Short-term Prediction of MCI to AD conversion based on Longitudinal MRI analysis and neuropsychological tests

Juan Eloy Arco[1], Javier Ramírez[2], Carlos G. Puntonet[3], Juan Manuel Górriz[2], and María Ruz[1]

[1] Department of Experimental Psychology
`jearco@correo.ugr.es`
[2] Department of Signal Theory Networking and Communications.
[3] Department of Architecture and Computer Technology
University of Granada, Granada 18071, Spain

Abstract. Nowadays, 35 million people worldwide suffer from some form of dementia. Given the increase in life expectancy it is estimated that in 2035 this number will grow to 115 million. Alzheimer's disease is the most common cause of dementia and it is of great importance diagnose it at an early stage. This is the main goal of this work, the development of a new automatic method to predict the mild cognitive impairment (MCI) patients who will develop Alzheimer's disease within one year or, conversely, its impairment will remain stable. This technique will analyze data from both magnetic resonance imaging and neuropsychological tests by utilizing a t-test for feature selection, maximum-uncertainty linear discriminant analysis (MLDA) for classification and leave-one-out cross validation (LOOCV) for evaluating the performance of the methods, which achieved a classification accuracy of 73.95%, with a sensitivity of 72.14% and a specificity of 73.77%.

1 Introduction

The ability to diagnose and predict Alzheimer's disease (AD) at an early stage has great impact on the possibility for improving treatment choices of this disease. For this reason, in recent years there has been a large increase in the number of studies attempting to develop systems that help in the diagnosis of AD ([1], [2], [3], [4]). In [5], a technique was proposed for predicting future clinical changes of MCI patients by using both baseline and longitudinal multimodality data. The main drawback of this study is the request of having multimodality data across different time points for each subject, which limits the size of subjects that can be used for the study. Most existing research focuses on only a single modality of biomarkers for diagnosis of AD and MCI, although recent studies have shown that different biomarkers may provide complementary information for the diagnosis of AD and MCI ([6], [7]).

Despite the brilliant solutions presented by these approaches, all of them are focused on classifying AD or MCI patients from healthy controls. Since the earlier

the diagnosis of this disease, the more effective the treatment, in this study we propose a method to compare between MCI patients who had converted to AD within 12 months and MCI patients who had not converted to AD within 12 months, in order to predict whether the patient will develop the disease or not. Once the images have been preprocessed, the whole brain is then partitioned into 116 regions of interest (ROIs) in terms of the Automated Anatomical Labeling (AAL) atlas, and the mean intensity and standard deviation value of each region was acquired by averaging the intensities and standard deviations values within that region. These extracted values along with the two neuropsychological tests (MMSE and ADAS-Cog) make the dataset. To reduce its size, a Student's t-test selects the most important features, i.e. those with greater discrimination power. Once this is achieved, MLDA algorithm is used for classification, evaluating the performance with a Leave-One-Out cross validation technique (LOOCV).

The organization of the rest of the paper is as follows. Details or our method based on both MR images and neuropsychological tests and MLDA algorithm for classification are mentioned below. A description of the data used in the preparation of this article is done in Section 2. The method consists of four stages. In Section 3.1 we focus on the source of information, which is based on an anatomical atlas for MRI and direct scores in the case of neuropsychological tests. It is necessary to select the right features not to overtrain the system, since this can cause a decrease in its performance and an increase in the computation time. That is the purpose of Student's t-test, which is described in Section 3.1. The classification algorithm and the techniques to evaluate its performance are available in Section 3.3. The experimental results are provided in Section 5, and a discussion of research contributions and practical advantages in addition to the conclusions are available in Section 6 and Section 7.

2 Database

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. In this paper, only ADNI subjects with all corresponding MRI, MMSE and ADAS-Cog baseline data are included. This yields a total of 134 MCI subjects who had at least three longitudinal scans (baseline image, and two subsequent images six months and twelve months after) including 73 MCI converters who had converted to AD within 12 months and 61 non-converters who had not converted to AD within 12 months.

3 Methods

3.1 Feature extraction

In Alzheimer's disease, the hippocampus is one of the first regions of the brain to become damaged and that is why it is used as a marker of early AD in a vast number of studies, therefore it is logical that some approaches focus on the study of changes in it. In [8], the classification accuracy of a system was tested using the hippocampal volume as an only feature. Volumes were normalized by the total intracranial volume computed by summing SPM5 segmentation, averaging left and right volumes for more robustness with respect to segmentation errors as proposed in [9]. Hippocampal shape is another feature used in other approaches. More specifically, [10] described a new method to automatically discriminate between patients with Alzheimer's disease or mild cognitive impairment using spherical harmonics (SPHARM) coefficients to model the shape of the hippocampi. These coefficients are a mathematical approach to represent surfaces with spherical topology, which can be seen as a 3D analog of Fourier series expansion.

Another approach is based on a labeled atlas for grouping the voxels into anatomical regions and was employed in [11]. The number of available atlases is large but this work uses AAL (Automated Anatomical Labeling), [12], a predefined anatomical atlas formed by 116 regions of interest (ROI), meaning which has not been specifically designed for studying patients with AD so its areas do not necessarily represent pathologically homogeneous regions. Once the structural images were segmented into gray matter density (GMD) and white matter density (WMD), individual GMD and WMD maps were partitioned into the 116 regions of AAL. Then both the mean and standard deviation of the GMD/WMD values of each region was then extracted by averaging the GMD/WMD values of all voxels within that region. Thus, each subject has a total of 464 features from grey matter and white matter images and 2 more features from neuropsychological tests, resulting on 466 features for each session.

3.2 Feature selection

Not all the features are equally effective. Some of them may become irrelevant or redundant for the classification process. From this arises the necessity to select a small set of features with the greatest discriminative power to improve the performance of the final classifier ([13], [14]) and to speed up computation ([15] and [16]). PCA (Principal component analysis, [17]) and ICA (Independent component analysis, [18]) are two methods widely used in literature ([19], [20]). The former is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, which are known as principal component. The latter focus on separating a multivariate signal into additive non-Gaussian and statistically independent subcomponents. However, a different approach was

adopted in this work, using a filter ranking based on two-sample two-tailed t-tests. For each feature of the complete data set, a decision test was performed for the null hypothesis that the data in features vector of both classes (MCI-C and MCI-NC) come from independent random samples from normal distributions with equal means but unknown variances, at the 5% significance level. The alternative hypothesis is that the data in both vectors come from populations with unequal means. Mathematically, the test statistic is:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \tag{1}$$

where $\overline{x}$ and $\overline{y}$ are the means of each group, S_1 and S_2 are the sample standard deviations and n and m are the number of features for each group. This process was developed on the training set of each LOOCV fold. Thus, features whose p-values were less than the significance level were selected meaning that from the whole 116 brain regions, only 26 of them had an adequate discriminative power for use in the classification process.
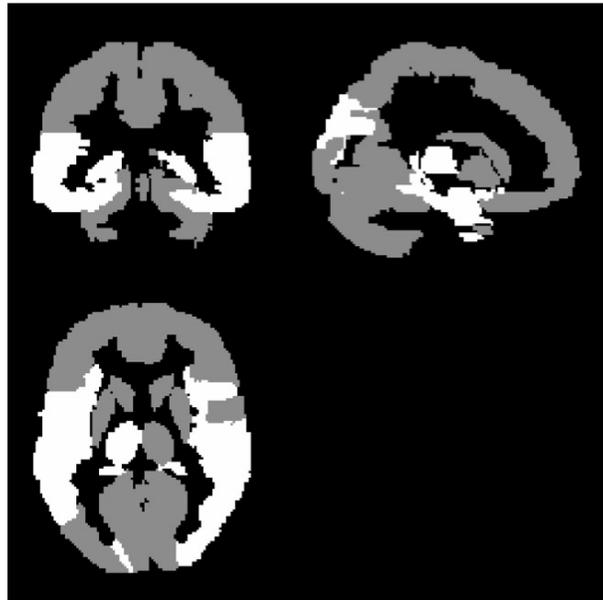
Fig. 1. Map of the brain regions chosen by the feature selection process (white colour). Some of the 28 areas with greater power of discrimination are the hippocampus, amygdala, thalamus, insula, temporal medial, temporal superior and occipital inferior.

3.3 Classification

Although in the literature there are more commonly used algorithms (i.e, SVM), in this work a variation of Linear Discriminant Analysis (LDA) was employed. LDA is a classification method that projects high-dimensional data onto a line and performs classification in this one-dimensional space. The projection maximizes the distance between the means of the two classes while minimizing the variance within each class. This defines the Fisher criterion, which is maximized over all linear projections, w:

$$J(w) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2} \tag{2}$$

where m represents a mean, s^2 represents a variance, and the subscripts denote the two classes. Therefore, the main objective of LDA is to find a projection matrix that maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix. As in PCA, the eigenvalues are of great importance in the correct separation of the classes. Following the complete mathematical procedure described in [21], Equation 2 can be rewritten as follows:

$$J(w) = \frac{w_k^T S_B w}{w_k^T S_W w} = \frac{\lambda_k w_k^T S_B w_k}{w_k^T S_W w_k} = \lambda_k \quad con \quad k = 1 \ldots d \tag{3}$$

where S_B is the "between classes scatter matrix", S_W is the "within classes scatter matrix" and w_k is the eigenvector associated to the eigenvalue λ_k. Consequently, to maximize the solution the eigenvector associated with the largest eigenvalue must be considered.

However, the traditional LDA cannot be directly used when the within-class scatter matrix is singular, as in the case of limited samples and high dimensional feature space. In this work, the dimension of feature space was still higher than the number of samples. In order to avoid these critical issues, [22] proposed a maximum uncertainty LDA-based approach (MLDA) to overcome the instability of the S_W matrix. It is based on the maximum entropy covariance selection method developed to improve quadratic classification performance on limited sample size problems.

The proposed method considers the issue of stabilizing the S_W estimate with a multiple of the identity matrix by selecting the largest dispersions regarding the S_W average eigenvalue. This selection algorithm expands only the smaller and consequently less reliable eigenvalues of within-class scatter matrix S_W. Thus, it is necessary to replace S_W matrix, as follows:

$$S_W^* = S_P^*(N - g) = (\Phi \Lambda^* \Phi^T)(N - g) \tag{4}$$

where S_P is the covariance matrix, Φ and Λ are the eigenvalues and eigenvectors of the covariance matrix, respectively, N is the number of training patterns from both classes and g is the total number of classes. It is a straightforward method that overcomes both the singularity and instability of the within-class

scatter matrix S_W when LDA is used in limited sample and high dimensional problems, so that's the reason why we chose it for this work.

4   Performance evaluation

In a general classification problem, the goal is to learn a classifier that performs well on unseen data drawn from the same distribution as the available data. One common way to estimate generalization capabilities is to measure the performance of the learned classifier on test data that has not been used to train the classifier. When a large test data set cannot be held out or easily acquired, resampling methods, such as cross validation, are commonly used to estimate the generalization error ([23]). Leave-one-out cross validation (LOOCV) was used to estimate the performance of the classifier. LOOCV involves separating the data so in each iteration there is only a test data while the remaining data are used to train the classifier. This means that on every fold of LOOCV, the most discriminative features are calculated and projected onto a one-dimensional space to properly determine the label of the testing sample.

Other measures to evaluate the performance of a classifier can be extracted from the confusion matrix. Accuracy is the proportion of the total number of predictions that are correct. Secondly, sensitivity (or true positive rate) measures the proportion of actual positives which are correctly identified. And finally, specificity (also called the true negative rate) measures the proportion of negatives which are correctly identified. It is desirable to have a classifier that gives high values of these three measures. In [24], ROC curve illustrates the performance of a binary classification as its discrimination threshold is varied. The curve is created by plotting the true positive rate (i.e. sensitivity) against the false positive rate (that is, 1-specificity) at various thresholds settings. This area can be interpreted as the probability that given a couple of patients (in our case, a mci converter and a non-converter patient), our algorithm classify them properly.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{5}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

5   Results

The aim of this work was the development of a completely automatic method for prediction of Alzheimer's disease and it has been broadly achieved. The experiments carried out on the database composed by both structural MRI (segmented into gray matter density and white matter density) and neuropsychological tests

(MMSE and ADAS-Cog). Thus, in each session there are two measures (mean and standard deviation) for each region of both segmented images besides the two neuropsychological tests. Several trials were made combining the different features yielding a value of accuracy equal to 73.95 %, with a sensitivity of 74.14%, a specificity of 73.77% and an area under the ROC curve of 0.7923.

Table 1. Results obtained using gray matter, white matter and neuropsychological tests from sessions 6/12 months before conversion (MCI-converters) and combining data from both sessions.

GRAY MATTER + WHITE MATTER				
6 months before conversion				
Features used	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Means + Tests	65.67	72.13	68.75	0.7913
Deviations + Tests	65.67	70.49	67.97	0.7962
Means + Deviations + Tests	65.67	72.13	68.75	0.7839
12 months before conversion				
Features used	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Means + Tests	67.24	63.93	65.55	0.7671
Deviations + Tests	68.97	62.3	65.55	0.7646
Means+ Deviations + Tests	67.24	65.57	66.39	0.7674
6 + 12 months before conversion				
Features used	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
Means + Tests	74.14	73.77	73.95	0.7923
Deviations + Tests	72.41	73.77	73.11	0.7911
Means + Deviations + Tests	74.14	73.77	73.95	0.7925

Table 1 shows the results obtained by the LDA classification algorithm when data from MRI images (means and deviations) and neuropsychological tests (MMSE and ADAS-Cog) are used as input features. For MCI converters patients, the data from one and two sessions before their conversions (i.e. six and twelve months before the conversion session respectively) can be used separately and in combination of both sessions. Besides, the average conversion session was calculated for all these patients, resulting that this was the fourth session (month 18 of the longitudinal analysis). Therefore, the data used for MCI non converters patients were those relating to the second and the third sessions.

The results show that we can predict more reliably the development of Alzheimer's disease 6 months before it appears instead of 12 months before the diagnosis of this disease, something which otherwise is logical. Regarding the use of means and deviations of each atlas region, there are no major differences between choosing one or the other. However, combining the data from two

previous sessions to the diagnosis, both accuracy and specificity and sensitivity increases considerably (almost 10 percentage points), while the area under the ROC curve remained almost unchanged. A t-test was used to compare the different experiments, resulting in that they are statistically significant, with a *t-value* exceeding 600 and a *p-value* less than 0.0001.

6 Discussion

In this study, we introduce a new method for discriminating MCI patients who will be diagnosed with Alzheimer's disease (up to a year before that it happens) from MCI patients whose impairment will remain constant in this period of time. Using an atlas (AAL) for partitioning the brain into 116 anatomical regions, a t-test for feature selection and LDA as classification algorithm our method achieved a high accuracy (73.95%), and the AUC was 0.79. Other methods ([20], [10]) achieved accuracies above 90%, which are superior to those achieved in our work. It is necessary to clarify the complexity of the problem we faced and the great potential our method has shown. The development of an automated system for prediction of Alzheimer's disease in an early stage is not a new challenge. However, this work represents a step further in predicting this disease because only patients with mild cognitive impairment are considered, because subjects who developed Alzheimer's disease at some point in our study are only considered from sessions before conversion. Therefore, our method is able to find significant differences in patients whose clinical diagnosis is identical, using a very simple approach in which the features are based on statistical measures from anatomical regions of the brain.

Thus, this system can be used as an aid in the diagnosis of Alzheimer's disease because is fully automatic, so it is not required to choose a prior anatomical region where focuses the analysis since the entire brain is considered. The fact that it is only needed a nuclear magnetic resonance which is available in most of the diagnostic centers and that only a few minutes are necessary to collect the data from both neuropsychological tests are important advantages for using this technique for clinical diagnosis. A suggestion for future research might be to consider the contribution of each voxel separately, as Multi-Voxel Pattern Analysis (MVPA) proposes, a technique which allows to detect differences with higher sensitivity than conventional univariate analysis.

7 Conclusion

In the current study, we have developed a system to predict if MCI patients will develop Alzheimer's disease within a period of one year by combining data from magnetic resonance images (the mean and the standard deviation of each brain region proposed by the atlas AAL) and the results from two neuropsychological tests (MMSE and ADAS-Cog), yielding an excellent performance (73.95% accuracy and AUC=0.79). This promising discrimination power suggests that this technique could be used as an aid in the diagnosis of AD, becoming a promising

starting point for other more complex methods as multivariate pattern analysis. We conclude that it would be so interesting to repeat all the procedure of this work extracting the features from an anatomical atlas based on the most damaged regions by this disease instead of an anatomical atlas, since it could improve significantly the results.

Acknowledgments

References

1. R. S. Deskian, H. J. Cabral, F. Settecase, C. P. Hess, W. P. Dillon, C. M. Glastonbury, M. W. Weiner, N. J. Schmansky, D. H. Salat, and Alzheimer's Disease Neuroimaging Initiative B. Fischl. Automated mri measures predict progression to alzheimer's disease. *Neurobiology of Aging*, 31(8):1364–1374, 2010.

2. Y. Liy, T. paajanen, y. Zhang, E. Westman, L-O. Wahlund, A. Simmons, C. Tunnard, T. Sobow, P. Mecocci, M. Tsolaki, B. Vellas, S. Muehlboeck, A. Evans, C. Spenger, s. Lovestone, and H. Soininen. Automated mri measures predict progression to alzheimer's disease. *Neurobiology of Aging*, 31(8):1375–1385, 2010.

3. A. Chincarini, P. Bosco, P. Calvini, G. Gemme, M. Esposito, C. Olivieri, L. Rei, S. Squarcia, G. Rodriguez, R. Bellotti, P. Cerello, I. de Mitri, A. Retico, and F. Nobili. Local mri analysis approach in the diagnosis of early and prodromal alzheimer's disease. *Neuroimage*, 102(2):657–665, 2011.

4. A. Nazeri, H. Ganjgahi, T. Roostaei, T. Nichols, and M. Zarei. Imaging proteomics for diagnosis, monitoring and prediction of alzheimer's disease. *Neuroimage*, 58(12):469–480, 2011.

5. D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. *Plos one*, 7(3), 2012.

6. L. G. Apostolova, K. S. Hwang, j. P. Andrawis, A. E. Green, S. Babakchanian, J. H. Morra, J. L. Cummings, A. W. Toga, J. Q. Trojanowski, L. M. Shaw, C. R. Jack Jr., R. C. Petersen, P. S. Aisen, W. J. Jagust, R. A. Koeppe, C. A. Mathis, M. W. Weiner, and P. M. Thompson. 3d pib and csf biomarker associations with the hippocampal atrophy in adni subjects. *Neurobiology Aging*, 31:1284–1303, 2010.

7. A. M. Fjell, K. B. Walhovd, C. Fennema-Notestine, L. K. McEvoy, D. J. Hagler, D. Holland, J. B. Brewer, and A. M. Dale. Csf biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and alzheimer's disease. *J. Neurosci.*, 30:2088–2101, 2010.

8. R. Cuingnet, E. Gerardin, j. Tessieras, G. Auzias, S. Lehricy, M. Habert, M. Chupin, H. Benali, O. Colliot, and The Alzheimer's Disease Neuroimaging Initiative. Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the adni database. *Neuroimage*, 56:766–781, 2011.

9. M. Chupin, A. Hammers, R.S. Liu, O. Colliot, J. Burdett, E. Bardinet, J.S. Duncan, L. Garnero, and L. Lemieux. Automatic segmentation of the hyppocampus and the amygdala driven by hybrid constraints: method and validation. *Neuroimage*, 46(3):749–761, 2009a.

10. R. Geradin, G. Chtelat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, m. Niethammer, B. Dubois, S. Lehricy, L. Garnero, E. Francis, and O. Colliot. Multi-dimensional classification of hippocampal shape features discriminates alzheimer's disease ang mild cognitive impairment from normal aging. *Neuroimage*, 47(4):1476–1486, 2009.

11. B. Magnin, L. Mesrob, S. Kinkingnéhun, M.Pélégrini-Isaac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical mri. *Neuroradiology*, 51(2):73–83, 2009.

12. N. Tzourio-Mayer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15:273–289, 2002.

13. J. Yan, T. Li, H. Wang, H. Huand, J. Wan, K. Nho, S. Kim, S.L. Risacher, A. J. Saykin, and L. Shen. Cortical surface biomarkers for predicting cognitive outcomes using group l2,1 norm. *Neurobiology of Aging*, 36(1):S185–S193, 2015.

14. M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain. Dimensionality reduction using genetic algorithms. *IEEE transactions on evolutionary computation*, 4(2):164–171, 2002.

15. X. Cui, J. M. Beaver, J. St Charles, and T. E. Potok. Dimensionality reduction particle swarn algorithm for high dimensional clustering. *IEEE swarm intteligence symposium*, 1:1–6, 2008.

16. S. Salcedo-Sanz, A. Pastor-Snchez, L. Prieto, A. Blanco-Aguilera, and R. Garca-Herrera. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization - extreme learning machine approach. *Energy conversion and management*, 87:10–18, 2014.

17. I. T. Jolliffe. *Principal component analysis, Second Edition*. Springer, 1973.

18. A. Hyvärinen. Fast and robust fixex-point algorithms for independent component analysis. *IEEE transactions on neural networks*, 10:626–634, 1999.

19. I. —'Alvarez, J. M. Górriz, J. Ramírez, D. Salas, M. López, C.G. Puntonet, and F. Segovia. Independent component analysis of spect images to assist the alzhimer's disease diagnosis. *The sixth international symposium on neural networks(ISSN 2009), Advances in intelligent and soft computing*, 56:411–419, 2009.

20. Z. Dai, C. Yan, Z. Wang, J. Wang, M. Xia, K. Li, and Y. He. Discriminative analysis of early alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (m3). *Neuroimage*, 59(3):2187–2195, 2012.

21. Max Welling. Fisher linear discriminant analysis, 2010.

22. C.E. Thomaz, J.P. Boardman, D.L.G. Hil, J. V. Hajnal, A.D. Edwards, M.A. Rutherford, D.F. Gillies, and D. Ruckert. Whole brain voxel-based analysis using registration and multivariate statistics. *Proceedings of the 8th Medical Image Understanding Analysis MIUA'04*, 1:73–76, 2004.

23. R. B. Rao and G. Fung. On the dangers of cross-validation. an experimental evaluation, 2008.

24. J. A. Hanley and B. J. McNeill. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143 (1):29–36, 1982.