SVM-based computer-aided diagnosis of the Alzheimer’s disease using t-test NMSE feature selection with feature correlation weighting


Department of Signal Theory, Networking and Communications, University of Granada, Spain

Article info

Article history:
Received 15 May 2009
Received in revised form 16 June 2009
Accepted 17 June 2009

Keywords:
SPECT Brain Imaging Classification
Computer-aided diagnosis
Alzheimer’s disease
Support Vector machine

ABSTRACT

This letter shows a computer-aided diagnosis (CAD) technique for the early detection of the Alzheimer’s disease (AD) based on single photon emission computed tomography (SPECT) image feature selection and a statistical learning theory classifier. The challenge of the curse of dimensionality is addressed by reducing the large dimensionality of the input data and defining normalized mean squared error features over regions of interest (ROI) that are selected by a t-test feature selection with feature correlation weighting. Thus, normalized mean square error (NMSE) features of cubic blocks located in the temporoparietal brain region yields peak accuracy values of 98.3% for almost linear kernel support vector machine (SVM) defined over the 20 most discriminative features extracted. This new method outperformed recent developed methods for early AD diagnosis.

© 2009 Elsevier Ireland Ltd. All rights reserved.

Alzheimer’s disease (AD) is a progressive neurodegenerative disorder first affecting memory functions and then gradually affecting all cognitive functions with behavioral impairments and eventually causing death. According to the latest estimates, the global prevalence of AD will quadruple to 100 million by 2050. To date there is no single test or biomarker that can predict whether a particular person will develop the disease. Its diagnosis is based on the information provided by a careful clinical examination, a thorough interview of the patient and relatives, and a neuropsychological assessment. A regional cerebral blood flow (rCBF) study by means of single photon emission computed tomography (SPECT) is commonly used as a complimentary diagnostic tool in addition to the clinical findings [7]. However, in late-onset AD there are minimal perfusion alterations in the mild stages of the disease, and age-related changes, which are frequently seen in healthy aged people, have to be discriminated from the minimal disease-specific changes. These minimal changes in the images make visual diagnosis a difficult task that requires experienced explorers. Even with this problem still unsolved, the potential of computer-aided diagnosis (CAD) has not been explored in depth [10,13,16,6].

Since their introduction in the late seventies, Support Vector Machines (SVMs) [17] marked the beginning of a new era in the learning from examples paradigm. Recent developments in defining and training statistical classifiers make it possible to build reliable classifiers in very small sample size problems [3] since pattern recognition systems based on SVM circumvent the curse of dimensionality, and even may find nonlinear decision boundaries for small training sets. This paper shows a new feature extraction and selection method for SVM-based classification of SPECT images and the design of an AD CAD system.

The SPECT image acquisition and preprocessing and the statistical methods used for feature extraction and feature selection is presented in this article followed by the experiments that were conducted in order to evaluate the proposed methods and the conclusions. Finally, basic notions of SVM are explained in Appendix A.

The proposed CAD techniques were evaluated by means of a SPECT image database that was especially collected during a concurrent study focussing on early AD diagnosis. Each patient was injected with a gamma emitting technetium-99m labeled ethyl cysteinate dimer (99mTc-ECD) radiopharmaceutical and the SPECT scan was acquired by means of a 3-head gamma camera Picker Prism 3000. Brain perfusion images were reconstructed from projection data by filtered backprojection (FBP) in combination with a Butterworth noise filter. SPECT images were spatially normalized [12] in order to ensure that a given voxel in different images refers to the same anatomical position. This process was done by using Statistical Parametric Mapping (SPM) [5] yielding $69 \times 95 \times 79$ normalized SPECT images. The normalization method assumed a general affine transformation model with 12 parameters and a cost function which presents an extreme value when the template and the image are matched together. Once the image is normalized by means of an affine transformation, it is registered using a more complex non-rigid spatial transformation model. Finally, intensity level is normalized to the

* Corresponding author at: Dpto. Teoría de la Señal, Telemática y Comunicaciones, Periodista Daniel Saucedo Aranda 18071, Granada, Spain. Tel.: +34 958 240842; fax: +34 958240831.
E-mail address: javierrp@ugr.es (J. Ramírez).

0304-3940/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2009.06.052

maximum intensity following a procedure similar to that in [14,9]. The images were initially labeled by experienced clinicians of the Virgen de las Nieves Hospital (Granada, Spain) as normal (NOR) for subjects without any symptoms of the disease and ATD to refer to possible, probable or certain AD patients. In total, the database consists of 79 patients: 41 NOR and 38 ATD.

Not all the brain regions provide the same discriminative power for detecting the early AD. In fact, the posterior cingulate gyri and precunei, as well as the temporo-parietal region are typically affected by hypo-perfusion in the AD [8]. In a previous work, a study was carried out in order to identify the most discriminating slices of interest [SOI] [11]. The analysis showed the high discrimination ability of specific normalized mean square error (NMSE) features corresponding to coronal slices and defined to be:

$$\text{NMSE}_p(x, y, z) = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f(m, n) - g_p(m, n)]^2}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f(m, n)]^2}$$

where $f(m, n)$ defines the reference slice template (obtained by averaging the SPECT images of all the normal controls in the database) and $g_p(m, n)$ the slice intensities of the $p$ subject. Fig. 1 shows the differences in cerebral blood flow (CBF) provided by these SOI [11]. It shows three different coronal slices of a template SPECT obtained by averaging the functional SPECT of 41 controls together with the corresponding coronal slices of a normal subject and an AD patient (Fig. 1c). A SVM classifier was trained using only these SOI NMSE features. Fig. 2 shows the 3D input space and the ability of SVM classifiers to separate the two classes (normal controls in cir-

We aimed to find the most discriminative ROIs defined by the values of $x, y, z$ and the corresponding NMSE feature to train a SVM classifier. It is based on an absolute value two-sample $t$-test defined above clearly identifies the typical discriminative ROI coordinates that were found by using the proposed method and after introducing a weighting factor over $T$ that avoids selecting correlated NMSE features.

SVM classifiers (see Appendix A) were used to define the AD CAD system. The most discriminative 3D NMSE features identified by the proposed feature selection process were used as input data for defining the decision rule. The performance of the SVM classifier was tested on a set of 79 real SPECT images (41 normals and 38 AD). Since leave-one-out cross-validation can suffer from high variability in estimating the classification accuracy, the .632+ bootstrap method [4] combining low variance with only a moderate bias was used for the evaluation of the proposed methods.

The results using localized 3D NMSE and 2D slice NMSE features were compared. The shape of the decision rule strongly depends on the classification method and its associated parameters as well as the input features as shown in Fig. 2. Note that, the linear and polynomial SVM classifiers defined using only the three NMSE features corresponding to the most significant 2D slices are the classifiers that better separate the two classes as shown in Fig. 2a and b. These results are in agreement with the fact that, in applications where the number of instances is lower than the number of features, and when the number of features increases, a mapping of the input data into a high-dimension feature space using more complex kernels is unnecessary [10].

Similar results are obtained by considering 3D cubic NMSE features identified by the feature selection process described in this work. Fig. 5 shows the accuracy of the system as a function of the number of input features for linear, quadratic, RBF and polynomial kernel SVM systems. The accuracy of the system tends to increase with the number of features up to a maximum stable value. SVM classifiers using linear and polynomial kernels defined over the
20 most discriminative input features yield the best results with accuracy values of about 98%, thus outperforming recent developed AD CAD systems based on Principal Component Analysis (PCA) [1] or the voxels-feature (VAF) approach [15] that yields just an 80% classification accuracy by combining the voxel intensities as input features and linear SVM.

An additional experiment was carried out for testing the proposed methods. The selection of the optimum voxel size $v$ was
studied by evaluating the performance of a SVM-based classifier using the .632+ bootstrap method [4]. Fig. 6 shows the accuracy of a linear kernel SVM system as a function of the number of features for different values of the voxel size \( v \). Note that, reduced size cubic NMSE features yield the best results with an accuracy value of about 95\% for \( v = 2 \) that tends to be stable as the number of features increases.

In order to fully evaluate the proposed system, sensitivity and specificity values were also estimated. They provide complimentary information to accuracy and are statistical performance measures of a binary classification test. Sensitivity measures the proportion of actual positives which are correctly identified (e.g. the percentage of AD patients who are identified as having the condition); and the specificity measures the proportion of negatives which are correctly identified (e.g. the percentage of normal subjects who are identified as not suffering AD). The proposed method yielded excellent results for these three measures, yielding peak accuracy value of 98.6\% with a 97.3\% sensitivity of and 100\% specificity. As a conclusion, the proposed NMSE features defined over 3D voxel blocks together with a 97.3\% sensitivity of and 100\% specificity. As a conclusion, the proposed method yielded an accuracy value of up to 98.6\% for a linear kernel SVM classifier defined over the 20 most discriminative 5 × 5 × 5 NMSE features selected. Moreover, the proposed CAD technique outperformed several recently developed methods for early AD diagnosis.

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 Project TIC-02566.

Appendix A. Support vector machines (SVM)

SVM [17,2] separate a set of binary-labeled training data by means of a hyperplane (called maximal margin hyperplane). The objective is to build a decision function \( f : \mathbb{R}^N \rightarrow \{\pm 1\} \) using training data that is, \( N \)-dimensional patterns \( \mathbf{x}_i \) and class labels \( y_i \) so that \( f \) will correctly classify new unseen examples \( (\mathbf{x}, y) \):

\[
(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N) \in \mathbb{R}^N \{\pm 1\}
\]  

(4)

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

\[
g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0
\]  

(5)

where \( \mathbf{w} \) is the weight vector that is orthogonal to the decision hyperplane and \( w_0 \) is the threshold. The optimization task consists of finding the unknown parameters \( w_i \), \( i = 1, \ldots, N \) and \( w_0 \) that define the decision hyperplane. 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 \( \Phi \), the training algorithm only depends on the data through dot products in such an Euclidean space, i.e. on functions of the form \( \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}) \). If a “kernel function” \( K(\mathbf{x}_i, \mathbf{x}_j) \) is defined such that \( K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \), it is not neces-
sary to know the $\Phi$ function during the training process. In the test phase, a SVM is evaluated for each input vector $x$ by computing dot products of the test point $x$ with $w$, or more specifically by computing the sign of:

$$f(x) = \sum_{i=1}^{N_S} \alpha_i \cdot y_i \cdot \Phi(s_i) \cdot \Phi(x) + \omega_0 = \sum_{i=1}^{N_S} \alpha_i \cdot y_i \cdot K(s_i, x) + \omega_0$$

(6)

where $s_i$ are the support vectors, and the coefficients $\alpha_i$ and $\omega_0$ are obtained during a training process.

References


