Functional Brain Image Classification Techniques for Early Alzheimer Disease Diagnosis

J. Ramírez, R. Chaves, J.M. Górriz, I. Álvarez,
M. López, D. Salas-Gonzalez, and F. Segovia

Dept. of Signal Theory, Networking and Communications
University of Granada, Spain

Abstract. Currently, the accurate diagnosis of the Alzheimer disease (AD) still remains a challenge in the clinical practice. As the number of AD patients has increased, its early diagnosis has received more attention for both social and medical reasons. Single photon emission computed tomography (SPECT), measuring the regional cerebral blood flow, enables the diagnosis even before anatomic alterations can be observed by other imaging techniques. However, conventional evaluation of SPECT images often relies on manual reorientation, visual reading and semiquantitative analysis of certain regions of the brain. This paper evaluates different pattern classifiers including k-nearest neighbor (kNN), classification trees, support vector machines and feedforward neural networks in combination with template-based normalized mean square error (NMSE) features of several coronal slices of interest (SOI) for the development of a computer aided diagnosis (CAD) system for improving the early detection of the AD. The proposed system, yielding a 98.7% AD diagnosis accuracy, reports clear improvements over existing techniques such as the voxel-as-features (VAF) which yields just a 78% classification accuracy.

1 Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disease associated with disruption of neuronal function and gradual deterioration in cognition, function, and behavior [1]. It affects approximately 2-4 million individuals in the United States and more than 30 million worldwide. With the growth of the older population in developed nations, the prevalence of AD is expected to triple over the next 50 years. The major goals in treating AD currently are to recognize the disease early in order to initiate appropriate therapy and delay functional and cognitive losses. However, accurate diagnosis of the AD in its early stage still remains a challenge in the clinical practice.

Emission computed tomography techniques producing functional images of the brain, especially single-photon emission computed tomography (SPECT) and positron emission tomography (PET), provide a substantial aid in the diagnosis of the initial dementia and the Alzheimer. Cerebral SPECT, which is based on brain uptake of a technetium 99m-based lipid-soluble radionuclide, is a widely available technique for brain perfusion assessment with a rotating

J. Mira et al. (Eds.): IWINAC 2009, Part II, LNCS 5602, pp. 150–157, 2009.

gamma camera. AD patients typically demonstrate a relative paucity of activity in the temporoparietal regions, compared with the activity in control subjects [2]. However, early detection of AD still remains a challenge since conventional evaluation of SPECT scans often relies on manual reorientation, visual reading and semiquantitative analysis.

This paper evaluates different pattern classifiers for the development of an early AD SPECT-based computer aided diagnosis (CAD) system [3]. The proposed methods combining pattern recognition and advanced feature extraction schemes are developed with the aim of reducing the subjectivity in visual interpretation of SPECT scans by clinicians, thus improving the accuracy of diagnosing Alzheimer disease in its early stage.

2 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 $\mathbf{v}_i$ and class labels y_i:

$$(\mathbf{v}_1, y_1), (\mathbf{v}_2, y_2), ..., (\mathbf{v}_l, y_l) \in (R^N \times \{\text{Normal}, \text{ATD}\}), \tag{1}$$

so that a classifier is produced which maps an unknown object $\mathbf{v}_i$ to its classification label y_i. Several different classifiers are evaluated in this work.

2.1 Support Vector Machines

Support vector machines (SVM) separate binary labeled training data by a hyperplane that 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 so that f will correctly classify new examples $(\mathbf{v}, y)$. 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 $\Phi(\mathbf{v})$, the training algorithm only depends on the data through dot products in such an Euclidean space, i.e. on functions of the form $\Phi(\mathbf{v}_i) \cdot \Phi(\mathbf{v}_j)$. If a "kernel function" K is defined such that $K(\mathbf{v}_i, \mathbf{v}_j) = \Phi(\mathbf{v}_i) \cdot \Phi(\mathbf{v}_j)$, it is not necessary to know the Φ function during the training process. In the test phase, an SVM is used by computing dot products of a given test point $\mathbf{v}$ with $\mathbf{w}$, or more specifically by computing the sign of

$$f(\mathbf{v}) = \sum_{i=1}^{N_S} \alpha_i y_i \Phi(\mathbf{s}_i) \cdot \Phi(\mathbf{v}) + w_0 = \sum_{i=1}^{N_S} \alpha_i y_i K(\mathbf{s}_i, \mathbf{v}) + w_0, \tag{2}$$

where $\mathbf{s}_i$ are the support vectors.

2.2 k-Nearest Neighbor

An object is classified by a majority vote of its neighbors, with the object being assigned to the most common class amongst its k nearest neighbors (kNN). k is a positive integer, typically small. For instance, if $k = 1$, then the object is simply assigned to the class of its nearest neighbor.

2.3 Neural Networks

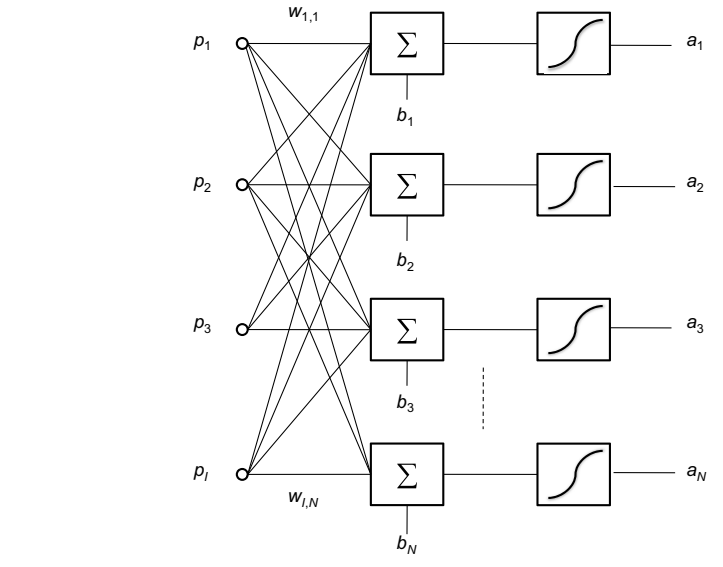
An Artificial Neural Network (ANN) [4,5] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, processes information. ANNs can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges (with weights) are connections 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. Feedforward 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 the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state.

Fig. 1.a) shows a feedforward network consisting of a single-layer network of N logsigmoid neurons having I inputs. 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.b). Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors.

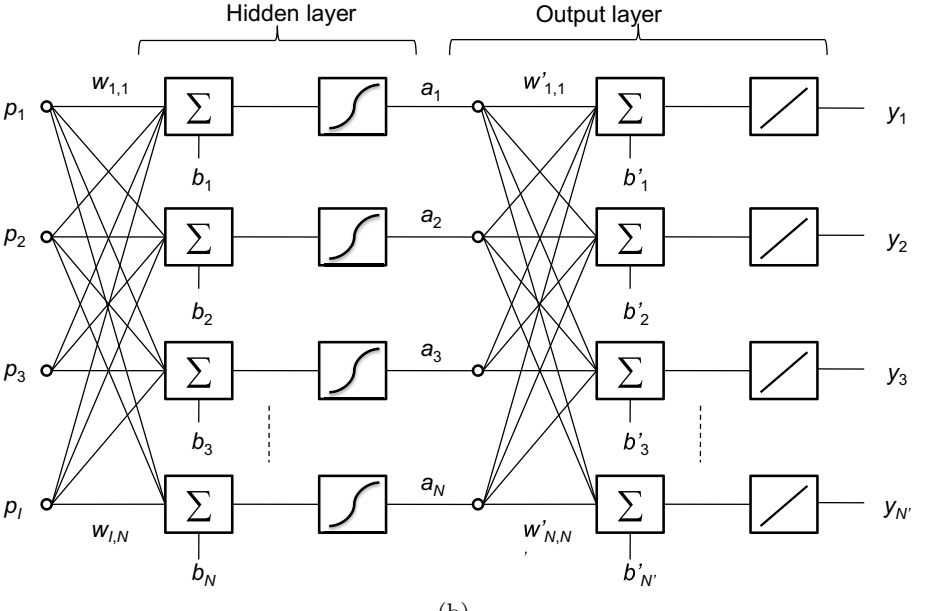
Learning process in the ANN context can be viewed as the problem of updating network architecture and connection weights so that a network can efficiently perform a specific task. The ability of ANNs to automatically learn from examples makes them attractive and exciting. The development of the backpropagation 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 feedforward 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.



(a)



(b)

Fig. 1. Feedforwad neural network architectures. a) A single layer of neurons, b) Hidden layer of neurons plus linear output layer.

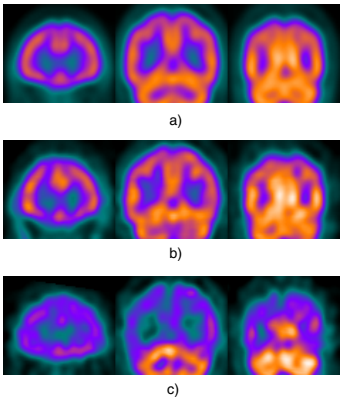


Fig. 2. Coronal slices of: a) Template, b) Normal subject, and c) AD patient

3 Image Acquisition, Preprocessing and Feature Extraction

SPECT scans are registered by means of a three-head gamma camera Picker Prism 3000 after injecting the patient a gamma emitting ^{99m}Tc-ECD radiopharmaceutical. 3D brain perfusion volumes are reconstructed from projection data using the filtered backprojection (FBP) in combination with a Butterworth noise filter. SPECT images are spatially normalized using the SPM software [6] in order to ensure that the voxels in different images refer to the same anatomical positions in the brain [7], giving rise to images of voxel size 69×95×79. Finally, intensity level of the SPECT images is normalized to the maximum intensity as in [8]. The images were initially labeled by experienced clinicians of the Virgen de las Nieves hospital (Granada, Spain), using 4 different labels: normal (NOR) for patients without any symptoms of ATD and possible ATD (ATD-1), probable ATD (ATD-2) and certain ATD (ATD-3) to distinguish between different levels of the presence of typical characteristics for ATD. In total, the database consists of 79 patients: 41 NOR, 20 ATD-1, 14 ATD-2 and 4 ATD-3.

Similarity measures between the functional activity of normal controls and each subject were used as features. First, the expected value of the voxel intensity of the normal subjects was computed by averaging the voxel intensities of all the normal controls in the database. Then, the Normalized Mean Square Error (NMSE) between slices of each subject and the template, and defined for 2-D slices of the volume to be:

$$NMSE = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[f(m,n) - g(m,n)]^2}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}[f(m,n)]^2} \tag{3}$$

where $f(m,n)$ defines the reference template and $g(m,n)$ the voxel intensities of each subject, was computed for coronal, transaxial, and sagittal slices.

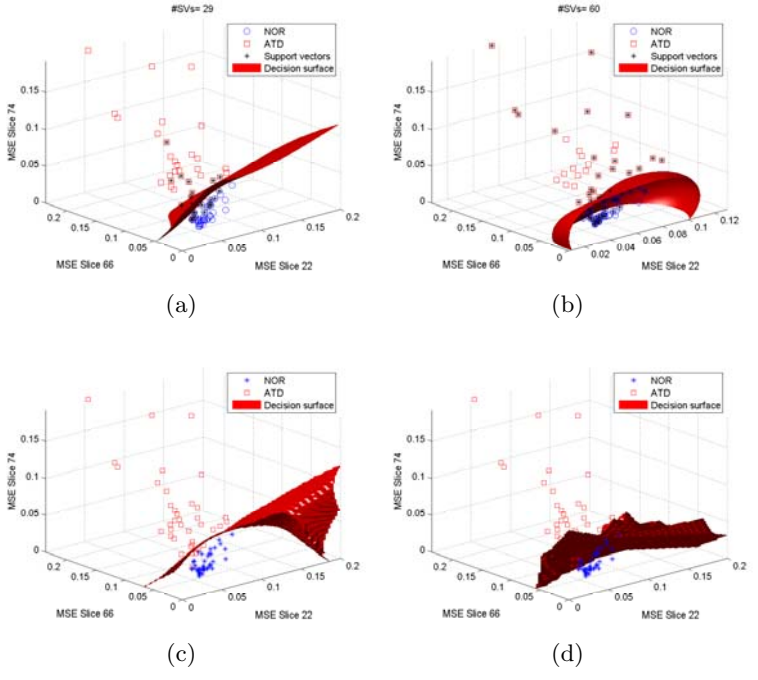


Fig. 3. Decision surfaces for different classifiers: a) SVM (polynomial), b) SVM (RBF), c) Feedforward networks (3 neurons in hidden layer), and d) kNN (k= 3)

A study was carried out in order to identify the most discriminating slices of interest (SOI) by means of a SVM-classified trained on NMSE features of transaxial, coronal and sagittal slices. The analysis showed the high discrimination ability of specific NMSE features of coronal slices. Fig. 2 shows the differences in regional cerebral blood flow (rCBF) provided by these SOI. It shows three different coronal slices of a template brain obtained by averaging the functional SPECT of 41 controls (Fig. 2.a) together with the corresponding coronal slices of a normal subject (Fig. 2.b) and an AD patient (Fig. 2.c). It can be concluded that the rCBF of patients affected by Alzheimer disease is significantly reduced when compared to the expected value for a normal subject. This reduction affects more to some specific cerebral regions. This result is in agreement with many other studies that have shown the temporo-parietal region to be practical for the early detection of the disease in patients that are no longer characterized by specific cognitive impairment but by general cognitive decline.

4 Evaluation Results

First of all, a baseline classifier using voxel-as-features (VAF) [9] was developed for reference. The dimension of the 95×69×79-voxel volume representing the rCBF of each subject was reduced by decimating the original 3D volume and

Table 1. Diagnosis accuracy for the different classifiers evaluated. Comparison to the VAF baseline.

SVM kernel	Lineal	Quadratic	RBF	Polynomial
Baseline VAF	78.5	68.4	51.9	53.2
NMSE (All)	93.7	69.6	35.4	83.5
NMSE (Coronal SOI)	97.5	93.7	94.9	97.5
kNN k =	3	5	7	9
NMSE (Coronal SOI)	96.2	93.7	94.9	93.7
Decision trees				
NMSE (Coronal SOI)	88.6			
Feedforward NN #Neurons in HL	1	3	5	7
	97.5	97.5	98.7	97.5

a SVM-based classifier was trained and tested based on a leave-one-out cross validation strategy. Accuracy of the baseline system was 78.5%. The reference VAF system was compared to different classifiers using the proposed NMSE features of the three most significant slices for AD detection that were shown and discussed in Fig. 2. Fig. 3 shows the 3-D input space defined by means of these three coronal NMSE features and the ability of kNN, SVM, feedforward neural network classifiers to separate the two classes (normal controls in blue *vs.* patients affected by DTA in red) 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. Among the different classification techniques considered, feedforward networks and SVM with almost linear polynomial kernels are the ones that better separate the two classes.

Table 1 shows the accuracy of the proposed and reference VAF systems evaluated by a leave-one-out cross-validation strategy. Results for the system using the NMSE features of all the coronal slices as well as the system using just the three most discriminative ones are included. It can be concluded that the proposed NMSE features using carefully selected coronal slices improves the performance of the system using information of all the brain volume corroborating the evidence that only selected brain areas are mainly affected by hypo-perfusion in patients suffering the Alzheimer disease. The best results are obtained for an almost linear kernel SVM system (97.5%) and a three-neuron in hidden layer feedforward neural network (98.7%), thus outperforming the VAF approach which obtains the best results for linear kernels (78.5%).

5 Conclusions

This paper showed an study for the selection of the optimum classification technique for the development of an AD CAD system. The analysis considered classifiers based on k-nearest neighbor, classification trees, support vector machines and feedforward neural networks. Template-based features defined in terms of the so called Normalized Mean Square error, which measures the difference in

regional cerebral blood flow of several coronal SOI, were found to be very effective for discriminating between AD patients and normal controls. With these and other innovations, the proposed system yielded a 97.5% diagnosis accuracy for almost linear SVM kernels and 98.7% for feedforward neural networks, thus outperforming the 78.5% accuracy of the classical baseline VAF approach .

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. Petrella, J.R., Coleman, R.E., Doraiswamy, P.M.: Neuroimaging and early diagnosis of Alzheimer disease: A look to the future. Radiology 226, 315–336 (2003)
2. Holman, B.L., Johnson, K.A., Gerada, B., Carvaiho, P.A., Sathn, A.: The scintigraphic appearance of Alzheimer's disease: a prospective study using Tc-99m HMPAO SPECT. Journal of Nuclear Medicine 33(2), 181–185 (1992)
3. Ramírez, J., Górriz, J.M., López, M., Salas-Gonzalez, D., Álvarez, I., Segovia, F., Puntonet, C.G.: Early detection of the Alzheimer disease combining feature selection and kernel machines. In: ICONIP 2008 Proceedings. LNCS, Springer, Heidelberg (2008)
4. McCulloch, W.S., Pitts, W.: A logical calculus of ideas immanent in nervous activity. Bull. Mathematical Biophysics 5, 115–133 (1943)
5. Rosenblatt, R.: Principles of Neurodynamics. Spartan Books, New York (1962)
6. 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)
7. 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)
8. Saxena, P., Pavel, D.G., Quintana, J.C., Horwitz, B.: An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimers disease. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 623–630. Springer, Heidelberg (1998)
9. 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)