Deformation based features for Alzheimer's disease detection with linear SVM

Alexandre Savio[1], Manuel Graña[1], Jorge Villanúa[2]

[1]Grupo de Inteligencia Computacional, www.ehu.es/ccwintco [2]Osatek, Hospital Donostia Paseo Dr. Beguiristain 109, 20014 San Sebastián, Spain.

Abstract. Detection of Alzheimer's disease over brain Magnetic Resonance Imaging (MRI) data is a priority goal in the Neurosciences. In previous works we have studied the accuracy of feature vectors obtained from VBM studies of the MRI data. In this paper we report results working on deformation based features, obtained from the deformation vectors computed by non-linear registration processes. Feature selection is based on the correlation between the scalar values computed from the deformation maps and the control variable. Results with linear kernel SVM reach accuracies comparable to previous best results.

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder, which is one of the most common cause of dementia in old people. Currently, due to the socioeconomic importance of the disease in occidental countries it is one of the most studied. The diagnosis of AD can be done after the exclusion of other forms of dementia but a definitive diagnosis can only be made after a post-mortem study of the brain tissue. This is one of the reasons why early diagnosis based on Magnetic Resonance Imaging (MRI) is a current research hot topic in Neuroscience. One of the current lines of research involves the application of machine learning algorithms to features extracted from brain MRI. We have already explored the application of various machine learning and computational intelligence algorithms to AD prediction [7,15,1,2]. Specifically we have performed these computational experiments on a subset of the Open Access Series of Imaging Studies (OASIS) database [12]. These works involved the use of Voxel-based Morphometry (VBM) [3] to select the Gray Matter (GM) voxels that would serve as discriminant features. We have made public the set of extracted features in order to allow for independent experimentation upon them [8].

Morphometry analysis has become a common tool for computational brain anatomy studies. It allows a comprehensive measurement of structural differences within a group or across groups, not just in specific structures, but throughout the entire brain. In this paper we use Deformation-based Morphometry (DBM) [10,16] and Tensor-based Morphometry (TBM) [5,11] to guide the feature extraction process. These morphometry methods analyze displacement vectors resulting from non-linear registration procedures with high number of degrees of freedom.

A similar study [14] with 50 subjects obtained 92% of accuracy when discriminating AD subjects from healthy controls using features extracted from displacement fields and different classification methods with SVM, Bayes statistics, and voting feature intervals (VFI). In addition, another study [16] obtained 83% of accuracy using similar approaches to detect subjects with mild cognitive impairment. Although their results can not be reproduced, this work confirms that the approach that we follow is a promising area of research.

In this experiment we obtain scalar measures of the voxel displacements and afterwards compute their correlation with the control variable, which indicates if the sample corresponds to a control subject or an AD patient. The voxel sites with high correlation are selected for the extraction of the feature vector values. We report the results of Support Vector Machine (SVM) with linear kernels performing the classification task.

Section Materials and Methods gives a description of the subjects selected for the study, the image processing, feature extraction details and the classifier system. Section Results gives our classification performance results and section Conclusions gives the conclusions of this work and further research suggestions.

2 Materials and Methods

A database of ninety eight women extracted from the freely available OASIS database were used in this AD detection experiment. The demographic and imaging details of the sample can be found elsewhere [8,15]. The implementation of the SVM used for this study is included in the libSVM (`http://www.csie.ntu.edu.tw/~cjlin/libsvm/`) software package and described in detail in [6]. The feature extraction step requires the data to be spatially normalized. The subjects in the database were already linearly registered to a MNI152 template[12]. Taking into account that, we need to non linearly transform them to a common template in order to obtain the deformation fields which will be used as starting point in the feature extraction process. For this non-linear registration step we could have used again the MNI152 standard template, but the registration algorithm used in this study [4] could not cope well with the large deformations required to register some subjects with enlarged ventricles. For this reason, a custom brain template volume was created with all the subjects in the database. This custom template was subsequently non linearly registered to all the study subjects. As a result from the non linear registration, displacement vectors for each subject are obtained. These displacement vector fields describe the effects of deformation of the template brain to the subject's. For each voxel i, the displacement field for one subject have a vector (x_i, y_i, z_i) representing the ending point of voxel i in the registration process.

Two measures have been extracted from the displacement vectors (see Fig. 1):

1. The displacement vector magnitudes, denoted DM in the results section

$$DM_i = \sqrt{x_i^2 + y_i^2 + z_i^2}, \tag{1}$$

2. The Jacobian determinant of the displacement field gradient matrices, denoted JD in the results section

$$\mathbf{J}_i = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(x-u_x)/\partial y & \partial(x-u_x)/\partial z \\ \partial(y-u_y)/\partial x & \partial(y-u_y)/\partial y & \partial(y-u_y)/\partial z \\ \partial(z-u_z)/\partial x & \partial(z-u_z)/\partial y & \partial(z-u_z)/\partial z \end{pmatrix}. \quad (2)$$

The Jacobian matrix in this case describes the velocity of the deformation procedure in the neighboring area of each voxel. To calculate this matrix, for each voxel site, we used the central difference using two adjacent voxels in one dimension. 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.

Much of the information about the shape change is lost using these measures, nevertheless a more complex multivariate approach would have to be performed in order to use all the information in the deformation gradient matrices or the Green strain tensors defined as $\mathbf{S}_i = \left(\mathbf{J}_i^T \mathbf{J}_i\right)^{1/2}$ [11]. Once the DM and JD maps were calculated, significant voxels were selected from a correlation measure of the voxels to the class labels of each subject. All the registration procedures in this study were performed using ANTS (http://www.picsl.upenn.edu/ANTS).

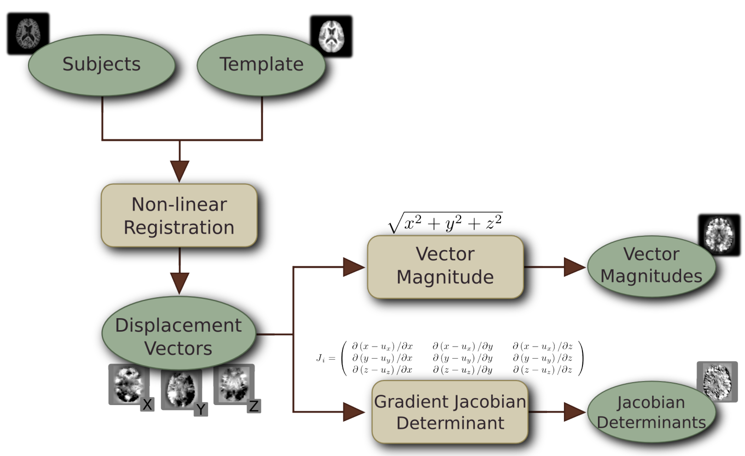

Fig. 1. Pipeline of the image pre-processing steps.

The feature values were extracted from the scalar DM and JD maps computed on the displacement vectors resulting from the registration processes, as described in figure 1. For each voxel site i we extracted one vector $\mathbf{v}_i$ with n components being the value of the voxel i of each one of the n subjects of this experiment. Afterwards correlation measures of these vectors with the control variable, specified by the vector containing the subject class label (-1 for control subject or 1 for patient) were performed. Pearson and Spearman correlation were used in this study. Volume masks containing the voxel sites whose correlation values where above some specified percentile (i.e. 0.990, 0.995) for each combination of map and correlation measure were applied to the corresponding DM or JD map to extract the feature vectors. Figure 2 illustrates the process. Features were normalized before training the SVM on them.

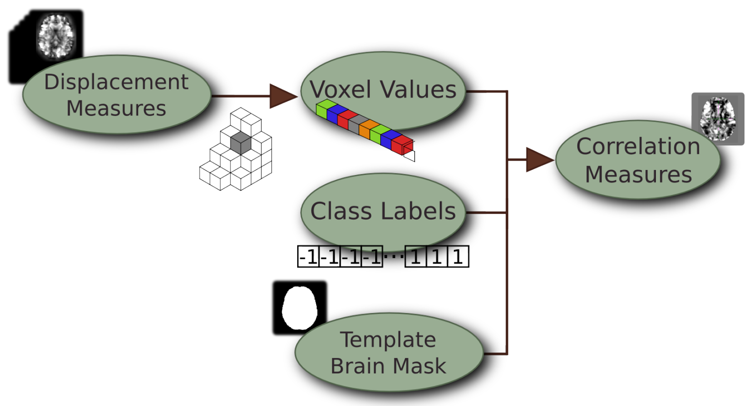

Fig. 2. Pipeline of the calculation of the correlation volumes.

2.1 Pearson's correlation

The Pearson's product-moment correlation coefficient [9] (PMCC or typically denoted by r) is a measure of the correlation (linear dependence) between two variables X and Y, giving a value between $+1$ and -1 inclusive. It is widely used in the sciences as a measure of the strength of linear dependence between two variables. It was developed by Karl Pearson from a similar but slightly different idea introduced by Francis Galton in the 1880s. The correlation coefficient is sometimes called "Pearson's r."

Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \tag{3}$$

where $\bar{X}$ and$\bar{Y}$ are the mean of the variables X and Y respectively.

2.2 Spearman's correlation

The Spearman's correlation coefficient is often thought of as being the Pearson's correlation coefficient between the ranked variables. In practice, however, a simpler procedure is normally used to calculate ρ. The n raw scores X_i, Y_i are converted to ranks x_i, y_i and the differences $d_i = x_i - y_i$ between the ranks of each observation on the two variables are calculated.

If there are no tied ranks, then ρ is given by [13]:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{4}$$

If tied ranks exist, Pearson's correlation coefficient between ranks should be used for the calculation 3. One has to assign the same rank to each of the equal values. It is an average of their positions in the ascending order of the values.

3 Results

Classifiers performance was measured with the architecture strategies mentioned before using a 10-fold cross-validation methodology. In this section the following data for each experiment is presented: the number of features extracted from each subject, classification accuracy, sensitivity, which is related to AD patients and specificity, which is related to control subjects. The results shown are the mean and standard deviation (stdev) values of the classification results from the cross-validation process. In this paper we report the results of the linear kernel SVM. Each of the tables contain the results of using all the selected voxel sites with correlation value above the selected percentile (top row), and the results selecting a fixed number of voxel sites in descending order of correlation value. The objective is to see if the feature vector size reduction based only on the correlation magnitude is an efficient feature selection method.

The results of the features extracted on the basis of the Spearman's correlation are systematically worse the results obtained by the SVM on the features selected from the VBM analysis [15]. These results are in tables 1, 2 and 3. Results performing the feature extraction on the DM map with the 0.995 percentile (table 1) improve over the 0.999 percentile (table 2), reaching values close to the reference values. The application of the process to the JD map (table 3) does not improve the results.

The results of the feature vectors extracted on the basis of the Pearson's correlation improve on the results of the Spearman's correlation selection. These results are in tables 4 and 5. The results on the DM map (table 4) are comparable to the reference results in [15]. Note that the use of the full feature vector give the same result as the reduced vector of 250 voxel sites, suggesting that a strong feature vector size reduction can be achieved. Results on the JD map (table 5) are worse than the DM results.

#Features	Accuracy	Sensitivity	Specificity
12229	0.76 (0.15)	0.77 (0.28)	0.75 (0.17)
2000	0.79 (0.14)	0.82 (0.12)	0.75 (0.20)
1000	0.79 (0.10)	0.87 (0.13)	0.70 (0.16)
500	0.77 (0.13)	0.87 (0.17)	0.67 (0.20)
250	0.79 (0.10)	0.90 (0.13)	0.67 (0.17)

Table 1. Results using linear SVM on DM features obtained from the 0.995 percentile of the Spearman correlation.

#Features	Accuracy	Sensitivity	Specificity
1861	0.66 (0.14)	0.70 (0.20)	0.62 (0.21)
1000	0.66 (0.14)	0.77 (0.18)	0.55 (0.20)
500	0.71 (0.12)	0.85 (0.17)	0.57 (0.17)
250	0.72 (0.15)	0.80 (0.11)	0.65 (0.29)

Table 2. Results using linear SVM on DM features obtained from the 0.999 percentile of the Spearman correlation.

#Features	Accuracy	Sensitivity	Specificity
17982	0.76(0.14)	0.77(0.27)	0.75(0.16)
2000	0.65(0.15)	0.65(0.21)	0.65(0.24)
1000	0.60(0.16)	0.62(0.27)	0.57(0.31)
500	0.58(0.08)	0.70(0.20)	0.47(0.18)
250	0.61(0.09)	0.65(0.21)	0.57(0.20)

Table 3. Results using linear SVM on normalized JD features obtained from the 0.995 percentile of the Spearman correlation measures.

#Features	Accuracy	Sensitivity	Specificity
27474	0.84 (0.10)	0.90 (0.17)	0.77 (0.14)
2000	0.79 (0.12)	0.85 (0.17)	0.72 (0.08)
1000	0.79 (0.10)	0.90 (0.13)	0.67 (0.17)
500	0.79 (0.13)	0.85 (0.21)	0.72 (0.14)
250	0.84 (0.10)	0.92 (0.12)	0.75 (0.17)

Table 4. Results using linear SVM on DM features over the 0.995 percentile of the Pearson correlation measures.

#Features	Accuracy	Sensitivity	Specificity
43967	0.66 (0.19)	0.70 (0.20)	0.62 (0.24)
2000	0.75 (0.13)	0.75 (0.26)	0.75 (0.17)
1000	0.69 (0.15)	0.77 (0.14)	0.60 (0.24)
500	0.66 (0.16)	0.72 (0.30)	0.60 (0.17)
250	0.66 (0.17)	0.70 (0.20)	0.62 (0.24)

Table 5. Results using linear SVM on JD features over the 0.990 percentile of the Pearson correlation measures.

4 Conclusions

In this paper we report classification results on the application of a feature extraction process based on the deformation vectors obtained from non-linear registration processes. The sample is a subset of the OASIS database carefully selected to be pairwise comparable. From the displacement vectors we computed two scalar measures: the magnitude of the displacement vector and the Jacobian determinant of the displacement gradient matrix. We compute the Spearman's and Pearson's correlations of the voxel values with the control variable, selecting voxel sites with the higher correlation. Results show that the deformation vector magnitude features provide better classification accuracy reaching the values of the reference results. Results for the Pearson features are better than for the Spearman features. Although we did not use any combination of classifiers in this study [17], it is important to note how the fusion of information from different images was used to extract relevant features for the classification task. This process of fusion of images of different modalities can be found in a wide variety of medical imaging studies. We are working on the application of non-linear SVM, using RBF kernels, and extending the experimental exploration to other classifiers and combinations.

Acknowledgments

We thank the Washington University ADRC for making MRI data available.

References

1. C. Hernández M. Graña J. Villanúa A. Savio, M. García-Sebastián. Classification results of artificial neural networks for Alzheimer's disease detection. *Intelligent Data Engineering and Automated Learning- IDEAL 2009, Emilio Corchado, Hujun Yin (eds) LNCS 5788*, pages 641–648, 2009.
2. M. Graña J. Villanúa A. Savio, M. García-Sebastián. Results of an Adaboost approach on Alzheimer's disease detection on MRI. *Bioinspired applications in Artificial and Natural Computation. J. Mira, J. M. Ferrández, J.R. Alvarez, F. dela Paz, F.J. Tolede (Eds.) LNCS 5602*, pages 114–123, 2009.
3. J. Ashburner and K. J. Friston. Voxel-Based Morphometry: The Methods. *Neuroimage*, 11(6):805–821, 2000.
4. B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with Cross-Correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, February 2008. PMID: 17659998 PMCID: 2276735.

5. Matias Bossa, Ernesto Zacur, and Salvador Olmos. Tensor-based morphometry with stationary velocity field diffeomorphic registration: Application to ADNI. *NeuroImage*, 51(3):956–969, July 2010.
6. 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`.
7. D. Chyzhyk, M. Graña, A. Savio, and J. Maiora. Hybrid Dendritic Computing with Kernel-LICA applied to Alzheimer's disease detection in MRI. *Neurocomputing*, accepted, 2011.
8. D. Chyzhyk and A. Savio. Feature extraction from structural MRI images based on VBM: data from OASIS database. Technical Report GIC-UPV-EHU-RR-2010-10-14, Grupo de Inteligencia Computacional UPV/EHU, 2010.
9. Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic, 2 edition, January 1988.
10. E. Gerardin, G. Chetelat, M. Chupin, R. Cuingnet, B. Desgranges, Ho-Sung Kim, M. Niethammer, B. Dubois, S. Lehericy, L. Garnero, F. Eustache, and O. Colliot. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47(4):1476–1486, October 2009.
11. N Lepore, C Brun, Y Y Chou, M C Chiang, R A Dutton, K M Hayashi, E Luders, O L Lopez, H J Aizenstein, A W Toga, J T Becker, and P M Thompson. 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.
12. D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner. 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.
13. J.S. Maritz. *Distribution-Free Statistical Methods, Second Edition*. Chapman and Hall/CRC, 2 edition, April 1995.
14. C. Plant, S. J. Teipel, A. Oswald, C. Böhm, T. Meindl, J. Mourao-Miranda, A. W. Bokde, H. Hampel, and M. Ewers. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *NeuroImage*, 50(1):162–174, March 2010.
15. A. Savio, M. García-Sebastián, D. Chyzhyk, C. Hernández, M. Graña, A. Sistiaga, A. Lopez de Munain, and J. Villanúa. Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI. *Computers in Biology and Medicine*, accepted with revisions, 2011.
16. S. J. Teipel, C. Born, M. Ewers, A.L.W. Bokde, M. F. Reiser, H-J. Möller, and H. Hampel. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *NeuroImage*, 38(1):13–24, October 2007.
17. Michal Wozniak and Marcin Zmyslony. Designing fusers on the basis of discriminants: Evolutionary and neural methods of training. In Manuel Graña Romay, Emilio Corchado, and M. Garcia Sebastian, editors, *Hybrid Artificial Intelligence Systems*, volume 6076 of *Lecture Notes in Computer Science*, pages 590–597. Springer Berlin / Heidelberg, 2010. 10.1007/978-3-642-13769-3_72.