Classification Results of Artificial Neural Networks for Alzheimer's Disease Detection

Alexandre Savio[1], Maite García-Sebastián[1], Carmen Hernández[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 on brain Magnetic Resonance Imaging (MRI) is a highly sought goal in the Neurosciences. We used four different models of Artificial Neural Networks (ANN): Backpropagation (BP), Radial Basis Networks (RBF), Learning Vector Quantization Networks (LVQ) and Probabilistic Neural Networks (PNN) to perform classification of patients of mild Alzheimer's disease vs. control subjects. Features are extracted from the brain volume data using Voxel-based Morphometry (VBM) detection clusters. The voxel location detection clusters given by the VBM were applied to select the voxel values upon which the classification features were computed. We have evaluated feature vectors computed from the GM segmentation volumes using the VBM clusters as voxel selection masks. The study has been performed on MRI volumes of 98 females, after careful demographic selection from the Open Access Series of Imaging Studies (OASIS) database, which is a large number of subjects compared to current reported studies.

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 the neurosciences.

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. Voxel-based Morphometry (VBM) is a computational approach to neuroanatomy that measures differences in local concentrations of brain tissue through a voxel-wise comparison of multiple brain images [2]. For instance, VBM

⋆ Research partially supported by Saiotek research projects BRAINER and S-PR07UN02, and the MEC research project DPI2006-15346-C03-03.

has been applied to study volumetric atrophy of the grey matter (GM) in areas of neocortex of AD patients vs. control subjects [3,18,7]. The procedure involves the spatial normalization of subject images into a standard space, segmentation of tissue classes using *a priori* probability maps, smoothing to reduce noise and small variations, and voxel-wise statistical tests. Statistical analysis is based on the General Linear Model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability. Classical statistical inference is used to test hypotheses that are expressed in terms of GLM estimated regression parameters. This computation of given contrast provides a Statistical Parametric Map (SPM), which is thresholded according to the Random Field theory.

Machine learning methods have become very popular to classify functional or structural brain images to discriminate them into normal or a specific neurodegenerative disorder. The Artificial Neural Networks (ANN) used for this study were the Feedforward Networks (sometimes called Multilayer Perceptron) trained with the Backpropagation of errors algorithm (BP), Radial Basis Networks (RBF), Learning Vector Quantization (LVQ) and Probabilistic Neural Networks (PNN) [10]. Support Vector Machine (SVM) both with linear [12] and non-linear [14] kernels have been tested in a previous work for the same task [8]. There are different ways to extract features from MRI for classification: based on morphometric methods [5], based on regions of interest (ROI) [15,14] or GM voxels in automated segmentation images [12]. Our approach is to use the VBM detected clusters as a mask on the Grey Matter (GM) segmentation images to select the potentially most discriminating voxels. Feature vectors for classification are either the voxel values or some summary statistics of each cluster. We considered the feature vector computed from all the VBM clusters together.

A work using ANNs and VBM for AD detection have been reported in [11], where a single three-layer, feed-forward ANN trained with a backpropagation algorithm was used as a classifier over a small set of unpublished proprietary MRI data. They perform data dimensionality reduction applying a Principal Component Analysis (PCA) to improve the efficiency of the classifier. Although their results can not be reproduced, this work confirms that the approach that we follow is a promising area of research.

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.

Materials and Methods

Subjects Ninety eight right-handed women (aged 65-96 yr) were selected from the Open Access Series of Imaging Studies (OASIS) database [16]. OASIS data set has a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease. We have ruled out a set of 200 subjects whose demographic, clinical or derived anatomic volumes information was incomplete. For the present study there are 49 subjects

who have been diagnosed with very mild to mild AD and 49 non-demented. A summary of subject demographics and dementia status is shown in table 1.

	Very mild to mild AD	Normal
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5 / 1 / 2)	31 / 17 / 1	0
MMSE	24 (15-30)	28.96 (26-30)

Table 1. Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best).

Imaging protocol Multiple (three or four) high-resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired [6] on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session. Image parameters: TR= 9.7 msec., TE= 4.0 msec., Flip angle= 10, TI= 20 msec., TD= 200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256×256 (1×1mm).

Image processing and VBM We have used the average MRI volume for each subject, provided in the OASIS data set. These images are already registered and re-sampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [16]. The Statistical Parametric Mapping software (SPM5) [1] was used to compute the VBM which gives us the spatial mask to obtain the classification features. Images were reoriented into a right-handed coordinate system to work with SPM5. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the Full-Width at Half-Maximum (FWHM) of the Gaussian kernel to 10mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data is modulated, but it is close enough for the mask to be reasonable. We designed the statistical analysis as a Two-sample t-test in which the first group corresponds with AD subjects. In SPM software terms: the contrast has been set to [-1 1], a right-tailed (groupN > groupAD), correction FWE, p-value=0.05. The

VBM detected clusters are used for the feature extraction for the classification procedures.

Backpropagation Backward propagation of errors or Backpropagation (BP) [17,10,9] is a supervised learning method, and it is a non-linear generalization of the squared error gradient descent learning rule for updating the weights of the artificial neurons in a single-layer perceptron, generalized to feed-forward networks . Backpropagation requires that the activation function used by the artificial neurons (or "nodes") is differentiable with its derivative being a simple function of itself. The backpropagation of the error allows to compute the gradient of the error function relative to the hidden units. It is analytically derived using the chain rule of calculus. During on-line learning the weights of the network are updated at each input data item presentation. We have used the resilient backpropagation, which uses only the derivative sign to perform the weight updating.

Radial Basis Function Networks Radial basis function networks (RBF) [4] are ANN that use radial basis functions as activation functions. RBF's consist of a two layer neural network, where each hidden unit implements a radial activated function. The output units compute a weighted sum of hidden unit outputs. Training consists of the unsupervised training of the hidden units followed by the supervised training of the output units' weights.

Probabilistic Neural Networks A Probabilistic Neural Network (PNN) [20] is a special type of neural network that uses a kernel-based approximation to form an estimate of the probability density function of categories in a classification problem. The distance is computed from the point being evaluated to each of the other points, and a *radial basis function* (RBF) is applied to the distance to compute the weight (influence) for each point.

Different types of radial basis functions could be used, but the most common is the Gaussian function. The sigma value of the function determines the *spread* of the RBF function; that is, how quickly the function declines as the distance increased from the point. With larger sigma values the function has more spread, so that distant points have a greater influence. PNN are a kind of Nearest Neighbor classifier that uses all the data samples as reference values and the only functional transformation is the computation of the posterior probability of the classes as a combination (sum/average) of the evidence given by each data sample through its RBF window.

The tuning of a PNN network depends on selecting the optimal sigma value of the spread of the RBF functions. In this paper an exhaustive search for the optimal spread value in the range (0, 1) for each training set has been done. The results shown in Table 4 correspond to the best spread value found.

Learning Vector Quantization Learning vector quantization (LVQ) [13,19] provides a method for training competitive layers in a supervised manner. The

system is composed of an unsupervisedly trained competitive layer which performs a partitioning of the input space. The supervisedly trained output layer provides the labeling of the input data according to its belonging to an input region (crisp clustering) or to its degree of membership (soft clustering). In the original proposition of the LVQ, the competitive units were cluster centers with the Euclidean distance as the similitude measure. Training of the competitive units can be performed by Kohonen's Self Organizing Map. Supervised training was simply the assignment of a label to a competitive unit according to a majority voting on the data samples falling in the partition corresponding to the unit. LVQ provides fine tuning of the competitive units using class information. The basic versions proposed by Kohonen are known as the LVQ1 and LVQ2. Both start with the unsupervised learning of the competitive units, and its initial majority voting labeling. In the LVQ1 a supervised training is performed as follows: for each data sample we compare its label with the one of its corresponding competitive unit, if the labels match (the data item is correctly classified) then the competitive unit is moved towards the input data sample, otherwise it is moved in the opposite direction. This rule may cause an unstable and oscillatory behavior if the discriminant boundary among classes is very complex. The LVQ2 rule is proposed to improve the learning, sometimes it is recommended to apply it *after* the LVQ1. In LVQ2, for each input data sample we find the two closest competitive units. If one correctly classifies the input and the other belongs to a wrong class, and the input data lies in a window around the mid-plane between them, then the correct class unit is moved towards the input and the incorrect unit is moved away from the input. We have used the simplest implementations.

Feature extraction We have tested two different feature vector extraction processes, based on the voxel location clusters detection obtained from the VBM analysis. The features were extracted from the output volumes of the segmentation step in the VBM analysis, they are a GM density volume for each subject.

1. The first feature extraction process computes the mean and standard deviation of the GM voxel values of each voxel location cluster, we denote these features as MSD in the result tables given below.
2. The second feature extraction process computes a very high dimensional vector with all the GM segmentation values for the voxel locations included in each VBM detected cluster. The voxel values were ordered in this feature vector according to the coordinate lexicographical ordering. We denote these features as VV in the result tables below.

Results

We evaluated the performance of the classifiers built with the diverse training and architecture strategies using 10 times the 10-fold cross-validation methodology. In this section we present for each experiment the following data: 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 values of the classification results from the 10-fold crossvalidation process, also the standard deviation (stdev) is shown. We will give results of each different classifiers: Backpropagation (Table 2), RBF (Table 3), PNN (Table 4), LVQ1 (Table 5) and LVQ2 (6).

The best accuracy result (Table 6) is 83% with the LVQ2, but this result is not far from the results of LVQ1 and PNN. Which is a very encouraging result, given that we have not removed critical subjects from the data collection: very mildly demented subjects who could end in a false positive diagnosis and . Regarding the usefulness of the features extracted, it is difficult to make an assessment, because some algorithms work better with VV than with MSD, and other have the inverse performance. Training and validation on MSD features is obviously more time efficient, and the best result corresponds to this feature extraction process.

Feature extracted	#Features	#Hidden units	%Accuracy	Sensitivity	Specificity
MSD	24	10	78.0 (0.12)	0.69 (0.14)	0.88 (0.13)
VV	3611	10	78.0 (0.11)	0.72 (0.17)	0.84 (0.18)

Table 2. Classification results with a BP network with resilient backpropagation. Mean (Standard deviation) of 10 cross-validations.

Feature extracted	#Features	Spread	%Accuracy	Sensitivity	Specificity
MSD	24	0.02	66.00 (0.13)	0.65 (0.24)	0.68 (0.14)
VV	3611	0.852	72.5 (0.10)	0.65 (0.21)	0.80 (0.17)

Table 3. Classification results with a RBF network. Mean (Standard deviation) of 10 cross-validations.

Feature extracted	#Features	Spread	%Accuracy	Sensitivity	Specificity
MSD	24	0.02;	77.8 (0.09)	0.62 (0.14)	0.94 (0.1)
VV	3611	0.852	74.2 (0.14)	0.68 (0.20)	0.81 (0.17)

Table 4. Classification results with a PNN network. Mean (Standard deviation) of 10 cross-validations.

Conclusions

In this work we have studied several ANN classifiers applied to classify MRI volumes of AD patients and normal subjects. The feature extraction processes

Feature extracted	#Features	#Hidden units	%Accuracy	Sensitivity	Specificity
MSD	24	10	81.0 (0.18)	0.72 (0.27)	0.90 (0.14)
VV	3611	10	79.3 (0.13)	0.76 (0.23)	0.82 (0.19)

Table 5. Classification results with a LVQ1 network . Network training parameters:MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 150 epochs, goal: 0.10 and learning rate: 0.010.Mean (Standard deviation) of 10 cross-validations.

Feature extracted	#Features	#Hidden units	% Accuracy	Sensitivity	Specificity
MSD	24	10	83.0 (0.12)	0.74 (0.23)	0.92 (0.1)
VV	3611	10	77.0 (0.15)	0.76 (0.23)	0.78 (0.17)

Table 6. Classification results with a LVQ2 network . Network training parameters: MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 50 epochs, goal: 0.01 and learning rate: 0.005. Mean (Standard deviation) of 10 crossvalidations.

is based on VBM analysis. After examining different designs for the SPM of the VBM we have found that the basic GLM design without covariates can detect subtle changes between AD patients and controls that lead to the construction of ANN classifiers with a discriminative accuracy of 83% in the best case as shown in table 6. A result of 83% of accuracy is really encouraging considering the number of subjects in the database. Improvements could be obtained using Adaptive Boosting including different types of ANNs and Support Vector Machines. The problem we have found is that the subjects wrongly classified maybe the most critical ones: old control subjects classified as AD (false positives) and subjects with a very early or mild dementia classified as normal (false negatives), exactly the ones which are the target in these studies that try to perform early detection of AD. Post-mortem confirmation data of AD diagnosed subjects could improve the results.

Further work may address the use of disease specific templates or other type of morphometric measures, such as Deformation-based Morphometry.

Acknowledgments

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

References

1. http://www.fil.ion.ucl.ac.uk/spm/.
2. J. Ashburner and K. J. Friston. Voxel-based morphometry: The methods. *Neuroimage*, 11(6):805–821, 2000.
3. G. F. Busatto, G. E. J. Garrido, O. P. Almeida, C. C. Castro, C. H. P. Camargo, C. G. Cid, C. A. Buchpiguel, S. Furuie, and C. M. Bottino. A voxel-based morphometry study of temporal lobe gray matter reductions in alzheimer's disease. *Neurobiology of Aging*, 24(2):221–231, 2003.

4. S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2):302–309, 1991.
5. C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of prodromal alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging*, 29(4):514–523, 2008.
6. A. F. Fotenos, A. Z. Snyder, L. E. Girton, J. C. Morris, and R. L. Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64(6):1032–1039, 2005.
7. G. B. Frisoni, C. Testa, A. Zorzan, F. Sabattoli, A. Beltramello, H. Soininen, and M. P. Laakso. Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry. *Journal of Neurology, Neurosurgery & Psychiatry*, 73(6):657–664, 2002.
8. Maite García-Sebastián, Alexandre Savio, Manuel Graña, and Jorge Villanúa. On the use of morphometry based features for Alzheimer's disease detection on MRI. *IWANN 09 Proceedings*, 2009.
9. Martin T. Hagan, Howard B. Demuth, and Mark H. Beale. *Neural Network Design*. PWS Pub. Co., Har/Dsk edition, December 1995.
10. Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition, July 1998.
11. Chengzhong Huang, Bin Yan, Hua Jiang, and Dahui Wang. Combining voxel-based morphometry with artifical neural network theory in the application research of diagnosing alzheimer's disease. volume 1, pages 250–254, May 2008.
12. S. Kloppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr, J. Ashburner, and R. S. J. Frackowiak. Automatic classification of MR scans in alzheimer's disease. *Brain*, 131(3):681, 2008.
13. T. Kohonen. *Self-organization and associative memory: 3rd edition*. Springer-Verlag New York, Inc., 1989.
14. Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1):46–57, 2004.
15. Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C. S. Carter, V. A. Stenger, S. Davis, H. Aizenstein, and J. T. Becker. Discriminative MR image feature analysis for automatic schizophrenia and alzheimer's disease classification. *Lecture Notes in Computer Science*, pages 393–401, 2004.
16. Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy 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.
17. D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning internal representations by error propagation*, pages 318–362. MIT Press, 1986.
18. R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox. Mapping the evolution of regional atrophy in alzheimer's disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, 99(7):4703, 2002.
19. Panu Somervuo and Teuvo Kohonen. Self-Organizing maps and learning vector quantization for feature sequences. *Neural Process. Lett.*, 10(2):151–159, 1999.
20. Donald F. Specht. Probabilistic neural networks. *Neural Netw.*, 3(1):109–118, 1990.