Development of Soft-Computing techniques capable of diagnosing Alzheimer's Disease in its pre-clinical stage combining MRI and FDG-PET images.

Olga Valenzuela, Francisco Ortuño, Belen San-Roman, Victor Martin-Rodriguez, Jose Luis Bernier, M. Jose Saez, Fernando Rojas, Ignacio Rojas

Department of Computer Architecture and Technology. University of Granada. 18060 Granada, Spain. olgavc@ugr.es

Abstract. In this paper an intelligent classifier was development, in which information provided from MRI and FDG-PET images are combined in order to obtain an automatic classifier. In the first step was to develop a classification method to tag simultaneously MR and FDG-PET images as either normal or with the Alzheimer's disease (AD). With the methodology obtained, and using similar features, the next step was the identification and classification in normal subjects, MCI (Middle Cognitive Impairment) patients and AD patients. The last step was the possibility to classify in Middle Cognitive Impairment Converters (MCI-C, i. e. , people that suffer a MCI and in the future will suffer from Alzheimer's disease within 18 months), and Middle Cognitive Impairment Non Converters (MCI-NC, i. e., people that suffer a MCI and in the future will not suffer from Alzheimer's disease). It is noteworthy that with this last study we could offer a tool to assist the early diagnosis of dementia.

Keywords: Gene Expression Microarray, Parallel Genetic Algorithm, Support Vector Machines

1 Introduction

There are several possible causes for dementia, but Alzheimer Disease (AD) is leading cause of dementia in the world, and increasing in the following years (Fig.1)

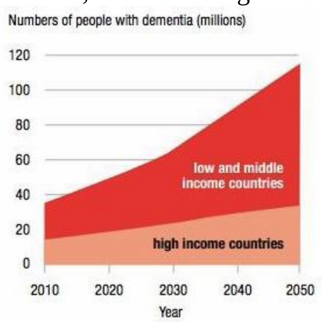


Fig1. Behaviour of the number of patient with dementia in high income countries and low middle income countries.

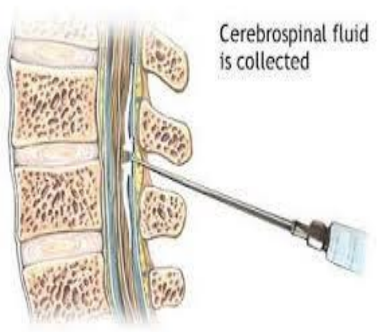
There are an estimated 35.6 million people in the world with dementia and more than 18 million of them have Alzheimer disease, which represents more than 50% of the total people with dementia.

Besides, the number of people with dementia is expected to increase to 65.7 million in 2030 and 115.4 million in 2050. One of the main questions that in the field of diagnose Alzheimer's is the possibility to forecast this pathology or at least to diagnose it before symptoms started.

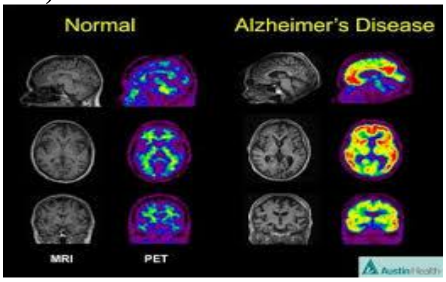
Today, the diagnosis of Alzheimer disease is made by using clinical criteria; however these criteria are not capable of diagnosing the disease in its pre-clinical stage, not allowing for an early diagnosis.

If an automatic system can forecast the diagnosis of the Alzheimer disease, future treatments could then focus the disease in its earliest stages, before irreparable brain damage has occurred. In this research field, different heterogeneous data base can be used. To mention some of them, important and significant results have been obtained using the technologies presented in Figure 2 for earlier Alzheimer disease detection.

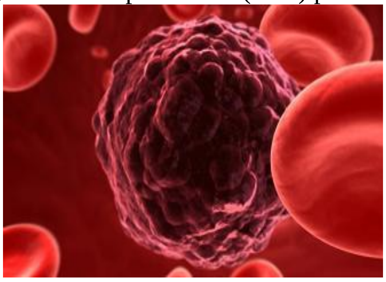
a) Biomarkers for earlier detection



b) Cerebrospinal fluid (CSF) proteins



c) Brain imaging/neuroimaging. MRI, PET FDG-PET images can be used.



d) Proteins in blood

Figure 2. Technologies used in the bibliography for Alzheimer disease detection

2 Brain imaging/neuroimaging for Alzheimer disease detection

One of the more promising research field for Alzheimer disease detection is the Neuroimaging processing [1][2][3]. In order to perform this contribution, the Alzheimer's Disease Neuroimaging Initiative (ADNI) database is used. ADNI researchers collect, validate and utilize data such as MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors for the disease. Data from the North American ADNI's study participants, including Alzheimer's disease patients, mild cognitive impairment subjects and elderly controls, are available from this site.

In this paper, images from 1.5-T and 3-T were simultaneously used for MRI. Also, information from PET images is combined. It is important to note that a pre-processing (normalization) of all the images obtained has been carried out, and all the images were visualized one by one, taking into account both the normal image, the gray matter and white matter. Abnormal or defectives images were considered corrupted and therefore discarded.

3 Feature Extraction

In this paper, we need to extract from an image the feature vector that characterizes it. Thus, our features are the approximate wavelet coefficients, using them to generate a classification rule to assist with diagnosis. As described in the following sections, the number of features is not as important as robustness to get the best classification accuracy (being *robustness* in an image application understood as the consistency of the results that certain feature provides across the entire application). Wavelets are mathematical functions that decompose data into different frequency components and then study each component with a resolution matched to its scale.

While the Fourier Transform only provides representation of an image based on its frequency content (losing time information of the signal), the Wavelet Transform provides both time and frequency information. Therefore, the Wavelet Transform is a better tool for feature extraction from images. The Discrete Wavelet Transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into numerically different frequency components, and then studies each component with resolution matched to its scale.

Assume *x(t)* is a square-integrable function; then the continuous wavelet transform of *x(t)* relative to a given wavelet ψ(t) is defined as:

$$W_{\psi}(a,b) = \int_{-\infty}^{+\infty} x(t)\psi_{a,b}(t)\,dt \tag{1}$$

Where:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right) \tag{2}$$

To get the DWT, previous equation can be discretized by restraining a and b to a discrete lattice (a = 2b; a > 0; a, b $\in \Re$). Then, the DWT can be expressed as follows:

$$ca_{j,k}(n) = DS\left[\sum_{n} x(n) g_j^*(n - 2^j k)\right] \tag{3}$$

$$cd_{j,k}(n) = DS\left[\sum_{n} x(n) h_j^*(n - 2^j k)\right] \tag{4}$$

Here $ca_{j,k}$ and $cd_{j,k}$ refer to the coefficients of the approximation components and detail components, respectively. *g(n)* and *h(n)* denote the low-pass filter and high-pass filter, respectively. *j* and *k* represent the wavelet scale and translation factors, respectively; and *DS* operator means the “down sampling” or decimation operation. The above decomposition process can be iterated decomposing successively the approximations in turn, so that the signal is broken down into various levels of resolution.

In case of images, the DWT is applied to each dimension separately, decomposing an image into four sub-bands, which are low-low (LL), low-high (LH), high-high (HH) and high-low (HL); where the LL sub-band can be regarded as the approximation component and it is used for the next level of the 2D-DWT, whereas the other sub-bands would be regarded as the detailed component of the image. A 2D-DWT scheme is shown in fig. 3.

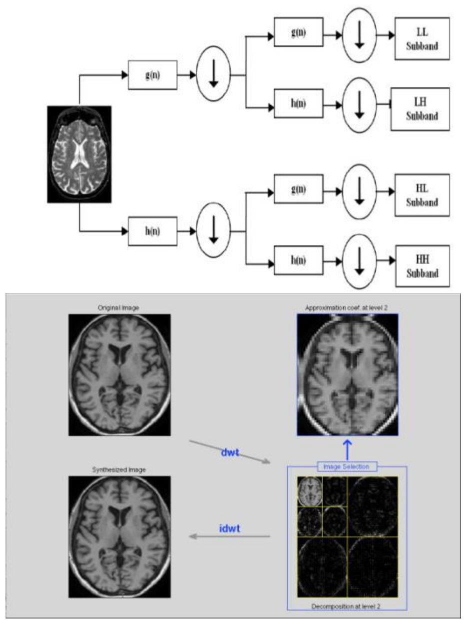


Fig. 3. A) 2D DWT decomposition scheme; B) Level 2 decomposition of an image using DWT

At each decomposition level, the half band filters produce signals spanning only half of the frequency bands. As the level decomposition is increased, a more compact and less resolution image is obtained.

4 Results and conclusion

First, all the MRI images are processed and the corresponding features are selected. Then each FDG-PET image [4] was motion-corrected as necessary, converted to a 30-minute static image, examined for different kind of noise or outlayer, and aligned with the corresponding MRI using the Image Registration Toolkit. An affine transformation was favored over a rigid one because it can account for scaling or voxel size errors remaining after phantom correction of the MRI. The next step was combining all information for being used for an intelligent classifier. In this phase, different techniques are taking into account:

4.1 SVM

The theory of SVMs was initially developed by V. Vapnik [6] in the early 80's and focuses on what is known as Statistical Learning Theory. SVM initially appeared to separate two classes, although its application has extended to any finite number of classes [5]. These techniques perform a linear classification, upon vectors processed on a higher dimensional space, ie in the transformed space, separating the different classes using an optimal hyperplane to maximize the margin between classes. We have evaluated SVM using different algorithms like SMO-Sequential Minimal Optimization- and using LIbSVM library, both with different types of kernel function.

4.2 Decision Trees

A classification tree consists of nodes, arcs and leaves. Each node represents a decision on a particular attribute value, being the terminal nodes where a decision is made about the class map. When classifying a new case will have to be compared the values of the attributes with the decisions taken at the nodes, following the branch, that match those values, in every decision. Eventually you will reach a terminal node or leaf that predicts the class for the case treated. One of the algorithms based on decision trees most used is the C4.5 [7] whose approach is TDIDT (Top Down Induction of Decision Trees), which is characterized by using a strategy of divide and conquer descending. Another algorithm, is the Random Forest introduced by Breiman in 1999, uses a set of trees (forest) classification. To classify a new data, is taken as input for each tree and produces the corresponding output classification. As a final decision for the whole of trees, takes the class with the most votes. Other algorithms have been evaluated like CART[8]- Classification and Regression Trees- and NBTREE [9] which is a Naïve Bayes decision tree.

4.3 Nearest Neighbor

The basics of neighborhood classification were established by [10] in the early 50's. The nearest neighbor method and its variants are based on the intuitive idea that similar objects belong to the same class, so that the class to which it belongs an object can be inferred from the class they belong objects of the learning sample that most resembles him. The idea of similarity is reflected formally in the concept of distance. The algorithm k-NN [11], is included within the so called lazy learning techniques, and does not generate a knowledge structure, which shapes the information inherent in the training set, but the dataset itself, represents the model, i.e. is not built any model, the model is the database itself or the training set. We have evaluated KNN algorithm using 2 different variants, one using k= 1 and the other one with k =50

It can be concluded that although all paradigms have a good result, which obtains better accuracy (with higher data rates to 90%) was SVM. It also noticed that the unification of MRI and PET images obtain better results than the single use of this information for building isolated classifiers.

5 ACKNOWLEDGE

This work has been partially supported by the projects: Spanish Ministry of Science and Innovation SAF2010-20558 and Excellence projects Junta de Andalucia P12-TIC-2082 reference.

REFERENCES

[1] Katherine R. Gray, Paul Aljabar, Rolf A. Heckemann, et.al "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease", NeuroImage 65 (2013) 167–175

[2] Heckemann, R.A., Keihaninejad, S., Aljabar, P., Gray, K.R., Nielsen, C., Rueckert, D. Hajnal, J.V., Hammers, A., the Alzheimer's Disease Neuroimaging Initiative, 2011. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. Neuroimage 56, 2024–2037.

[3] Aljabar, P., Rueckert, D., Crum, W.R., 2008. Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. Neuroimage 43, 225–235.

[4] Gray, K.R., Wolz, R., Heckemann, R.A., Aljabar, P., Hammers, A., Rueckert, D., for the Alzheimer's Disease Neuroimaging Initiative, 2012. Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease. Neuroimage 60, 221–229.

[5] Urquiza, J. M., I. Rojas, et al. (2012). "Using machine learning techniques and genomic/proteomic information from known databases for defining relevant features for PPI classification." Computers in Biology and Medicine 42(6): 639-650.

[6] Vapnik, V. N. The Nature of Statistical Learning Theory. Springer. 1995

[7] Quinlan J. C4.5: Programs for machine learning, Morgan Kaufmann Pub., 1993 (ISBN: 1558602380).

[8] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and regresion trees. Wadsworth Int. Group, Belmont, CA, 1984.

[9] R. Kohavi. Scaling up the accuracy of naive-Bayes classifiers: a decision-tree hybrid. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp 202-207, 1996.

[10] E. Fix and J. Hodges. Discriminatory analysis, nonparametric discrimination consistency properties. Technical Report 4, US Air Force, School of Aviation Medicine, Randolph Field, TX, 1951

[11] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, IT-13(1):21–27, 1967