An ensemble of classifiers guided by the AAL brain atlas for Alzheimer's disease detection

Alexandre Savio*, Manuel Graña

Grupo de Inteligencia Computacional (GIC), Universidad del País Vasco (UPV/EHU), San Sebastián, Spain

Abstract. Detection of Alzheimer's disease based on Magnetic Resonance Imaging (MRI) still is one of the most sought goals in the neuroscientific community. Here, we evaluate a ensemble of classifiers each independently trained with disjoint data extracted from a partition of the brain data volumes performed according to the 116 regions of the Anatomical Automatic Labeling (AAL) brain atlas. Grey-matter probability values from 416 subjects (316 controls and 100 patients) of the OASIS database are estimated, partitioned into AAL regions, and summary statistics *per* region are computed to create the feature sets. Our objective is to discriminate between control subjects and Alzheimer's disease patients. For validation we performed a leave-one-out process. Elementary classifiers are linear Support Vector Machines (SVM) with model parameter estimated by grid search. The ensemble is composed of one SVM per AAL region, and we test 6 different methods to make the collective decision. The best performance achieved with this approach is 83.6% accuracy, 91.0% sensitivity, 81.3% specificity and 0.86 of area under the ROC curve. Most discriminant regions for some of the collective decision methods are also provided.

1 Introduction

Alzheimer's Disease (AD) is one of the most important causes of disability in the elderly and with the increasing proportion of elderly in many populations, the number of dementia patients will rise also. Due to the socioeconomic importance of the disease in occidental countries there is a strong international effort focus in AD. In the early stages of AD brain atrophy may be subtle and spatially distributed over many brain regions, including the entorhinal cortex, the hippocampus, lateral and inferior temporal structures, as well as the anterior and posterior cingulate.

Machine learning methods have become very popular to classify functional or structural brain images to discriminate them into two classes: normal or a specific neurodegenerative disorder [1]. The development of automated detection procedures based in Magnetic Resonance Imaging (MRI) and other medical

* This work has been partially supported by the "Ayudas para la Formación de Personal Investigador" fellowship from the Gobierno del País Vasco.

imaging techniques [2] is of high interest in clinical medicine. It is important to note that these techniques are aimed to help clinicians with more statistical evidence for the diagnosis, it is not intended to substitute any other existing diagnosis procedure.

Most published classification methods working on MRI data train a single classifier. However, it is challenging to train only a global classifier that can be robust enough to achieve good classification performance, mostly due to noise and small sample size of neuroimaging data. Other studies using ensemble of classifiers on brain anatomical MRI can be found in the literature. In [3] they propose a classification method via aggregation of regression algorithms fed with histograms of deformations generated from the Open Access Series of Imaging Studies (OASIS) database obtaining a 0.04 test error rate. Another study shows a local patch-based subspace ensemble method which builds multiple individual classifiers based on different subsets of local patches with the sparse representation-based classifier obtaining an accuracy of 90.8% on the ADNI database [4]. In [5] subsets of ranked features from neuroimaging data are used to in an ensemble of linear Support Vector Machine (SVM) classifiers obtaining 0.94 of Area Under the Receiver Operating Characteritic (ROC) Curve (AUC) when detecting a AD patients vs. control subjects.

In this paper we use modulated Grey-Matter (GM) maps partitioned according to the regions from the Automatic Anatomical Labeling (AAL) atlas to create datasets of statistical features of these GM maps within each of these regions for each subject. These datasets are put into a leave-one-out with grid search process for classifier validation. After that, an ensemble collective decision is made in order to obtain a classification result. We report the results of an ensemble of linear Support Vector Machine (SVM) classifiers.

Section 2 gives a description of the subjects selected for the study, the image processing, feature extraction details, cross-validation and classifier algorithms. Section 3 gives the classification performance results and in section 4 we provide conclusions of this work and further research suggestions.

2 Materials and Methods

In summary the procedure we have followed in this work was: (1) segment the subjects in 3 tissue volume estimation maps, (2) nonlinearly register each subject to the MNI template, (3) calculate the Jacobian determinant of the corresponding displacement fields, with this, (4) modulate the GM partial volume estimation maps. (5) Extract from the GM voxels inside each region of interest (ROI) in the AAL atlas a set of statistical values and finally, (6) use the feature set of each ROI in a leave-one-out classification procedure with parameter grid search. In figure 1 we show a pipeline of the experiment procedure.

The implementation of the feature extraction, classification and result measures have been done in Python with scikit-learn [6]. The source code and preprocessing scripts are freely available for download in

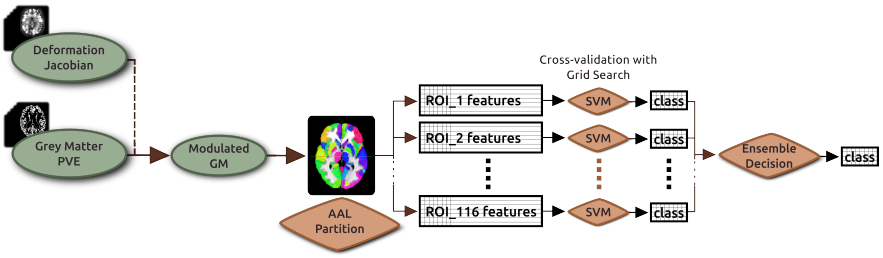

Fig. 1. Flow diagram of the feature extraction and ensemble of classifiers.

http://www.ehu.es/ccwintco/index.php/Usuario:Alexsavio. The data can also be shared through an email to the corresponding author.

2.1 Data

In this study we use all the subjects of a public available brain MRI database, the first Open Access Series of Imaging Studies (OASIS) [7]. These subjects were selected from a larger database of individuals who had participated in MRI studies at Washington University, they were all right-handed and older adults had a recent clinical evaluation. Older subjects with and without dementia were obtained from the longitudinal pool of the Washington University Alzheimer Disease Research Center (ADRC). This release of OASIS consists of a cross-sectional collection of 416 male (119 controls and 41 patients) and female (197 controls and 59 patients) subjects aged 18 to 96 years (218 aged 18 to 59 years and 198 subjects aged 60 to 96 years). Further demographic and image acquisition details can be found in [7].

This database includes at least 3 raw anatomical MP-RAGE images from each subject as well as post-processed images: (a) corrected for inter-scan head movement and rigidly aligned to the Talairach and Tournoux space [8], (b) transformed to a template with a 12-parameter affine registration and merged into a 1-mm isotropic image, (c) skull-stripped and corrected for intensity inhomogeneity and (d) segmented by tissue type. To carry out our experiment we have started with the volumes from (c).

2.2 Preprocessing

The spatial normalization of each subject of the database is performed with the FMRIB Software Library (FSL) FNIRT [9]. A four step registration process with increasing resolution and a scaled conjugate gradient minimization method has been performed using the default parameters, nearest neighbour interpolation and the standard Montreal Nationa Institute (MNI) brain template.

In this registration process the subject S is nonlinearly registered to a template T, and a displacement vector $\overrightarrow{u}(\overrightarrow{r})$ is obtained such that $S(\overrightarrow{r} - \overrightarrow{u})$ corresponds with $T(\overrightarrow{r})$, where $\overrightarrow{r}$ denotes the voxel location. The Jacobian

matrix in this case describes the velocity of the deformation procedure in the neighboring area of each voxel and it is defined by

$$\mathbf{J}_i = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(x-u_x)/\partial y & \partial(x-u_x)/\partial z \\ \partial(y-u_y)/\partial x & \partial(y-u_y)/\partial y & \partial(y-u_y)/\partial z \\ \partial(z-u_z)/\partial x & \partial(z-u_z)/\partial y & \partial(z-u_z)/\partial z \end{pmatrix}. \quad (1)$$

The determinant of the Jacobian matrix is the most commonly used scalar measure of deformation for tensor-based brain morphometry analyses (TBM) [10]. The determinant of the Jacobian matrix $\mathbf{J}_i$ is commonly used to analyze the distortion necessary to deform the images into agreement. A value $\det(\mathbf{J}_i) > 1$ implies that the neighborhood adjacent to the displacement vector in voxel i was stretched to match the template (i.e., local volumetric expansion), while $\det(\mathbf{J}_i) < 1$ is associated with local shrinkage.

Apart, we segment the subjects with FSL FAST [11] into 3 volumes with estimation maps of brain tissues: grey (GM) and white (WM) matter and cerebral-spinal fluid (CSF). In this case we are interested in the GM maps, which we multiply by the Jacobians from the non-linear registration in order to get a modulated GM map in the standard MNI space. Subsequently these maps are smoothed with a 2mm Full-Width Half-Maximum (FWHM) Gaussian filter and used for feature extraction. A visual check has been performed for all images in every processing step carried out in this experiment.

2.3 Feature extraction

The Automatic Anatomical Labeling (AAL) atlas [12] is used to partition the GM maps into 116 brain anatomical regions. In this study we extract of each AAL anatomical region from each subject GM map 7 statistical measures: the maximum voxel value, the minimum, the mean, the variance, the median, the kurtosis and the skewness. Resulting in 116 sub-datasets of 416 subjects with 7 features each. One classifier for each of these, together, perform as an ensemble.

2.4 Support Vector Machines

The Support Vector Machine (SVM) [13] algorithm used for this study is included in the libSVM software package [14][6]. Given training vectors $\mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, l$ of the subject features of the two classes, and a vector $\mathbf{y} \in \mathbb{R}^l$ such that $y_i \in \{-1, 1\}$ labels each subject with its class, in our case, for example, patients were labeled as -1 and control subject as 1. To construct a classifier, the SVM algorithm tries to maximize the classification margin. To this end it solves the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l} \boldsymbol{\xi}_i \quad (2)$$

subject to $y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq (1 - \xi_i),\ \xi_i \geq 0, i = 1, 2, \ldots, n$. The dual optimization problem is

$$\min_{\alpha} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{Q}\boldsymbol{\alpha} - \mathbf{e}^T\boldsymbol{\alpha},$$

subject to $\mathbf{y}^T\boldsymbol{\alpha} = 0,\ 0 \leq \alpha_i \leq C,\ i = 1, \ldots, l$, where $\mathbf{e}$ is the vector of all ones, $C > 0$ is the upper bound on the error, $\mathbf{Q}$ is an $l \times l$ positive semi-definite matrix, $Q_{ij} \equiv y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is the kernel function that describes the behavior of the support vectors. Here, the training vectors $\mathbf{x}_i$ are mapped into a higher (maybe infinite) dimensional space by the function $\phi(\mathbf{x}_i)$. C is a regularization parameter used to balance the model complexity and the training error.

The option we choose to address the problem with the optimization function when unbalanced datasets are present is to adjust a weight for each class with a value inversely proportional to class frequencies. This converts the optimization equation 2 into:

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}^T\mathbf{w} + u_y C \sum_{i=1}^{l} \boldsymbol{\xi}_i, \tag{3}$$

where u_y is the weight value for a given class y.

The chosen kernel function results in different kinds of SVM with different performance levels and the choice of the appropriate kernel for a specific application is a difficult task. In this study we only tested for a linear kernel, another kernel type and its parameter values in the grid search increased exponentially the experiment computation time.

2.5 Cross-validation and parameter grid search

A leave-one-out cross-validation is carried out to calculate the results. In each validation fold, one subject is kept out, a grid search through classifiers parameters is performed against the training set and the kept out subject is then tested against the previously trained classifier with best performance in the grid search. In this grid search we perform a 3-fold cross-validation against the training set using each possible combination of parameter values. The parameter value grid for SVM was C in $[1e-3,\ 1e-2,\ 1e-1,\ 1,\ 1e1,\ 1e2,\ 1e3]$.

2.6 Ensemble decisions

After obtaining the result from all the classifiers in the ensemble we have to make an aggregate decision to have only one class decision. Here we test 6 ensemble majority voting decision criteria using (1) all of the ROI classifiers, the 20 with best average (2) training AUC and (3) training F1-score, (4) an *a priori* set of temporal brain regions and hippocampus only in the left hemisphere known to be affected by AD, (5) another set of brain regions which include temporal and parietal areas of the brain, cingulum, insula and hippocampus from both hemispheres and (6) the same previous set of regions plus frontal lobe regions. In figure 2 we show the regions included in each of these *a priori* ROI maps.

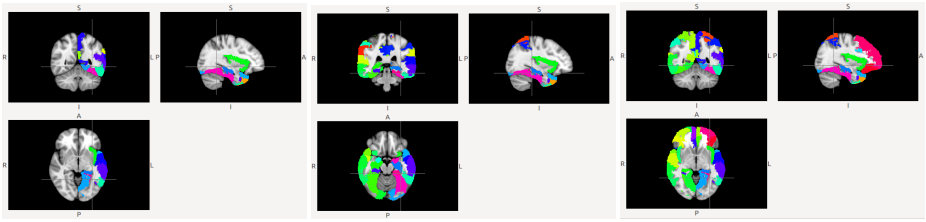

Fig. 2. Slices of the MNI standard template showing the *a priori* maps of affected regions in the left temporal area (left), corresponding to mild AD (middle) and to normal AD (right) used for ensemble decision.

3 Results

In this section we present the linear kernel SVMs performance using the leave-one-out cross-validation. We report accuracy ($(TP+TN)/N$), sensitivity ($TP/(TP+FN)$), specificity ($TN/(FP+TN)$) and area under the ROC curve (AUC) [15] for each ensemble decision. In table 1 we show the classification performance of the ensemble with linear SVMs.

	Accuracy	Sensitivity	Specificity	AUC
Majority Voting	83.6	75.0	86.4	80.7
Best 20 Training AUC	**83.6**	**91.0**	**81.3**	**86.2**
Best 20 Training F1-score	83.2	90.0	81.0	85.5
Left Hem. AD ROIs	82.4	81.0	82.9	81.9
Mild AD ROIs	82.0	74.0	84.5	79.9
Normal AD ROIs	83.2	79.0	84.5	81.7

Table 1. Support Vector Machine accuracy, sensitivity, specificity and ROC area of the leave-one-out cross-validation classification results.

The best classification performance is obtained using the majority voting of the 20 highest training AUC classifiers.

We performed the same experiments using Decision Trees (CART) [16] and Random Forests [17], and also against other deformation measures as the Jacobian determinant and trace, displacement vector magnitude and geodesic anisotropy, more detailed in [18]. Nevertheless we did not obtain in any of them a sensitivity measure $> 70\%$, which led us to discard those results. The parameter grid search values used for CART was criterion Gini or Entropy and maximum depth [None, 10, 20, 30], and for the RF we tested for number of estimators in [3, 5, 10, 30, 50, 100] and maximum number of feature to consider for best split [N, $\sqrt{N}$, $\log_2(N)$, automatic, 1, 3, 5, 7], where N is the number of features.

3.1 Discriminant ROIs

In figure 3 we show the 20 regions that have obtained best training AUC and F1-score during the leave-one-out. Most of the ROIs coincide with the manually selected ROIs, mainly with left hemisphere ROIs and frontal lobes.

Fig. 3. Slices of the MNI standard template where the 20 best training AUC (left) and F1-score (right) ROIs are colored.

4 Conclusions

In this paper we report classification results of an ensemble of SVM classifiers against GM data partitioned with the AAL atlas. The sample is the complete cross-sectional OASIS database. For each subject, the modulated GM data is partitioned into 116 regions and 7 statistical values from each are used as feature vectors. The results are in agreement with most of our previous classification experiments [1,18,19]. Although classification approaches using whole-brain features showed better performance.

It was of our interest to see the performance of the experiments which used *a priori* knowledge ROI maps against those experiments with supervised methods of feature selection. Unsupervised feature selection methods will most probably show worse classification performance than those which are supervised, but lead to more generalized systems and less fitted to the experimental database. We are aware that registration and segmentation errors can lead to biases in the accuracy of the classifiers. In addition, atrophy in brain structures can be interpreted as a late stage of AD and functional MRI could be used instead to detect previous stages of the disease. As future work we will be trying to successfully apply ensembles to deformation features, but we could also find more interesting approaches including functional MRI data.

Acknowledgments We thank the Washington University ADRC for making MRI data available. This work has been supported by the MICINN grant TIN2011-23823.

References

1. D. Chyzhyk, M. Graña, et al. Hybrid dendritic computing with kernel-LICA applied to Alzheimer's disease detection in MRI. *Neurocomputing*, 75(1):72–77, January 2012.
2. C. Davatzikos, Y. Fan, et al. Detection of prodromal alzheimer's disease via pattern classification of MRI. *Neurobiology of aging*, 29(4):514–523, April 2008. PMID: 17174012 PMCID: 2323584.
3. T. Chen, A. Rangarajan, et al. CAVIAR: Classification via aggregated regression and its application in classifying OASIS brain database. *Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro.*, 2010:1337–1340, April 2010.
4. M. Liu, D. Zhang, et al. Ensemble sparse classification of alzheimer's disease. *NeuroImage*, 60(2):1106–1116, April 2012. PMID: 22270352.
5. E. Varol, B. Gaonkar, et al. Feature ranking based nested support vector machine ensemble for medical image classification. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 146 –149, May 2012.
6. F. Pedregosa, G. Varoquaux, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011.
7. D. S. Marcus, T. H. Wang, et al. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, September 2007. PMID: 17714011.
8. J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. Thieme, January 1988.
9. S. M. Smith, M. Jenkinson, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1:S208–219, 2004. PMID: 15501092.
10. N. Lepore, C. Brun, et al. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Transactions on Medical Imaging*, 27(1):129–141, January 2008. PMID: 18270068.
11. Y. Zhang, M. Brady, et al. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57, January 2001. PMID: 11293691.
12. N Tzourio-Mazoyer, B Landeau, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, January 2002. PMID: 11771995.
13. V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, September 1998.
14. Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
15. D. Faraggi and B. Reiser. Estimation of the area under the ROC curve. *Statistics in Medicine*, 21(20):3093–3106, October 2002.
16. J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, March 1986.
17. L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.
18. A. Savio and M. Graña. Deformation based feature selection for computer aided diagnosis of alzheimer's disease. *Expert Systems with Applications*, 40(5):1619–1628, April 2013.
19. A. Savio, M.T. García-Sebastián, et al. Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI. *Computers in Biology and Medicine*, 41(8):600–610, August 2011.