Alzheimer's Diagnosis Using Eigenbrains and Support Vector Machines

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

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

Abstract. An accurate and early diagnosis of the Alzheimer's Disease (AD) is of fundamental importance for the patients medical treatment. Single Photon Emission Computed Tomography (SPECT) images are commonly used by physicians to assist the diagnosis, rating them by visual evaluations. In this work we present a computer assisted diagnosis tool based on a Principal Component Analysis (PCA) dimensional reduction of the feature space approach and a Support Vector Machine (SVM) classification method for improving the AD diagnosis accuracy by means of SPECT images. The most relevant image features were selected under a PCA compression, which diagonalizes the covariance matrix, and the extracted information was used to train a SVM classifier which could classify new subjects in an unsupervised manner.

1 Introduction

Distinguishing AD from other causes of dementia still remains a diagnostic challenge, specially during the early stage of the disease, that offers better opportunities to treat its symptoms. Thus, an accurate and early diagnosis of the AD by means of non-invasive methods, is of fundamental importance for the patients medical treatment. Nuclear imaging as Single Photon Emission Computed Tomography (SPECT) or Positron Emission Tomography (PET) are examples of non-invasive, three-dimensional functional imaging modalities that provide clinical information regarding biochemical and physiologic processes in patients, and are frequently used as a diagnostic tool in addition to the clinical findings.

The examination of the predictive abilities of nuclear imaging with respect to AD and other dementia illnesses is usually done through visual assessments performed by experts[1, 2]. However, statistical classification methods have not been widely used to assist the diagnosis, being Statistical Parametric Mapping (SPM) the most extended tool in the neuro-imaging community[3, 4]. It consists of doing a voxel-wise 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 [5]. This method suffers

J. Cabestany et al. (Eds.): IWANN 2009, Part I, LNCS 5517, pp. 973–980, 2009.

the inconvenient of local and mono-variate approaches and it was not developed specifically for the typical case of a single image study, but for comparing groups of images. One can use it for diagnostics by comparing the image under study to a group of normal images. This comparison has the disadvantages of assessing a population containing just one individual and being lacking in any information about the pathology, that the other population of AD affected images contain.

On the other hand, multivariate approaches that consider as one observation all the voxels in a single scan, suffer from the *curse of dimensionality* problem. This major problem, associated with pattern recognition systems, occurs when the number of available features for designing the classifier is very large compared with the number of available training examples. The importance of multivariate approaches is that the interactions among voxels and error effects are assessed statistically, while paying the price of losing capability of making statistical inferences about regionally specific changes.

Principal Component Analysis (PCA) is an example of multivariate technique that require more training examples than features and is not suitable for making any statistical inference about the characterizations that it obtains [3]. In the Alzheimer's disease aided diagnosis, it has been used as a mathematical device that simply identify prominent patterns of correlations or functional connectivity of brain regions, to be analyzed with other statistical tools as SPM, ANOVA or MANCOVA [6, 7, 8].

Our approach to the computer aided diagnosis (CAD) involves not statistical inference but machine learning techniques, which are appropriate for single image studies. In this work we used two combined techniques that, independently have been successful in solving several classification problems. Firstly, we made use of *eigenbrains* or eigenimages, that were obtained from the Principal Components [3], to reduce the dimension of the feature space to a set of projection coefficients (see Sect. 2), in a similar fashion as in face detection [9]. This process reduced the dimensionality of the feature space from $\sim 5 \cdot 10^5$ to $\sim 10^2$, thus facing the small sample size problem. Secondly, once a significant feature ensemble was selected, we built a SVM to manage the classification task.

Support Vector Machines (SVMs) have marked the beginning of a new era in the learning from examples paradigm [10]. Recently, SVMs have attracted attention from the pattern recognition community due to a number of theoretical and computational merits derived from [10]. These techniques have been successfully applied to many fields including voice activity detection (VAD) [11], content-based image retrieval [12], and medical imaging diagnosis [13, 14]. Somehow, the application of SVM to high dimensional and small sample size problems is still a challenge and improving the accuracy SVM based-approaches is still a field in development[15, 16].

The combination of these two methods grows a CAD system for the early detection of Alzheimer Type Dementia (ATD) tested over SPECT images, and developed with the aim of reducing the subjectivity in visual interpretation of these scans by clinicians, thus improving the accuracy of diagnosing Alzheimer's disease in its early stage.

2 PCA Application to SPECT Images: Eigenbrains

Principal Component Analysis is a standard technique for extracting the most significant features from a dataset. It is based on a linear transformation acting on a zero mean dataset, that diagonalizes its covariance matrix. In brain images, the dataset is an ensemble of 3D brain images $\mathbf{\Gamma}_i$, whose size M is typically $79 \times 95 \times 69 \sim 5 \cdot 10^5$ voxels. Let the full 3D brain image set be $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, ..., \mathbf{\Gamma}_N$, each understood as a vector of dimension M. The average brain image of the dataset is defined as $\mathbf{\Gamma} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{\Gamma}_n$. We need firstly to extract the average of the image set to each one, producing a new set $\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \mathbf{\Gamma}$ with $n = 1, 2, ..., N$. On this set, a PCA transformation is composed by M orthogonal vectors $\mathbf{u}_i$, such that

$$\lambda_i = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{u}_i^T\mathbf{\Phi}_n)^2 \tag{1}$$

is maximum, subject to the constrain

$$\mathbf{u}_i^T\mathbf{u}_j \;=\; \delta_{ij} \tag{2}$$

where δ_{ij} is the Kronecker delta. The resulting $\mathbf{u}_i$ and λ_i are the eigenvectors and eigenvalues respectively of the covariance matrix:

$$\mathbf{C} = \frac{1}{N}\sum_{i=n}^{N}\mathbf{\Phi}_n\mathbf{\Phi}_n^T = \mathbf{A}\mathbf{A}^T \tag{3}$$

where $\mathbf{A} = [\mathbf{\Phi}_1, ..., \mathbf{\Phi}_N]$. We will refer to this orthogonal eigenvector basis $\{\mathbf{u}_i\}, i = 1, ..., M$ as eigenbrains, because of its brain like appearance (see Fig. 1). To obtain them, it is necessary to diagonalize the $M \times M$ covariance matrix, which for brain images would be approximately a $5{\cdot}10^5 \times 5{\cdot}10^5$ matrix. The computational complexity of the diagonalization process can be significantly reduced by diagonalizing the matrix $\hat{\mathbf{C}} = \mathbf{A}^T\mathbf{A}$, whose size is $N \times N$, with $N << M$ [9]. This allows to obtain N of the M eigenvectors $\mathbf{u}_n$ of $\mathbf{C}$, from the eigenvectors $\mathbf{v}_n$ of $\hat{\mathbf{C}}$ as $\mathbf{u}_n = \mathbf{A}\mathbf{v}_n$, $n = 1, .., N$. Usually, the first few eigenbrains explain the whole variance, so only a number $M' < M$ is necessary to appropriately describe the dataset (in this case $M' = N$). The obtained eigenbrains span a new subspace which we refer to as the "eigenbrain space". For the classification task we projected each image into the previously defined eigenbrain space. Each projected image produced a vector of weights so that a matrix of weights can be constructed with the whole database. This matrix $\mathbf{\Omega}$ is given by:

$$\mathbf{\Omega}_{in} = \mathbf{u}_n^T\mathbf{\Phi}_i, \qquad n = 1, 2, ..., M', \quad i = 1, 2, ..., N \tag{4}$$

and describes the contribution of each eigenbrain in representing the input brain image $\mathbf{\Phi}_i$, treating the eigenbrains as a basis set for brain images (see Fig. 1). The matrix $\mathbf{\Omega}$ contains the most significant information extracted from the principal component analysis, stacked in a $N \times M'$ data ensemble. We used this matrix $\mathbf{\Omega}$ for the following classification task, as N M'-dimensional patterns:

$$\mathbf{x}_i = [\mathbf{\Omega}_{i1}, \mathbf{\Omega}_{i2}, ..., \mathbf{\Omega}_{iM'}], \quad i = 1, 2, ..., N \tag{5}$$

each of them with its corresponding class label $y_i \in \{\pm 1\}$.

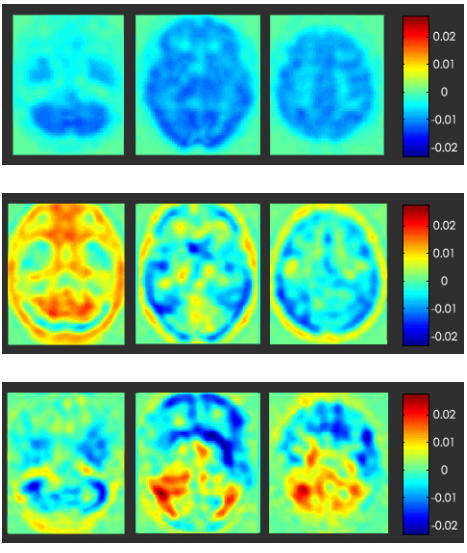

Fig. 1. Three representative transversal slices of the first three eigenbrains, ranked by its eigenvalue

3 Background on SVMs

The classification is achieved through a SVM, that separates a given set of binary labeled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyper-plane). The objective is to build a function $f : \mathbb{R}^{M'} \longrightarrow \{\pm 1\}$ using training data, consisting of M'-dimensional patterns $\mathbf{x}_i$ and class labels y_i:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_N, y_N) \in \left(\mathbb{R}^{M'} \times \{\pm 1\}\right), \tag{6}$$

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 using the *kernel trick*, so that the hyperplane defining the SVM corresponds to a non-linear decision boundary in the input space [10]. In this way the decision function f can be expressed in terms of the *support vectors* only [10]:

$$f(\mathbf{x}) = \text{sign}\{\sum_{i=1}^{N_S} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + w_0\}, \tag{7}$$

where $K(.,.)$ is the kernel function, α_i is a weight constant derived form the SVM process and $\mathbf{s}_i$ are the support vectors [10]. Common kernels that are used by SVM practitioners for the nonlinear feature mapping are:

– Polynomial

$$K(\boldsymbol{x}, \boldsymbol{y}) = [\gamma(\boldsymbol{x} \cdot \boldsymbol{y}) + c]^d. \tag{8}$$

– Radial basis function (RBF)

$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\gamma||\boldsymbol{x} - \boldsymbol{y}||^2). \tag{9}$$

as well as the linear kernel, in which $K(.,.)$ is simply a scalar product.

4 Experiments

The database consists of a set of 3D SPECT brain images produced with an injected gamma emitting ^{99m}Tc-ECD radio-pharmaceutical and acquired by a three-head gamma camera Picker Prism 3000. Images of the brain cross sections are reconstructed from the projection data using the filtered back-projection (FBP) algorithm in combination with a Butter-worth noise removal filter. The SPECT 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, in a process described in detail in [13, 17]. The images were initially labeled by experienced clinicians of the "Virgen de las Nieves" hospital (Granada, Spain), within two classes: NORMAL and AD. In total, the database consists of 79 patients: 41 NOR and 38 AD.

The size of the images was reduced by a factor $1/n^3$, with n ranging from 2 to 12, in order to reduce the effect of possible defective acquisitions of some brain regions and to simplify the computation of the eigenbrains. In average, only 76 eigenbrains were necessary to explain near the 100% of the variance, with about the 60% retained by the first eigenbrain. These 76 eigenbrains were taken in groups of $M' = 1, 2, ...$ to 76 items to construct the matrix of (4), grouping them according to:

– their covariance eigenvalues.
– selecting those which made the Fisher Discriminant Ratio (FDR) between the $\boldsymbol{\Omega}$ values higher.

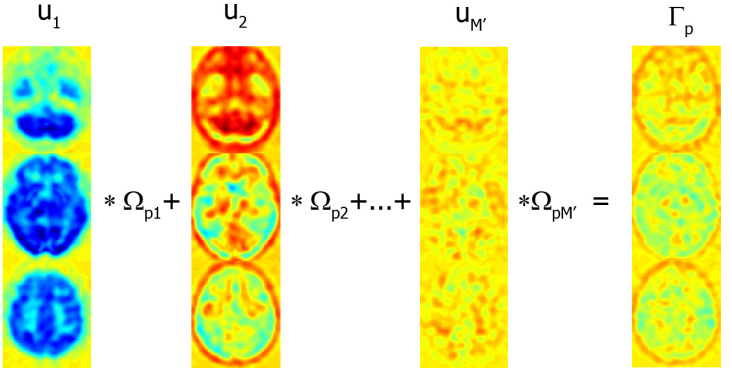

Fig. 2. Eigenbrain representation of three representative transversal slices of the p-th zero-mean brain image. The representation is encoded in the coefficients $(\Omega_{p1}, \Omega_{p2}, ..., \Omega_{pM'})$.

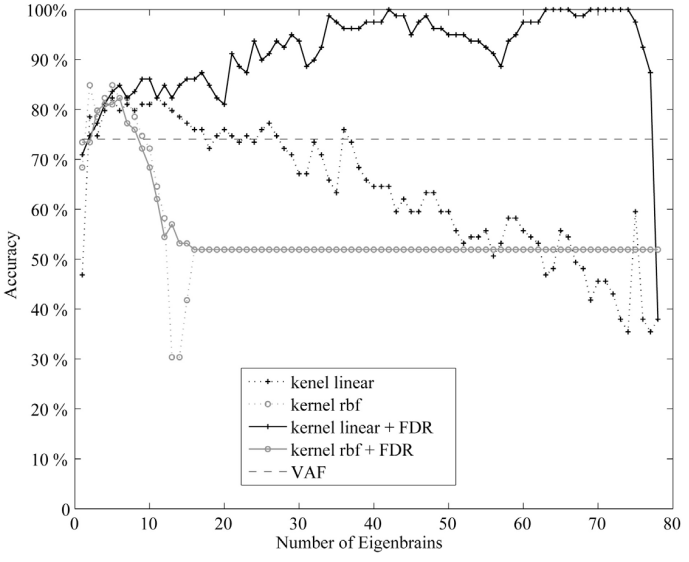

Fig. 3. Two different kernels accuracies versus the number of eigenbrains used to construct the matrix Ω for the two methods of selecting components

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \tag{10}$$

where μ_i and σ_i^2 denote the variance and the mean of the i-th class respectively. Once the matrix $\mathbf{\Omega}$ was obtained, a SVM was trained using 4 different kernels: linear, quadratic, RBF and polynomial, and was tested using a leave-one-out cross-validation strategy.

Results: Linear kernels are expected to perform better than others when the dimension of the feature space increases. If this increase is made with a covariance eigenvalue criterion, the relevant information is contained in the few first eigenbrains, and the rest only contributes with noisy information, as seen from Fig. 3. But enhancing the dimension of the feature space with a FDR criterion shows a fundamental increase of the effectiveness, if compared to ranked covariance eigenvalues which allows to conclude that only a small fraction of the eigenbrains contains noisy information. When a linear kernel SVM was built in combination with a large dimensional feature space, independently of the reducing factor, the method reaches notably the 100% accuray. This outperforms previous results, as the 74% accuracy Voxel-as-Features approach [18] [19].

5 Conclusions

Using statistical classification methods in SPECT images for assisting the AD diagnosis looks promising, but still faces problems to become a useful tool to

physicians. Two main problems are the high dimensionality of the feature space and the small sample size problem. In this work, we presented two solutions to each problem: we made use of the eigenbrain approach, inspired in the eigenface solution to the face detection problem, leading to the idea of collecting a small feature pattern which best describes the patient image characteristics. Secondly, we trained a SVM supervised learning classifier, which allowed us to automatically separate the patient database in normal and affected subjects. Furthermore, a classification of new subjects in a unsupervised manner was possible, without any concrete knowledge about the Alzheimer's disease. A significant advance is seen when estimating the performance of this learning process with a leave-one-out cross-validation test, reaching the 100% accuracy. PCA is then a simple and effective method to deal with the high dimensionality problem in SPECT images with a low computational cost, which shows its effectiveness in combination with a linear kernel SVM and high dimensional feature space, provided that the dimension is enhanced with a FDR criterion.

Acknowledgment

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] Cummings, J.L., Vinters, H.V., Cole, G.M., Khachaturian, Z.S.: Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities. Neurology 51(suppl. 1), S2–S17 (1998)

[2] Ng, S., Villemagne, V.L., Berlangieri, S., Lee1, S.-T., Cherk, M., Gong, S.J., Ackermann, U., Saunder, T., Tochon-Danguy, H., Jones, G., Smith, C., O'Keefe, G., Masters, C.L., Rowe, C.C.: Visual assessment versus quantitative assessment of 11c-pib pet and 18f-fdg pet for detection of alzheimer's disease. Journal of Nuclear Medicine 48, 547–552 (2007)

[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] Yin, T.K., Chiu, N.T.: Discrimination between alzheimer's dementia and controls by automated analysis of statistical parametric maps of 99mTc-HMPAO-SPECT volumes. In: Proceedings of the Fourth IEEE Symposium on Bioinformatics and Bioengineering, pp. 183–190 (2004)

[5] Adler, R.J.: The Geometry of random fields. Wiley, New York (1981)

[6] Scarmeas, N., Habeck, C.G., Zarahn, E., Anderson, K.E., Park, A., Hilton, J., Pelton, G.H., Tabert, M.H., Honig, L.S., Moeller, J.R., Devanand, D.P., Stern, Y.: Covariance pet patterns in early alzheimer's disease and subjects with cognitive impairment but no dementia: utility in group discrimination and correlations with functional performance. NeuroImage 23(1), 35–45 (2004)

[7] Salmon, E., Kerrouche, N., Perani, D., Lekeu, F., Holthoff, V., Beuthien-Baumann, B., Sorbi, S., Lemaire, C., Collette, F., Herholz, K.: On the multivariate nature of brain metabolic impairment in alzheimer's disease. Neurobiology of Aging 30(2), 186–197 (2009)
[8] Nobili, F., Salmaso, D., Morbelli, S., Girtler, N., Piccardo, A., Brugnolo, A., Dessi, B., Larsson, S.A., Rodriguez, G., Pagani, M.: Principal component analysis of fdg pet in amnestic mci. Eur. J. Nucl. Med. Mol. Imaging 35(12), 2191–2202 (2008)
[9] Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of congnitive neuroscience 3(1), 71–86 (1991)
[10] Vapnik, V.N.: Statistical Learning Theory. John Wiley and Sons, Inc., New York (1998)
[11] Ramírez, J., Yélamos, P., Górriz, J.M., Segura, J.C.: SVM-based speech endpoint detection using contextual speech features. Electronics Letters 42(7), 877–879 (2006)
[12] Tao, D., Tang, X., Li, X., Wu, X.: Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(7), 1088–1099 (2006)
[13] Ramírez, J., Górriz, J.M., Gómez-Río, M., Romero, A., Chaves, R., Lassl, A., Rodríguez, A., Puntonet, C.G., Theis, F., Lang, E.: Effective emission tomography image reconstruction algorithms for SPECT data. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2008, Part I. LNCS, vol. 5101, pp. 741–748. Springer, Heidelberg (2008)
[14] Ramírez, J., Górriz, J.M., Romero, A., Lassl, A., Salas-Gonzalez, D., López, M., Alvarez, I., Gómez-Río, M., Rodríguez, A.: Computer aided diagnosis of alzheimer type dementia combining support vector machines and discriminant set of features. In: Information Sciences (2008) (accepted)
[15] Lassl, A., Górriz, J.M., Ramírez, J., Salas-Gonzalez, D., Puntonet, C.G., Lang, E.W.: Clustering approach for the classification of spect images. In: Medical Imaging Conference, Dresden. IEEE, Los Alamitos (2008)
[16] 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. IEEE, Los Alamitos (2008)
[17] Salas-Gonzalez, 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)
[18] 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)
[19] 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)