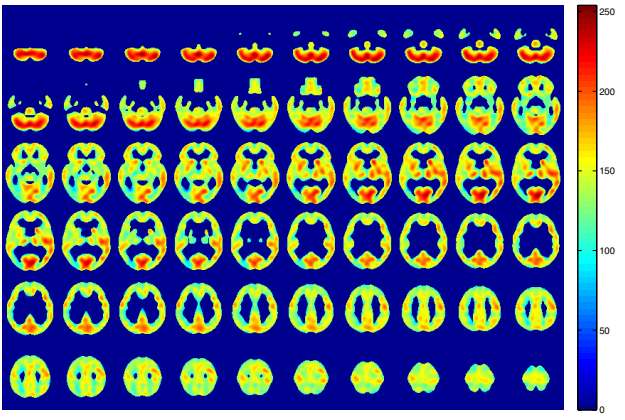

Fig. 1. SPECT average image of the dataset once the mask has been applied

using affine and non-linear spatial normalization, thus the basic assumptions are met.

After the normalization steps, a $79 \times 95 \times 69$ brain volume representation is obtain for each subject. However, not all these voxels are significant in the AD diagnosis. Voxels in the black edges outside the brain zone or voxels with low gray level (both in normal and AD patients) can be rejected for the classification task. As proposed in [4], we first construct a binary mask which selects the voxels of interest and discards the rest. This is done by taking the voxels whose mean intensity value averaged over all images exceeds the half of the maximum mean intensity value, and this mask is applied to the original images. In the resulting masked images, the irrelevant information has been removed or reduced. Fig. 1 represents the masked SPECT average image along the transaxial axis. The mask application has rejected those voxels whose intensity values are lower than 128.

3 Principal Component Analysis and Eigenbrains

PCA generates a set of orthonormal basis vectors, known as principal components, which maximize the scatter of all the projected samples. After the preprocessing steps, the N remaining voxels 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, each of dimensionality N. 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 [6]

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

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

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 [6], 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:

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

4 Bayes Classifier

For pattern recognition, the Bayes classifier is the best classifier in terms of minimum Bayes error, therefore the *a posteriori* probability functions will be evaluated [7]. Let ω_1 and ω_2 denote the object classes (AD and NORMAL), and $\mathbf{Z}$ a patient voxels vector in the reduced PCA subspace. The *a posteriori* probability function of ω_i given $\mathbf{Z}$ is defined as

$$P(\omega_i|\mathbf{Z}) = \frac{p(\mathbf{Z}|\omega_i)P(\omega_i)}{p(\mathbf{Z})}, \quad i = 1, 2. \tag{8}$$

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

$$p(\mathbf{Z}|\omega_i)P(\omega_i) = \max_j \{p(\mathbf{Z}|\omega_j)P(\omega_j)\}, \quad \mathbf{Z} \in \omega_i \tag{9}$$

The brain projected data $\mathbf{Z}$ is classified to ω_i of whom the *A Posteriori* probability given $\mathbf{Z}$ is the largest between the classes. Usually there are not enough samples to estimate the conditional probability density function for each class (within class density). The within class densities are usually modeled as normal distributions

$$p(\mathbf{Z}|\omega_i) = \frac{1}{(2\pi)^{m/2}|\mathbf{\Sigma}_i|^{1/2}} \times exp\left\{-\frac{1}{2}(\mathbf{Z}-\mathbf{M}_i)^t\mathbf{\Sigma}_i^{-1}(\mathbf{Z}-\mathbf{M}_i)\right\} \tag{10}$$

where $\mathbf{M}_i$ (see Eq. 11) and $\mathbf{\Sigma}_i$ are the mean and covariance matrix of class ω_i, respectively.

4.1 Probabilistic Reasoning Model (PRM)

Under the unified Bayesian framework, two new probabilistic reasoning models, PRM-1 and PRM-2 are derived in [8], which utilize the within class scatters to derive averaged estimations of within class covariance matrices.

In particular, let ω_1, ω_2 and N_1, N_2 denote the classes and number of patients within each class, respectively. Let $\mathbf{M}_1, \mathbf{M}_2$ be the means of the classes in the reduced PCA subspace span $[\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_m]$. We then have

$$\mathbf{M}_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{Z}_j^{(i)}, \quad i = 1, 2. \tag{11}$$

where $\mathbf{Z}_j^{(i)}, j = 1, 2, ..., N_i$ represents the sample voxels vector for the ω_i class.

PRM-1. The PRM-1 model assumes the within class covariance matrices are identical and diagonal

$$\mathbf{\Sigma}_i = diag\left\{\sigma_1^2, \sigma_2^2, ..., \sigma_m^2\right\} \tag{12}$$

Each component σ_i^2 can be estimated by sample variance in the one dimensional PCA subspace

$$\sigma_i^2 = \frac{1}{L}\sum_{k=1}^{L}\left\{\frac{1}{N_k - 1}\sum_{j=1}^{N_k}\left(z_{ij}^{(k)} - m_{ki}\right)^2\right\} \tag{13}$$

where $z_{ij}^{(k)}$ is the i-th element of the sample $\mathbf{Z}_j^{(k)}$, m_{ki} the i-th element of $\mathbf{M}_k$, and L the number of classes (two in our approach).

PRM-2. The PRM-2 model estimates the within class scatter matrix Σ_ω in the reduced PCA space as

$$\Sigma_\omega = \frac{1}{L}\sum_{k=1}^{L}\left\{\frac{1}{N_k}\sum_{j=1}^{N_k}\left(Z_j^{(k)} - M_k\right)\left(Z_j^{(k)} - M_k\right)^t\right\} \tag{14}$$

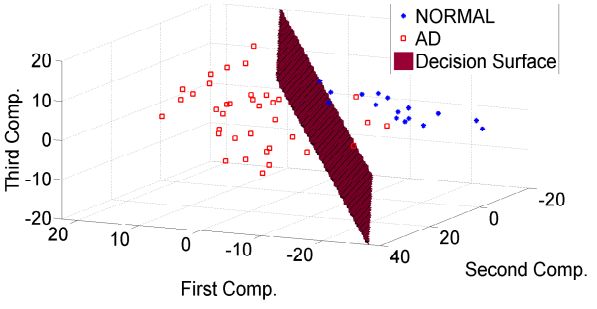

Fig. 2. For PET images, distributions of the three first principal coefficients for AD and NORMAL patients and the decision surface

The Σ_ω is diagonilized using singular value decomposition (SVD)

$$\Sigma_\omega = USV^t \tag{15}$$

where U and V are unitary matrices, S is a diagonal matrix

$$S = diag\{s_1, s_2, ..., s_m\} \tag{16}$$

with non-negative singular values as diagonal elements. The squared diagonal elements are ordered in decreasing value

$$\left(s^2_{(1)}, s^2_{(2)}, ..., s^2_{(m)}\right) = order\{s^2_1, s^2_2, ..., s^2_m\} \tag{17}$$

Finally, the within class covariance matrix is derived as

$$\Sigma_I = diag\left\{s^2_{(1)}, s^2_{(2)}, ..., s^2_{(m)}\right\} \tag{18}$$

5 Evaluation Results

SPECT and PET images used in this work were taken with a PRISM 3000 machine and a SIEMENS ECAT 47 respectively. Initially they were labeled by experienced clinicians of the "Virgen de las Nieves" Hospital (Granada, Spain) and "Clínica PET Cartuja" (Seville, Spain) respectively. The database consists of 91 SPECT patients (41 labeled as NORMAL and 50 labeled as AD) and 60 PET patients (18 NORMAL and 42 AD). Initially, the original brain image $79 \times 95 \times 69$ voxel sized is reduced by averaging over subsets of $4 \times 4 \times 4$ voxels. After applying the mask, the remaining voxels are rearranged into a vector form so that PCA can be applied to the training set and eigenbrains are obtained. The new patient to be classified is projected into the eigenbrain space, the *a posteriori* probabilities $P(\mathbf{Z}|\omega_1)$ and $P(\mathbf{Z}|\omega_2)$ are computed (where ω_1 = NORMAL and ω_2 = AD) and

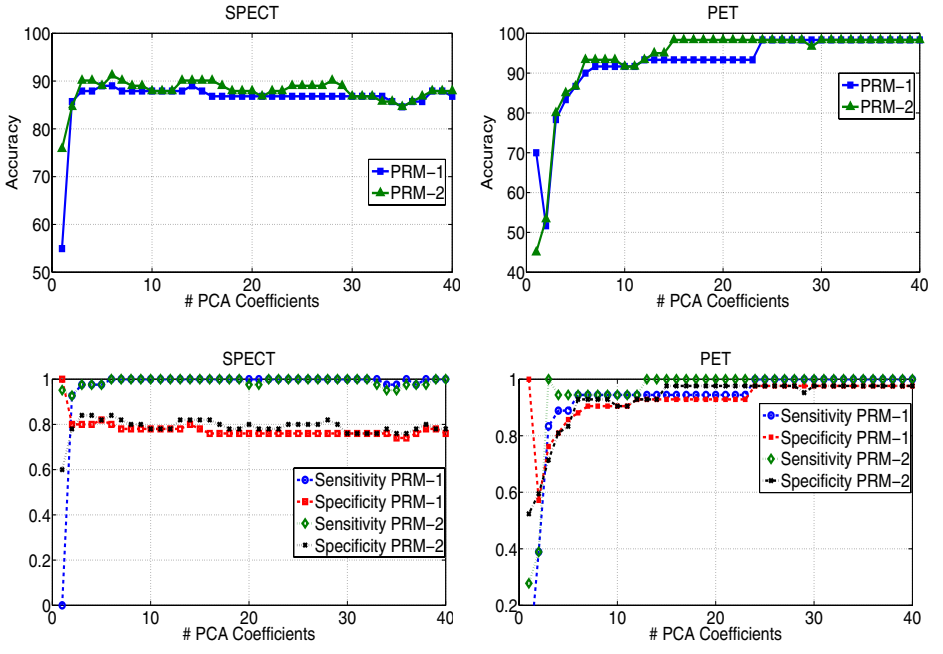

Fig. 3. Accuracy (up) and Sensitivity/Specificity (down) values for SPECT (left) and PET (right) images using bayesian classifiers when the number m of considered principal components increases

Table 1. Accuracy results obtained by the proposed methods and by the work taken as reference (VAF)

	SPECT	PET
VAF approach	85.71%	96.67%
Eigenbrains PRM-1	89.01%	98.33%
Eigenbrains PRM-2	**91.21%**	**98.33%**

the MAP rule (Eq. 9) is applied, both for PRM-1 and PRM-2 models. Fig. 2 represents the three main PCA coefficients as 3D points for NORMAL and AD patients separated by the classifier decision surface. Fig. 3 shows the accuracy and the sensitivity and specificity values obtained for SPECT and PET images when the number of considered principal component m increases.

As it is usually done in cases where the number of available samples is relatively small, the classifier was tested with the Leave-One-Out method, that is, the classifier is trained using all but one patient, which is used in the test phase. This procedure is repeated for each patient and an average accuracy value for all the experiments is obtained. The VAF approach was also implemented and tested by the same cross-validation strategy. Results are shown and compared in Table 1.

6 Conclusions

A computer aided diagnosis system for the early detection of the Alzheimer disease was shown in this paper. The system was developed by performing the principal components analysis to a subset of remaining voxels after some preprocessing steps, which allows us to face the small size problem and reduce drastically the feature space dimension. The most important components in terms of ability to separate are chosen to make up a bayesian classifier. The *a posteriori* information is used when a new patient is needed to be classified. With this approach, 91.21% and 98.33% accuracy values are obtained for SPECT and PET images respectively, which mean an improvement over the reference work.

References

1. Fung, G., Stoeckel, J.: SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease Using Spatial Information. Knowledge and Information Systems 11(2), 243–258 (2007)
2. Ramírez, J., Górriz, J.M., Salas-Gonzalez, D., Lassl, A., López, M., Puntonet, C. G., Gómez, M., Rodríguez, A.: Computer Aided Diagnosis of Alzheimer Type Dementia Combining Support Vector Machines and Discriminant Set of Features. Accepted in Information Sciences (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. 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: Medical Imaging Conference, Dresden (2008)
5. Álvarez, I., López, M., Górriz, J.M., Ramírez, J., Salas-Gonzalez, D., Puntonet, C.G., Segovia, F.: Automatic Classification System for the Diagnosis of Alzheimer Disease Using Component-Based SVM Aggregations. In: Proc. International Conference on Neural Information Processing (2008)
6. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 13(1), 71–86 (1991)
7. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press, London (1991)
8. Chengjun, L., Wechsler, H.: A Unified Bayesian Framework for Face Recognition. In: Proceedings in International Conference on Image Processing (1998)