Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees

Article *in* Physics in Medicine and Biology · April 2010

DOI: 10.1088/0031-9155/55/10/002 · Source: PubMed

CITATIONS
38

READS
72

8 authors, including:

Juan M Gorriz
University of Granada
353 PUBLICATIONS **2,952** CITATIONS
SEE PROFILE

Javier Ramírez
University of Granada
327 PUBLICATIONS **3,724** CITATIONS
SEE PROFILE

Miriam López
University of Granada
78 PUBLICATIONS **1,217** CITATIONS
SEE PROFILE

Carlos G. Puntonet
University of Granada
362 PUBLICATIONS **2,888** CITATIONS
SEE PROFILE

Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees

2010 Phys. Med. Biol. 55 2807

(http://iopscience.iop.org/0031-9155/55/10/002)

IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **55** (2010) 2807–2817 doi:10.1088/0031-9155/55/10/002

Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees

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

[1] Department of Signal Theory Networking and Communications, ETSIIT, 18071, University of Granada, Spain
[2] Department of Computer Architecture and Computer Technology, ETSIIT, 18071, University of Granada, Spain

E-mail: dsalas@ugr.es

Received 29 September 2009, in final form 25 March 2010
Published 22 April 2010
Online at stacks.iop.org/PMB/55/2807

Abstract

This paper presents a computer-aided diagnosis technique for improving the accuracy of early diagnosis of Alzheimer-type dementia. The proposed methodology is based on the selection of voxels which present Welch's *t*-test between both classes, normal and Alzheimer images, greater than a given threshold. The mean and standard deviation of intensity values are calculated for selected voxels. They are chosen as feature vectors for two different classifiers: support vector machines with linear kernel and classification trees. The proposed methodology reaches greater than 95% accuracy in the classification task.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Distinguishing Alzheimer's disease (AD) remains a diagnostic challenge, especially during the early stage of the disease. Furthermore, in this early stage, the disease offers better opportunities to be treated. With the growth of the older population in developed nations, the prevalence of AD is expected to quadruple over the next 50 years (Evans *et al* 1989, Brookmeyer *et al* 1998, 2007).

Single photon emission computed tomography (SPECT) is a widely used technique to study the functional properties of the brain. SPECT brain imaging techniques employ radioisotopes which decay, emitting predominantly a single gamma photon. When the nucleus of a radioisotope disintegrates, a gamma photon is emitted in a random direction which is uniformly distributed in the sphere surrounding the nucleus. If the photon is unimpeded by a collision with electrons or other particles within the body, its trajectory will be a straight line or ray. In order for a photon detector external to the patient to discriminate the direction that

a ray is incident from, a physical collimation is required. Typically, lead collimator plates are placed prior to the the detector's crystal in such a manner that the photons incident from all but a single direction are blocked by the plates. This guarantees that only photons incident from the desired direction will strike the photon detector. SPECT has become an important diagnostic and research tool in nuclear medicine.

In SPECT images, the differences between different brains make the normalization of the images with respect to a reference template necessary. The general affine model, with 12 parameters, is usually chosen as a first step in a normalization algorithm before proceeding with a more complex non-rigid spatial transformation model (Salas-Gonzalez *et al* 2008, Friston *et al* 2007).

Alzheimer's diagnosis is usually based on the information provided by a careful clinical examination, a thorough interview of the patient and relatives and a neuropsychological assessment (Braak and Braak 1997, Cummings *et al* 1998, Hoffman *et al* 2000, Ng *et al* 2007). A SPECT study is frequently used as a complimentary diagnostic tool in addition to the clinical findings (Silverman *et al* 2001, Higdon *et al* 2004). However, in late-onset AD there is minimal perfusion in the mild stages of the disease, and age-related changes, which are frequently seen in healthy aged people, have to be discriminated from minimal disease-specific changes. These minimal changes in the images make visual diagnosis a difficult task that requires experienced experts.

This paper presents a computer-aided diagnosis system for the early detection of Alzheimer-type dementia (ATD) using support vector machines (SVM) and classification trees (CT). SVM is a powerful tool which has recently been the focus of attention for the classification of tomography brain images (Ramírez *et al* 2010, Górriz *et al* 2008, Fung and Stoeckel 2007). In this work, we compare the performance of both classifiers, SVM and CT, using the accuracy rate, sensitivity and specificity.

In the proposed methodology, Welch's *t*-test is performed in each voxel of the brain image for mean normal and mean Alzheimer's disease images. The set of voxels which present greater values than different thresholds are selected. The mean and standard deviation of these selected voxels are used as features of the classifiers.

This work is organized as follows: in section 2 we present an overview of the support vector machines and the classification trees; in section 3, the SPECT image acquisition and image preprocessing steps are explained; we discuss the construction of the feature vector in section 4; in section 5, we summarize the classification performance obtained for linear support vector machine and classification trees and, finally, conclusions are drawn in section 6.

2. Overview of the classifiers

The images we work with belong to two different classes: normal controls and Alzheimer-type dementia (ATD). The goal of the classification task 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:

$$(\mathbf{v}_1, y_1), (\mathbf{v}_2, y_2), \ldots, (\mathbf{v}_l, y_l) \in (R^N \times \{-1, +1\}), \tag{1}$$

so that a classifier is produced which maps an object $\mathbf{v}_i$ to its classification label y_i where y_i is either 1 or -1, indicating the class to which the point $\mathbf{v}_i$ belongs. This classifier will have the ability to classify new test samples.

2.1. Support vector machines with linear kernels

Linear discriminant functions define decision hypersurfaces or hyperplanes in a multidimensional feature space:

$$g(\mathbf{v}) = \mathbf{w}^T\mathbf{v} + w_0 = 0 \tag{2}$$

where $\mathbf{w}$ is the weight vector and w_0 is the threshold. $\mathbf{w}$ is orthogonal to the decision hyperplane. The goal is to find the unknown parameters $w_i, i = 1, \ldots, N$, which define the decision hyperplane (Vapnik 1998).

Let $\mathbf{v}_i$, $i = 1, 2, \ldots, l$, be the feature vectors of the training set. These belong to two different classes, ω_1 or ω_2. If the classes are linearly separable, the objective is to design a hyperplane that classifies correctly all the training vectors. This hyperplane is not unique and it can be estimated maximizing the performance of the classifier, that is, the ability of the classifier to operate satisfactorily with new data. The maximal margin of separation between both classes is a useful design criterion. Since the distance from a point $\mathbf{v}$ to the hyperplane is given by $z = |g(\mathbf{v})|/\|\mathbf{w}\|$, the optimization problem can be reduced to the maximization of the margin $2/\|\mathbf{w}\|$ with constraints by scaling $\mathbf{w}$ and w_0 so that the value of $g(\mathbf{v})$ is +1 for the nearest point in w_1 and -1 for the nearest point in w_2. The constraints are as follows:

$$\mathbf{w}^T\mathbf{v} + w_0 \geqslant 1, \qquad \forall\, \mathbf{v} \in w_1 \tag{3}$$

$$\mathbf{w}^T\mathbf{v} + w_0 \leqslant 1, \qquad \forall\, \mathbf{v} \in w_2, \tag{4}$$

or, equivalently, minimizing the cost function $J(\mathbf{w}) = 1/2\|\mathbf{w}\|^2$ subject to

$$y_i(\mathbf{w}^T\mathbf{v}_i + w_0) \geqslant 1, \qquad i = 1, 2, \ldots, l. \tag{5}$$

Let us relate the mathematical symbols in this section with the actual variables used in this work (see sections 4 and 5 for a more detailed explanation of these variables). In the case in which the mean and standard deviation of selected voxels are chosen as features, vector $\mathbf{v}$ is a two-dimensional ($N = 2$) vector. Therefore, for a given image i the vector of the features is $\mathbf{v}_i = (\mu_i, \sigma_i)$, where μ_i is the mean of intensity values for selected voxels (those voxels with Welch's t-test value greater than a given threshold value ε) in image i. Furthermore, σ_i is the standard deviation of selected voxels. Lastly, l denotes the number of images ($l = 79$).

2.2. *Classification trees*

Classification trees is a non-parametric technique that produces classification of categorical dependent variables (Breiman *et al* 1993). Binary tree-structured classifiers are constructed by repeated splits of subsets of X into two descendant subsets, beginning with X itself, where X denotes the measurement space to be defined as containing all possible measurement vectors. For instance, in this work in the case in which two features are chosen, X is a two-dimensional space such that the first coordinate (mean of selected voxels) ranges over all real values between [0–255]. The second coordinate, standard deviation of selected voxels, is defined as continuously ranging from $[0, +\infty]$ although in practice, in our work it takes values from 10 to 70 (see the axis range in figures 6 and 7).

This process is plotted in figure 1 for a two-class tree. In the figure, X_2 and X_3 are disjoint with $X = X_2 \cup X_3$. Those subsets which are not split, in this case X_2, X_4 and X_5, are called terminal nodes. Each terminal node is designated by a class label. There may be more than one terminal subset with the same class label.

The first problem in tree construction is how to use the learning sample to determine the binary splits of X into smaller pieces. In order to build a classification tree, three questions need to be solved: how to select the splits, when to declare a node terminal or split and how to assign a class to each terminal node. Once a good split of X is found, then a search is made for good splits of each of the two descendant nodes.

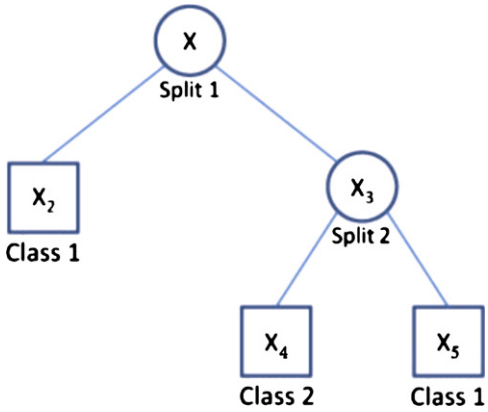


Figure 1. Hypothetical two-class tree.

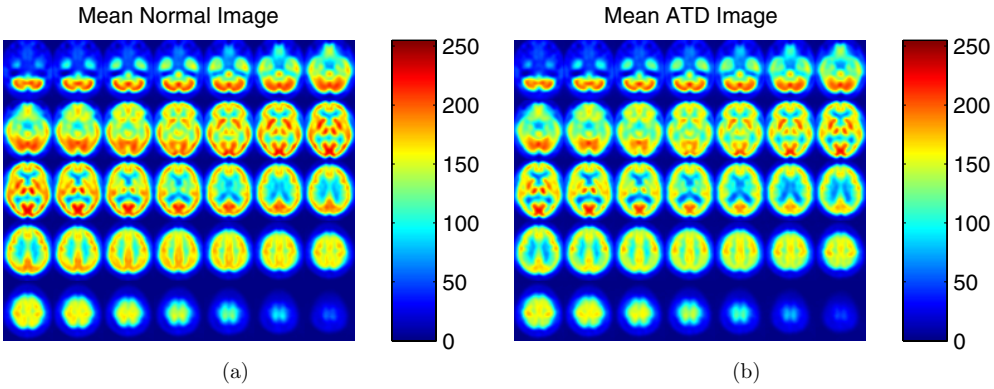
Classification and regression trees (CART) was developed by Leo Breiman (Breiman *et al* 1993). We refer to this book to get a deep insight into CART techniques and their mathematical foundations. The first problem in tree construction is how to use the learning sample to determine the binary splits of X into smaller and smaller pieces. The fundamental idea is to select each split of a subset so that the data in each of the descendant subsets are 'purer' than the data in the parent subset. Therefore, a mathematical measure of 'impurity' needs to be addressed such that the node impurity is largest when all classes are equally mixed together in it, and smallest when the node contains only one class.

Analogous to the last subsection, for the sake of clarity, let us relate the mathematical symbols and terminology in this section with the actual variables used in this work, for instance, figure 7 where the pairwise relationship between the mean and standard deviation of selected voxels for threshold $\varepsilon = 7$ is plotted. This figure suggests a possible classification rule which could be obtained automatically using CART to explain the first split in figure 1: (i) X is the feature mean of selected voxels. (ii) split 1 is the following condition for each image i, $\mu_i < 150$. Let us note that figure 7 shows that all the samples which fulfill this condition are ATD images (denoted as red crosses). Therefore, in this example, after the training procedure, a given test image will be classified as ATD if the mean of selected voxels (with $\varepsilon = 7$) is $\mu_i < 150$. Therefore, 'class 1' under the X_2 box would be ATD. On the other hand, if $\mu_i > 150$, the CART algorithm automatically calculates other classification rule, which is not reported here, denoted as split 2 in figure 1.

3. SPECT image acquisition and preprocessing

The patients were injected with a gamma emitting ^{99m}Tc-ECD radiopharmaceutical and the SPECT raw data were acquired by a three head gamma camera Picker Prism 3000. A total of 180 projections were taken for each patient with a 2° angular resolution. The images of the brain cross sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter (Vandenberghea *et al* 2001, Bruyant 2002, Zhou and Luo 2006, Ramírez *et al* 2008).

The complexity of brain structures and the differences between brains of different subjects make the normalization of the images with respect to a common template necessary. This ensures that the voxels in different images refer to the same anatomical positions in the brain.



(a) (b)

Figure 2. (a) Normal mean SPECT brain image $\bar{I}_{\mathrm{NOR}}$. (b) Alzheimer-type dementia mean SPECT brain image $\bar{I}_{\mathrm{ATD}}$.

In this work, the images have been normalized using a general affine model, with 12 parameters (Hill *et al* 2001, Salas-Gonzalez *et al* 2008, Woods *et al* 1998).

After the affine normalization, the resulting image is registered using a more complex non-rigid spatial transformation model. The deformations are parameterized by a linear combination of the lowest frequency components of the three-dimensional cosine transform bases (Ashburner and Friston 1999). A small-deformation approach is used, and regularization is by the bending energy of the displacement field. Then, we normalize the intensities of the SPECT images with respect to the maximum intensity, which is computed for each image individually by averaging over 3% of the highest voxel intensities, similarly as in Saxena *et al* (1998).

4. Feature selection: Welch's *t*-test

We study the main differences between normal and Alzheimer-type dementia images. Voxels which provide a higher difference between both groups will be considered as training vectors of two classifiers: support vector machines with linear kernel and classification trees. Thus, this statistical study will allow us to select discriminant regions of the brain to establish whether a given SPECT image belongs to a normal or an ATD patient.

Figure 2(a) shows the mean SPECT brain image of normal subjects $\bar{I}_{\mathrm{NOR}}$ and 2(b) the mean ATD image $\bar{I}_{\mathrm{ATD}}$. They are very similar; nevertheless, there are some regions of the mean normal brain which present higher intensity values.

The root-mean-square deviation of the images from their mean for normal controls and ATD images are defined as I_σ^{NOR} and I_σ^{ATD}. They are plotted in figures 3(a) and (b). These figures help us to discriminate those areas of the brain which present lower dispersion between images.

Welch's *t*-test for independent samples is a modification of the *t*-test that does not assume equal population variances. Using the information provided by the mean and standard deviation normal and ATD images, the *t*-test can be calculated for each voxel by

$$I^t = \frac{\bar{I}_{\mathrm{NOR}} - \bar{I}_{\mathrm{ATD}}}{\sqrt{\frac{I^\sigma_{\mathrm{NOR}}}{N_{\mathrm{NOR}}} + \frac{I^\sigma_{\mathrm{ATD}}}{N_{\mathrm{ATD}}}}}, \tag{6}$$

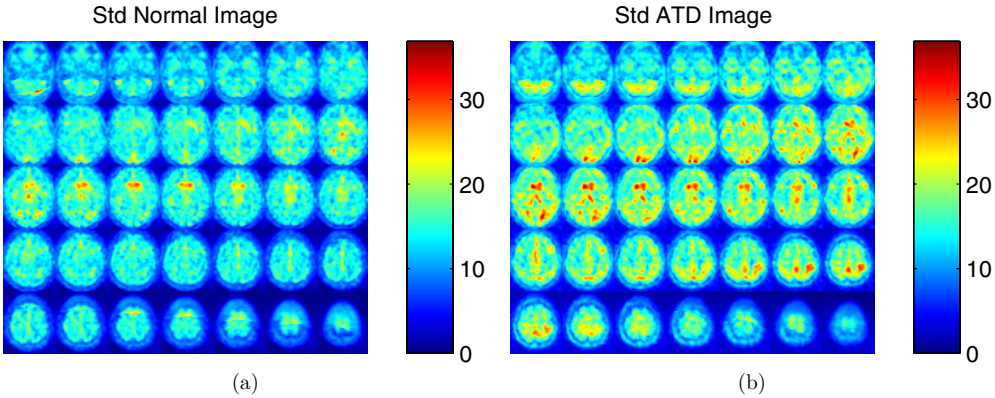


Figure 3. (a) Standard deviation image of normal subjects $I^{\sigma}_{\mathrm{NOR}}$. (b) Standard deviation image of ATD subjects $I^{\sigma}_{\mathrm{ATD}}$.

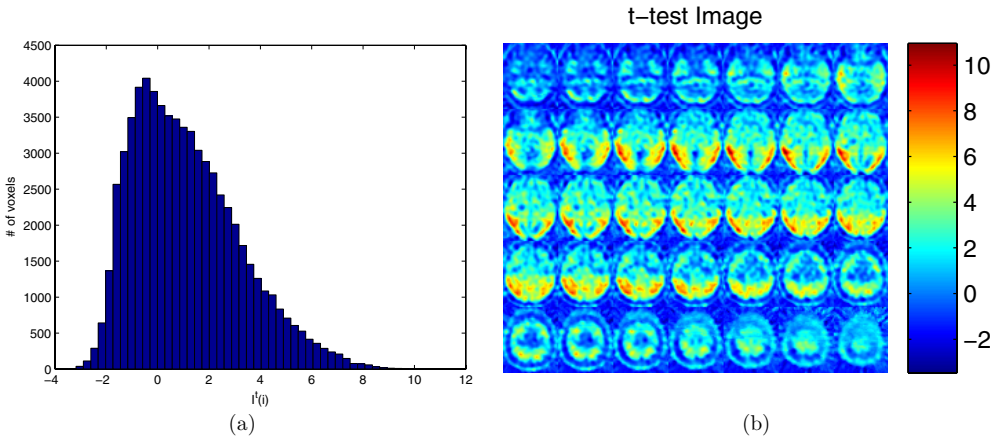


Figure 4. (a) Histogram representing the distribution of the t-values for the image $I^t(i)$. (b) Welch's t-test image I^t.

where I^t denotes the resulting image with the value given by Welch's t-test calculated in each voxel. Figure 4 depicts this image I^t. The histogram with the distribution of I^t is plotted in figure 4(b).

Welch's t-test gives us a value which allows us to measure the voxel intensity difference between the mean normal ($\bar{I}_{\mathrm{NOR}}$) and mean ATD image ($\bar{I}_{\mathrm{ATD}}$). Those voxels which present a t-test value greater than a given threshold will be selected for the classification task.

5. Results

The performance of the classification is tested on a set of 79 real SPECT images (41 normal and 38 ATD) using the leave one-out method: the classifier is trained with all but one image of the database. The remaining image, which is not used to define the classifier, is then categorized. In that way, all SPECT images are classified and the success rate is computed from the number

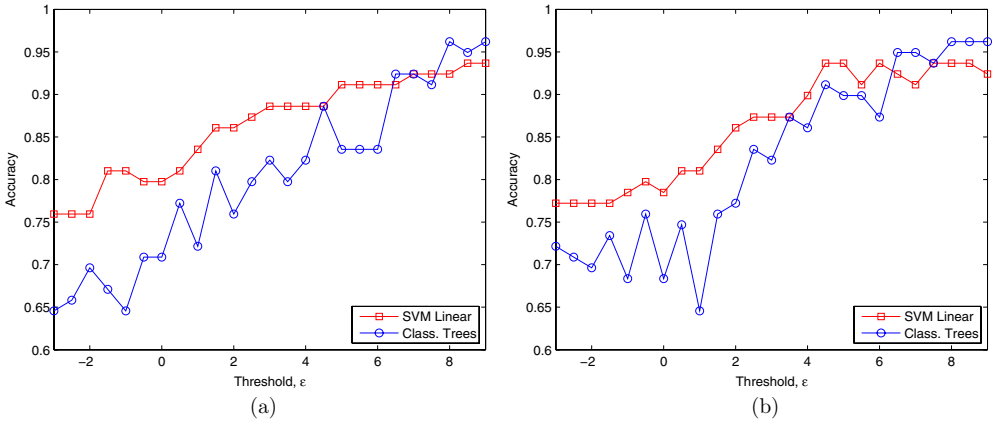


Figure 5. Accuracy rate versus threshold value ε. Feature vector: (a) mean of selected voxels. (b) Mean and standard deviation of selected voxels.

of correctly classified subjects. This process is repeated such that each observation in the sample is used once as the validation data.

First of all, we select those voxels i which fulfill the following condition:

$$\mathrm{i} \,/\, \{I^t(i) > \varepsilon\}, \tag{7}$$

where ε is a threshold. We choose 25 threshold values equally spaced from -3 to 9. The number of selected voxels decreases as ε increases. Increasing ε allows us to select those voxels which present a greater difference between the mean normal and ATD images according to Welch's t-test. Two different features will be used: on the one hand, mean of selected voxels and, on the other hand, a vector with two components: mean and standard deviation of selected voxels.

The classification performance obtained using support vector machines with linear kernel and classification trees versus the threshold value ε is plotted in figure 5. As expected, the classification accuracy is lower for small threshold values ε and increases concomitantly with ε. Using both set of features, it is easily seen that support vector machines present, in general, a higher correct rate than classification trees except for very high values of ε. On the other hand, better results are obtained using mean and standard deviation as features. For instance, this fact is remarked in the case in which $\varepsilon > 4$. The best accuracy rate (96.2%) was obtained when the threshold was set to $\varepsilon > 8$ and classification was performed using classification trees. Other classifiers such as support vector machines with quadratic kernel, radial basis function and Fisher-linear discriminant analysis have also been used for classification purposes. They present very similar performance to the SVM with linear kernels used in this work. These results have not been included in the figure for the sake of clarity.

As expected, higher accuracy rates are obtained for higher values of ε. This is due to the fact that feature vectors separate better both classes (normal and ATD) as the threshold value increases. This fact is shown by means of scatter plots: in order to show the pairwise relationship between the mean and standard deviation of selected voxels for two different values of the threshold ($\varepsilon = 0$ and $\varepsilon = 7$) two matrix plots are drawn (figures 6 and 7, respectively). These figures show that, as it was pointed out before, the selected features are more separated for normal and ADT images when $\varepsilon = 7$ than in the case $\varepsilon = 0$.

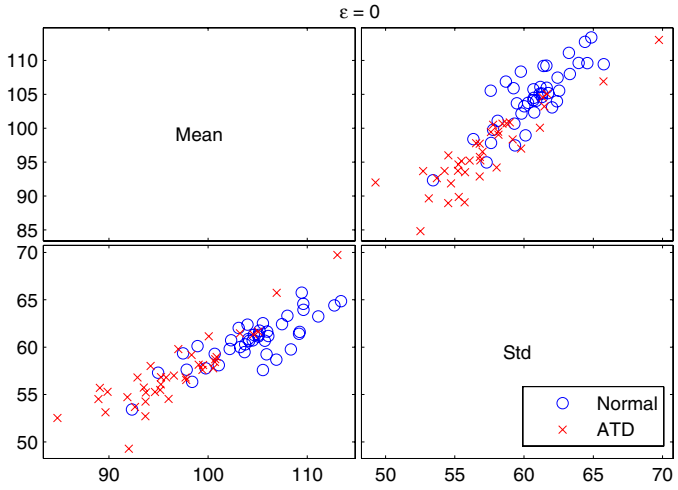


Figure 6. Matrix plot $\varepsilon = 0$.

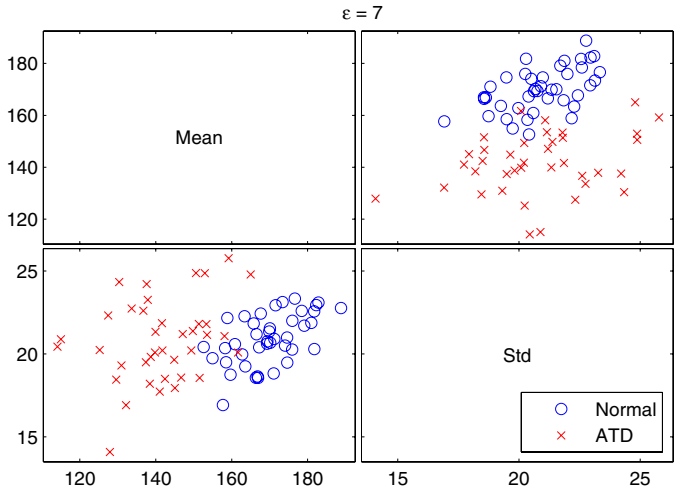


Figure 7. Matrix plot $\varepsilon = 7$.

The accuracy rate does not give us a complete overview of the performance of the classification, hence sensitivity and specificity have been calculated. The sensitivity measures the proportion of actual positives which are correctly identified while the specificity measures give us the proportion of negatives which are correctly identified. Figures 8 and 9 plot the sensitivity versus threshold values. In general, we obtain higher sensitivity using support vector machines than using decision trees. Nevertheless, decision trees achieve higher specificity than SVM for high threshold values.

In this work, we also choose $\varepsilon = 0$ or even voxels with Welch's t-test with $\varepsilon < 0$, only to estimate the performance of the classification task for different threshold values and to highlight the improvement of the accuracy of the proposed methodology concomitantly with the value of ε. This procedure allows us to easily estimate an 'optimal' value of the threshold

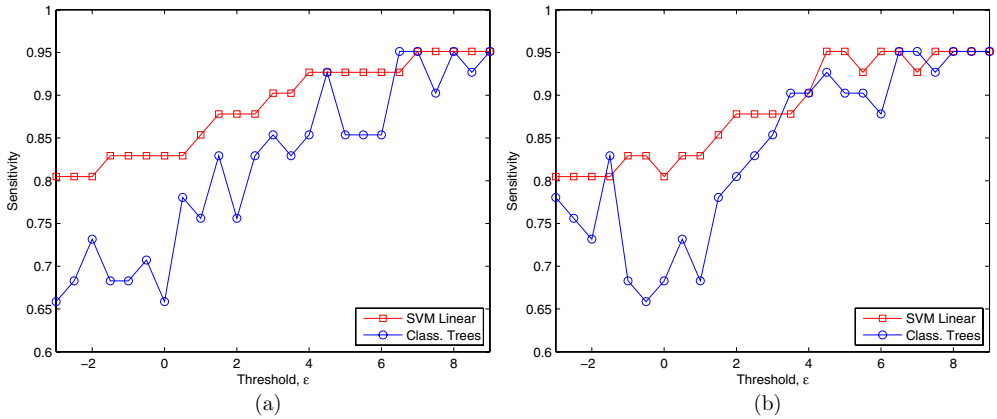


Figure 8. Sensitivity versus threshold value ε. Feature vector: (a) mean of selected voxels. (b) Mean and standard deviation of selected voxels.

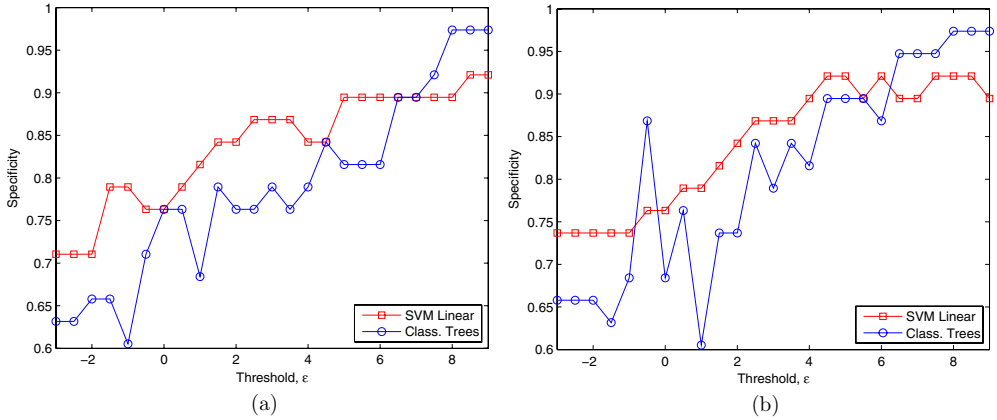


Figure 9. Specificity versus threshold value ε. Feature vector: (a) mean of selected voxels. (b) Mean and standard deviation of selected voxels.

in the sense that we are able to select those voxels with better discrimination ability. Thus, the plot of the accuracy rate versus the threshold enables us to visually select a useful value of the threshold for classification.

In this sense, an alternative could be to calculate an optimal ε value using an analytical procedure. Nevertheless estimating ε studying different threshold values is enough for our practical purposes for different reasons: on the one hand, the estimation of a very precise and specific threshold value is not critical, as the accuracy of the proposed methodology is over 90% for a wide range of values of the threshold for the dataset studied. On the other hand, if we choose to estimate the threshold value using an analytical procedure, the performance of this analytical selection procedure based on a statistical criterion should be checked in terms of accuracy, specificity and sensitivity for different values of the threshold, in order to prove the validity of the threshold estimation method. This is merely the experimental method we use in that work to estimate a useful value of the threshold. Furthermore, we would also like

to point out that the estimation of a useful value of ε, similarly to the training procedure, only needs to be performed once as a 'batch' process.

6. Conclusion

In this work, a computer-aided diagnosis technique for the diagnosis of Alzheimer-type dementia (ATD) is presented. Initially, the images were labeled by experts as normal or Alzheimer's dementia. After normalization of the brain images, we calculate the mean normal and mean ATD image. Welch's *t*-test between both images is calculated in each voxel. The resulting statistic *t* gives us a procedure to ranking voxels. Voxels with *t*-values greater than different threshold values are selected. The mean and standard deviation of selected voxels are used as feature vectors for two different classifiers: a support vector machine with linear kernel and a classification tree. The accuracy, sensitivity and specificity of both classifiers are analyzed using several threshold values. The proposed methodology reaches up to 96.2% accuracy in the classification task. The method proposed in this work allows us to classify the brain images in normal and affected subjects in a parsimonious way, with no prior knowledge about Alzheimer's disease.

Acknowledgments

This work was partly supported by the MICINN of Spain 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 Projects TIC-02566 and TIC-4530. We also thank the reviewers for their careful reading of the manuscript and helpful comments.

References

Ashburner J and Friston K J 1999 Nonlinear spatial normalization using basis functions *Hum. Brain Mapp.* **7** 254–66

Braak H and Braak E 1997 Diagnostic criteria for neuropathologic assessment of Alzheimer's disease *Neurobiol. Aging* **18** S85–8

Breiman L, Friedman J H, Olshen R A and Stone C J 1993 *Classification and Regression Trees* (London: Chapman & Hall)

Brookmeyer R, Gray S and Kawas C 1998 Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset *J. Am. Med. Assoc.* **88** 1337–42

Brookmeyer R, Johnson E, Ziegler-Graham K and Arrighi M 2007 Forecasting the global burden of Alzheimer's disease *Alzheimer's Dementia* **3** 186–91

Bruyant P P 2002 Analytic and iterative reconstruction algorithms in SPECT *J. Nucl. Med.* **43** 1343–58

Cummings J L, Vinters H V, Cole G M and Khachaturian Z S 1998 Alzheimer's disease: etiologies, pathophysiology, cognitive reserve, and treatment opportunities *Neurology* **51** (Suppl. 1) S2–17

Evans D, Funkenstein H, Albert M, Scherr P, Cook N, Chown M, Hebert L, Hennekens C and Taylor J 1989 Prevalence of Alzheimer's disease in a community population of older persons: higher than previously reported *J. Am. Med. Assoc.* **262** 2551–6

Friston K, Ashburner J, Kiebel S, Nichols T and Penny W (ed) 2007 *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (New York: Academic)

Fung G and Stoeckel J 2007 SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information *Knowl. Inf. Syst.* **11** 243–58

Górriz J M, Ramírez J, Lassl A, Salas-Gonzalez D, Lang E W, Puntonet C G, Álvarez I, López M and Gómez-Río M 2008 Automatic computer aided diagnosis tool using component-based SVM *Medical Imaging Conf., Dresden* (Piscataway, NJ: IEEE)

Higdon R *et al* 2004 A comparison of classification methods for differentiating fronto-temporal dementia from Alzheimer's disease using FDG-PET imaging *Stat. Med.* **23** 315–26

Hill D L G, Batchelor P G, Holden M and Hawkes D J 2001 Medical image registration *Phys. Med. Biol.* **46** R1–45

Hoffman J M, Welsh-Bohmer K A and Hanson M 2000 FDG PET Imaging in patients with pathologically verified dementia *J. Nucl. Med.* **41** 1920–8

Ng S *et al* 2007 Visual assessment versus quantitative assessment of ^{11}C-PIB PET and ^{18}F-FDG PET for detection of Alzheimer's disease *J. Nucl. Med.* **48** 547–52

Ramírez J, Górriz J M, Gómez-Río M, Romero A, Chaves R, Lassl A, Rodríguez A, Puntonet C G, Theis F and Lang E 2008 Effective emission tomography image reconstruction algorithms for SPECT data *ICCS 2008, Part I (LNCS* vol 5101*/2008)* (Berlin: Springer) pp 741–8

Ramírez J, Górriz J M, Romero A, Lassl A, Salas-Gonzalez D, López M, Gómez-Río M and Rodríguez A 2010 Computer aided diagnosis of Alzheimer type dementia combining support vector machines and discriminant set of features *Inf. Sci.* at press (http://dx.doi.org/10.1016/j.ins.2009.05.012)

Salas-Gonzalez D, Górriz J M, Ramírez J, Lassl A and Puntonet C G 2008 Improved Gauss–Newton optimization methods in affine registration of SPECT brain images *IET Electron. Lett.* **44** 1291–2

Saxena P, Pavel D G, Quintana J C and Horwitz B 1998 An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimer's disease *Medical Image Computing and Computer-Assisted Intervention—MICCAI (Lecture Notes in Computer Science* vol 1496*)* (Berlin: Springer) pp 623–30

Silverman D H, Small G W and Chang C Y 2001 Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome *J. Am. Med. Assoc.* **286** 2120–7

Vandenberghea S, D'Asselera Y, de Wallea R V, Kauppinenb T, Koolea M, Bouwensa L, Laerec K V, Lemahieua I and Dierckx R 2001 Iterative reconstruction algorithms in nuclear medicine *Comput. Med. Imaging Graph.* **25** 105–11

Vapnik V 1998 *Statistical Learning Theory* (New York: Wiley)

Woods R P, Grafton S T, Holmes C J, Cherry S R and Mazziotta J C 1998 Automated image registration: I. General methods and intrasubject, intramodality validation *J. Comput. Assist. Tomogr.* **22** 139–52

Zhou J and Luo L M 2006 Sequential weighted least squares algorithm for PET image reconstruction *Digit. Signal Process.* **16** 735–45