Classification of MRI images for Alzheimer's disease detection

Luis Javier Herrera*, Ignacio Rojas, H. Pomares, A. Guillén, O. Valenzuela, O. Baños
Computer Architecture and Computer Technology Department, University of Granada
Granada, Spain
*email: jherrera@atc.ugr.es

***Abstract*—Alzheimer's Disease (AD) is normally identified by several behavioral symptoms often mistakenly associated to age-related concerns or stress. However correct diagnosis and monitoring of the disease requires of additional resources. This paper presents a new methodology for classification of Alzheimer's disease from MR images for medical support. A large database with more than one thousand patients was used. Two different problems are tackled in this work: a first one where a classification method is developed to classify MR images as either normal or with the Alzheimer's disease and a second one for the identification and classification between normal subjects, MCI patients and AD patients. It is noteworthy that with this last study we could offer a tool to assist the early diagnosis of dementia. The outline of the methodology includes wavelet feature extraction from the MRIs, dimensionality reduction, training-test subdivision and classification using Support Vector Machines. Some concerns related to performance evaluation and dimensionality reduction are discussed.**

***Keywords*—Support Vector Machine (SVM); Alzheimer's Disease; Mild Cognitive Impairment (MCI); PCA; Wavelets; MRI**

I. INTRODUCTION

Alzheimer is a neurodegenerative disease typically identified by a memory loss and other mental capabilities decrease, as neurons die and different areas of the brain get atrophied. The disease generally presents a duration of 10 years after diagnosis, although this can vary depending on the severity of the disease and the moment of diagnosis. It is the most common cause of dementia in the world; it is incurable and terminal, appearing mostly in people older than 65 years old. The cause and progression of Alzheimer's disease are not well understood, although research indicates that the disease is associated with plaques and tangles in the brain. Current treatments only help with the symptoms of the disease, and there are no available treatments that stop or reverse its progression.

Today, the diagnosis of Alzheimer's disease is done by clinical criteria, however these criteria are not able to diagnose the disease in its pre-clinical stage, not offering an early diagnosis. Given this situation, it is necessary to develop methods and techniques to be included in the criteria that provide an early diagnosis, which would allow people with dementia to plan ahead while they still have the capacity to make important decisions about their future care as well as it would allow them to access available drug and non-drug therapies that may improve their cognition and enhance their quality of life.

Alzheimer's disease is characterized by loss of neurons and synapses in the cerebral cortex and certain subcortical regions. This loss results in gross atrophy of the affected regions, including degeneration in the temporal lobe and parietal lobe, and parts of the frontal cortex and cingulate gyrus. Degeneration is also present in brainstem nuclei like the locus coeruleus.

Many studies have focused on quantifying focal atrophy in the temporal lobe [1] [2] and even exist visual scales to quantify the degree of atrophy, which are quick and easy to use. Recently have been published validations of computerized methods to measure the degree of temporal atrophy. In comparison, these methods have a similar discriminatory power [3] with the advantage that they would facilitate measurements and would provide more objective results by standardizing the methods of analysis [4].

In parallel, there has been observed that AD patients may have a striking atrophy in posterior regions known as posterior cortical atrophy. This impairment seems to be more characteristic, but not exclusive, of AD with typical clinic manifestations and shows in cases of early onset (onset before 65 years). Recently have been proposed visual scales to quantify the degree of posterior cortical atrophy that seems to be useful to discriminate AD from other dementias, especially from frontotemporal dementia (FTD), which also may cause temporal lobe atrophy [5].

Computer-based studies look for the acquisition of certain features that lead a classification rule to classify different kind of MR images. In this way, the start point is the classification of MR images of the human brain as normal or abnormal. Using wavelets as image features to train a Support Vector Machine (SVM) classifier and using neural network self-organizing maps (SOM) good classification accuracy was achieved in [6].

The following studies continued the same line of work as most of them also used wavelets as features and a certain classifier. However, some of them introduced Principal Component Analysis (PCA) as feature reduction algorithm [7] [8]. These last two studies used Neural Networks (NN) for classification. Finally, the final step is to use the developed technique for the classification of specific conditions. So,

978-0-7695-5137-1/13 $26.00 © 2013 IEEE
DOI 10.1109/SocialCom.2013.127

wavelets and SVM were used to classify mammographic masses from digitalized mammograms as benign or malign [9], and in Alzheimer's disease field, volumetric and shape features together with PCA and SVM were used to classify MR images as having the disease or not [10].

However, most the articles published to date for the use of intelligent classification systems Alzheimer's disease either use a low number of data for both training and test (in some cases less than a hundred), and/or do not qualify, since it is a complex task, patients with Mild Cognitive Impairment (MCI)[20]. Imaging findings in patients with MCI are typically inconspicuous, and often considered as normal. The main problem is that do not exist standardized criteria to determine if the impairment observed in a MR image is caused by normal aging or if it can be admitted as pathological. After 65 years is considered normal to detect mild signs of atrophy and little impairments of the white matter, but the evaluation of the degree of damage, which will determine if it is considered pathological, remains subjective [11].

This paper presents results comparing the use of feature reduction methods in the classification of Alzheimer's disease, extending previous results in [16]. A large database of MR images is used, with more than one thousand patients, and different subdivisions of training-test sets has been used to show the robustness of the proposed approach. Then, two main objectives are approached: a SVM model is developed to classify MR images as either normal or with the Alzheimer's disease; then the identification and classification between normal subjects, MCI patients and AD patients is tackled. The results obtained show that these models could offer a tool to assist the early diagnosis of dementia.

II. DATA

ADNI (Alzheimer's Disease Neuroimaging Initiative) beginning in 2004 is a 7-year massive effort to support research in the discovery and development of treatments that slow or stop the progression of AD. ADNI is a multisite longitudinal clinical/imaging/genetic/biospecimen/biomarker study. Its goal is to determine the characteristics of AD as the pathology that evolves from normal aging to mild symptoms, to MCI, to dementia. ADNI is committed to establishing standardized methods for imaging/biomarker collection and analysis for use in clinical trials.

Figure 1 shows the eight cores that structure ADNI. In this paper we are only interested in the MRI core, based at the Mayo Clinic, Rochester, Minnesota, and responsible for all MRI procedures and for developing standardized imaging methods.

Describing the past, present and the future of the ADNI MRI core, it starts with the "ADNI-1" for the first 5 years (approximately, from 2005 to 2010). Focusing on the MR images, all subjects enrolled in this period received a 1.5T protocol examination at multiple point times which varied by baseline clinical diagnosis: MCI at 0, 6, 12, 24 and 36 months; AD at 0, 6, 12 and 24 months; and controls at 0, 6, 12, 24 and 36 months. A subset of participants (approximately 25%) was enrolled in a 3T arm, which involved MRI scanning at both 1.5T and 3T at each scheduled time. ADNI-1 ended October 2010.

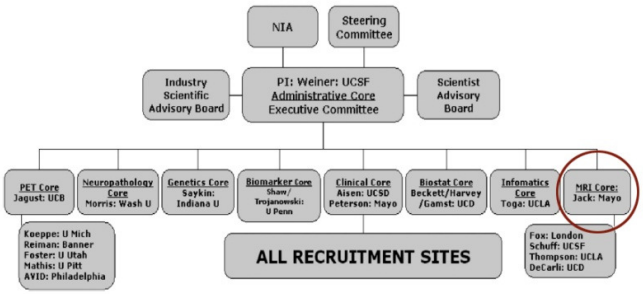

Figure 1. Governance and organization of ADNI

Plans for the second 5 years of ADNI running through 2015 are based on "ADNI-GO" (GO stands for "Grand Opportunity") and "ADNI-2", which is the 5-year competitive renewal of ADNI-1. These initiatives are following three cohorts of subjects:

Cognitively normal and late MCI subjects carried forward from ADNI-1 (followed at 1.5T).

• Early MCI enrolled in ADNI-GO and carried forward into ADNI-2 (scanned at 3T).

• Cognitively normal, late MCI and AD subjects newly enrolled in ADNI-2 (scanned at 3T).

Thus, the initiatives described above provide us a great database of MR images that are used in this paper. Specifically, for the realization of this study were downloaded 1500 images with a total size of approximately 240 GB. Once they were normalized and some images with errors were eliminated, 1350 images (443 are from cognitively normal subjects, 448 from MCI subjects and 459 from AD subjects) are left with a size of 15.7 MB each one, which makes a database of normalized MR images with a size of approximately 21 GB. The extension of the files is .nii, which corresponds to the format NIfTI (Neuroimaging Informatics Technology Initiative). This format is adapted from the widely used ANALYZE 7.5 using the "empty space" in its header to add new features. For more information see the official website of ADNI [12].

In studies that involve images of many patients, it is often useful (in this paper necessary) to coregister a brain image of a patient to that of another subject or a standard template. We refer to this process as spatial normalization.

To perform the normalization of our MR images we used the "SPM5" (currently available SPM8 release) toolbox for MATLAB. SPM (Statistical Parametric Mapping) entails the construction of spatially extended statistical processes to the test hypotheses about regionally specific effects [13]

III. FEATURE EXTRACTION

Feature vectors characterizing the images have to be extracted to operate them. Wavelet coefficients were used as features in the design of the classifiers.

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, so it loses time information of the signal, the Wavelet Transform provides both time and frequency information (see Figure 2). Therefore, the Wavelet Transform is a better tool for feature extraction from images.

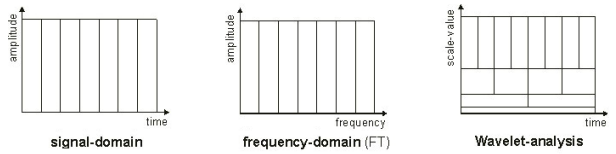

Figure 2. The development of signal analysis

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 a numerically different frequency components, and then studies each component with resolution matched to its scale.

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

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

where

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

To get the DWT, equation 1 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:

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

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

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 [14].

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, meanwhile the other sub-bands would be regarded as the detailed component of the image. A 2D-DWT scheme is shown in Figure 3.

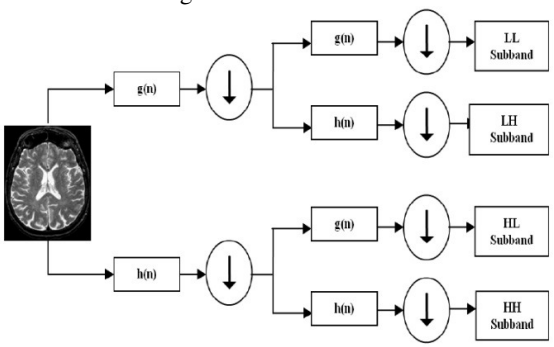

Figure 3. 2D DWT decomposition scheme

Figure 4 shows the decomposition up to level 2 of an image using the Wavelet Toolbox of MATLAB.

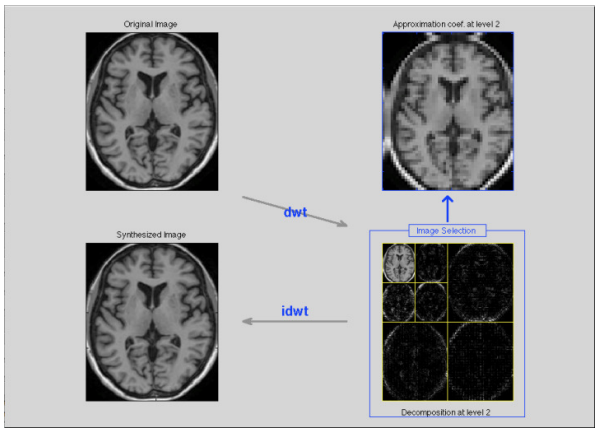

Figure 4. Level 2 decomposition of an image using DWT

At each decomposition level, the half band filters produce signals spanning only half the frequency bands. This makes the frequency resolution two when the indetermination in frequency becomes a half less [9], the size of the first level approximation coefficients of an N by N image is N/2 by N/2, the second level is N/4 by N/4 and so on. As the level decomposition is increased, a more compact and less resolution image is obtained [6].

Based on the studies cited earlier, we find that two wavelet families are mostly used in the literature: the Daubechies-4 wavelet up to level 2 and the Haar wavelet up to level 3. Both alternatives were considered in this study.

IV. FEATURE SELECTION

Dimensionality reduction is a complex problem, with a twofold objective. The objective of dimensionality reduction is twofold: decrease the model complexity (both in training and testing), and increase (or at least maintain) the performance (generalization capabilities) of the models. It offers two main alternatives: feature reduction and feature selection

The first deals with modifications of the original input data space in order to reduce the original dimensionality and obtain a new simpler representation of the problem including the largest part of the original data variability. Principal Component Analysis (PCA) is a traditional feature reduction technique with converts a set of observations into a set of values of linearly uncorrelated variables called principal components, being the number of principal components is less than or equal to the number of original variables.

On the other hand, feature selection approaches deal with the selection of the most relevant features (a subset of the original feature set) that obtain an optimal performance in the classification problem, or that obtain the best tradeoff between model simplicity and performance. Normalized Mutual Information Feature Selection (NMIFS)[21] algorithm is a powerful technique derived from the minimum Redundancy Maximum Relevance (mRMR) which has shown to present good results in selecting the most relevant features in a problem. It is based in the Mutual Information (MI) measure taken from the Information Theory. The fraser method was used in this work to perform the MI estimations needed.

V. SVM CLASSIFICATION

Support Vector Machines (SMVs)[18] were used as classification technique, with LIBSVM toolbox under MATLAB as simulation software [15]. SVMs are originally formulated for binary classification. Multi-classification is normally performed by combination of different binary classifiers in a one-vs-one approach. Final class is decided according to a voting scheme.

The SVM approach is a well-known paradigm which has shown to perform optimally (with similar or better performance than many other paradigms) in many applications. Another advantage over other methods is that it is less sensitive to the dimensionality of the problem in computational cost, which allows working with complex problems with a large number of variables involved.

Among the kernel functions alternatives, Gaussian Radial Basis Function kernel was chosen as it has proven to offer a good asymptotic behavior[17]. The estimation of the hyper-parameters of the SVM (C to control over-fitting, and γ to control the width of the Gaussian kernel) was done using grid search and cross-validation. Both hyper-parameters were logarithmically ranged between 2^{-15} and 2^{5}, in a 10x10 grid search. In the case of evaluating a large number of classifiers, however, this optimization technique can be computationally too costly. The Extreme Machine Learning approach points out[19] that it is possible to obtain successful classification results by using *reasonable* values of hyper-parameters. Thus in the ten-fold training-test subdivisions, the grid-search approach for hyper-parameters optimization was done only once, using the same optimal hyper-parameters for the rest of training-test subdivisions.

Traditionally a MRI classification process could include the following steps: 2D-DWT is used to perform the feature extraction, then a feature selection algorithm based on PCA is used and finally classification is performed. The structure of the procedure is schematically shown at Figure 5.

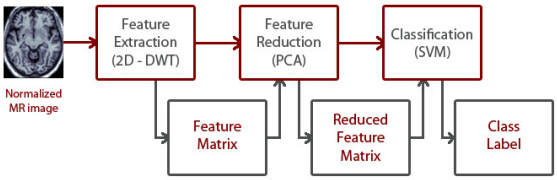

Figure 5. Structure of Experiment I schematically

In the datasets operated in this work, the complete procedure shown in Figure 5 was applied to every slice of every image to set up a ranking of accuracy. The 20 top slice or slices considered in the process where the best to extract the wavelets from them. This procedure led to the higher slices where only the hippocampus can be observed.

No. of Classes	Wavelet Family	No. of Patients	No. of Features	No of. PC.
2 (NAD)	Db4 L2	902	40144	369
2 (NAD)	Haar L3	902	8400	297
3 (NMCIAD)	Haar L3	1350	12000	466

Table 1. Matrices of features after Feature Extraction process

VI. RESULTS

Original wavelet features' data were first used for performance assessment. Random 10-fold cross validation was used to obtain different training-test subdivisions and obtain a more precise performance estimation on unseen data. Table 2 shows the CV-performance attained. As mentioned, grid-search was used only for one of the subdivisions, and the optimal hyper-parameters selected were used for the rest of subdivisions. The grid-search training process needed 10 hours in a core-i7 PC with 8GB of RAM under Linux. Rest of training processes with given optimal hyper-parameters needed 1 hour in total.

Afterwards and as suggested in previous works [7][8][16], PCA was used to reduce the dimensions of features to a higher degree. Once the principal components and their associated variances were calculated, a number of them that preserves 95% of total variance was kept [16]. See table 1 for summary of the datasets involved in this study. In this case the computational cost was around ten times lower for the PCA features, due to the lower dimensionality of the data involved in the problem.

Results show a very high performance in classification results in the NAD problem, higher than 95% for both types of wavelets transformations without the use of PCA feature reduction. Use of PCA, though recommended in some works, shows a diminishing in the performance results. Although the computational cost in the training process is much lower when using PCA features, it has to be noted that the computational cost of the use of the models (testing) is

Study	Mean CV Accuracy (%)
Db4 L2 Wavelet	95.01
Db4 L2 Wavelet + PCA	93.90
Haar L3 Wavelet	96.23
Haar L3 Wavelet + PCA	94.79

Table 2. Accuracy of the NAD studies

negligible. So it is to be taken into account that if training computational cost is affordable, using the original wavelets features is recommended from the performance point of view. It has to be recall that the use of ELM [18] reduction can lead to a very important decrease in computational cost with good performance results.

On the other hand, feature selection algorithms were applied to both types of feature spaces, original wavelet features and PCA features. Results showed that reduction in both cases led to no improvement in performance through feature selection. Moreover it was observed that large divergences occur in the identification of most relevant features (both in PCA and wavelets) for the different training-test subdivisions. Figure 6 shows the ranking distribution of the different PCA features in the ten executions of the NFMIS algorithm (one per training-test subdivision) for the Haar L3 Wavelet + PCA dataset. This widespread data cloud show that it is very difficult to detect the real relevancy of each of the PCA features extracted from the dataset, confirming that feature reduction in this dataset led to no improvement or preservation of the classification performance. A typical successful feature selection outcome would lead to a very diagonal point cloud, as different training-test subdivisions would lead to similar feature rankings. Similarly, Figure 7 shows the same results for the Haar L3 Wavelet dataset. Although relevancy of the features is slightly more clear for this problem, results in feature reduction were similarly fruitless.

In relation to the comparison between both types of wavelet transformations, it was observed that Haar L3 wavelet transform led to slightly better results than Db4 L2 wavelet transform. Due to the difficulty in claiming such results from a single 10-fold training test subdivision, this process was randomly repeated three times. T-test analysis was performed over the three executions (30 results in total) with significance level 0.1, confirming the first results shown in results in the shown in table 2. Wavelet features show to present better performance than PCA dimensionality reduction. MoreoverHaar L3 wavelet transform show to present better performance than DB4 L2 wavelet transform.

With respect to the MCI problem, results showed no significant difference between the use or not of dimensionality reduction. Thus, accuracy using original wavelet features showed a mean CV accuracy of 83.33%, while using PCA features, mean CV accuracy obtained reached 83.63%.

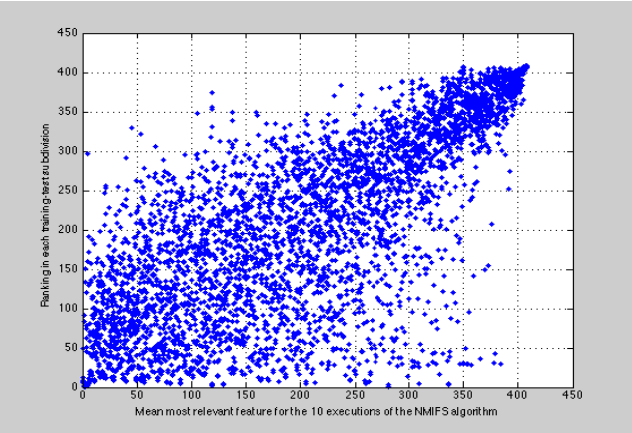

Figure 6. Ranking distribution of the different PCA features in the ten executions of the NFMIS algorithm (one per training-test subdivision) for the Haar L3 Wavelet + PCA dataset. X axis represents the mean ranking for the PCA features (independently of which one it is). Y axis represents the ten rankings obtained for such features.

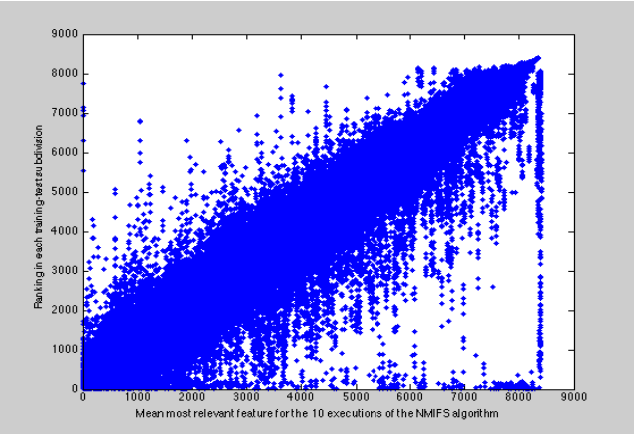

Figure 7. Ranking distribution of the different wavelets features in the ten executions of the NFMIS algorithm (one per training-test subdivision) for the Haar L3 Wavelet dataset.

A. *Sensitivity and specificity*

On the other hand, sensitivity and specificity were also calculated for the problem. In every medical study, each subject either has or does not have the disease. The test outcome can be positive (predicting that the person has the disease) or negative (predicting that the person does not have the disease). The test results for each subject may or may not match the actual status of the subject. Tables 6 show the sensitivity, specificity, PPV and NPV achieved in our studies for NAD classification. Table 7 shows the results attained in the NMCIAD study.It is noteworthy that as these measures belong to binary classification tests, they have to be adapted in the case of three classes. Thus, sensitivity and PPV referred to a certain disease are shown (MCI or AD).

Study	Sensitivity	Specificity	PPV	NPV
Db4 L2 Wavelet	0.953	0.947	0.947	0.952
Db4 L2 Wavelet PCA	0.953	0.929	0.929	0.951
Haar L3 Wavelet	0.963	0.961	0.965	0.962
Haar L3 Wavelet PCA	0.949	0.945	0.949	0.950

Table 6. Sensitivity, specificity, PPV and NPV of the NAD studies

Study	Sensitivity (MCI)	Sensitivity (AD)	Specificity	PPV (MCI)	PPV (AD)	NPV
Haar L3 + RBF Kernel	0.69	0.69	0.95	0.92	0.92	0.79

Table 7. Sensitivity, specificity, PPV and NPV of the NMCIAD study

VII. CONCLUSION

This paper deals with the important challenge of identification of Alzheimer's disease and the condition prior to dementia which is Mild Cognitive Impairment (MCI), developing intelligent classifiers, which using the information of magnetic resonance imaging, can successfully classify different patients according to their condition.

In this work Discrete Wavelet Transform (DWT) has been used for feature extraction, Principal Component Analysis (PCA) for feature reduction, and different methodologies, such as and Normalized Mutual Information Feature Selection (NMIFS) algorithm, as feature selection. Results have shown that the use of dimensionality reduction in this problem led to a worse classification accuracy, both using feature reduction (PCA) and feature selection (NMIRS algorithm), thus showing the importance of using all the available information to perform the classification, and the difficulty in selecting relevant information for the classification problem. This leads to a larger computational cost in the training of the classifiers, but that worth the cost in some cases according to the results attained for the NAD problem. SVM has been used as classification technique and results obtained have shown to be promising.

At this point it remains as future work the study on the optimal slices to perform the classification, the use of other dimensionality reduction algorithms which could attain a reduction in time complexity for the problem, and the study on other databases of the same algorithm proposed here.

ACKNOWLEDGMENT

This paper has been partially supported by the Spanish CICYT Project SAF2010-20558 and Regional Excellent Project P09-TIC-175476.

REFERENCES

[1] Scheltens, P., Frisoni, G. B., Galluzzi, S., Nobili, F. M., Fox, N. C., Robert, P. H., et al. (2003). Neuroimaging tools to rate regional atrophy, subcortical cerebrovascular disease, and regional cerebral blood flow and metabolism: consensus paper of the EADC. Neurol Neurosurg Psychiatry , 71, 1371-81.

[2] Koedan, E., Lehmann, M., Van der Flier, W. M., Scheltens, P., Pijnenburg, Y., Fox, N., et al. (2011). Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol , 21, 2618-2615.

[3] Westman, E., Cavallin, L., Muehlboeck, J. S., Zhang, Y., Mecocci, P., Vellas, B., et al. (2011). Sensitivity and Specificity of Medial Temporal Lobe Visual Ratings and Multivariate Regional MRI Classification in Alzheimer's Disease. *PLoS One, 6* (7).

[4] Jack, C. R., Barkhof, F., Barnstein, M. A., Cantillon, M., Cole, P. E., DeCarli, C., et al. (2011). Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic creiterion for Alzheimer's disease. *Alzheimer's & Dementia, 7*, 474-485.

[5] Koedan, E., Lehmann, M., Van der Flier, W. M., Scheltens, P., Pijnenburg, Y., Fox, N., et al. (2011). Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol, 21*, 2618-2615.

[6] Chaplot, S., Patnaik, L., & Jagannathan, N. R. (2006). Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical Signal Processing and Control, 1*, 86-92.

[7] El-Dahshan, E. A., Salem, A. M., & Younis, T. H. (2009). A hybrid technique for automatic MRI brain images classification. *Studia Univ. Babes-Bolyai, Informatica, 54* (1).

[8] Zhang, Y., Zhengchao, D., Wu, L., & Wang, S. (2011). A hybrid method for brain image classification. *Expert Systems with Applications, 38*, 10049-53

[9] Gorgel, P., Sertbas, A., Kilic, N., Ucan, O., & Osman, O. (2009). Mammographic mass classification using wavelet based suppont vector machine. *Journal of Electrical & Electronics Engineering, 9* (1), 867-875.

[10] Lee, J., Su, S., Huang, C., Wang, J. J., Xu, W., Wei, Y., et al. (2009). Combination of multiple features in support vector machine with principal component analsis in application for Alzheimer's disease diagnosis. *Lecture Notes in Computer Science, 5863*, 512-519.

[11] Pérez, D. A., Ramos, A., & Álvarez-Linera, J. (2010). *Neuroimagen. Diagnóstico, técnicas, secuencias 2*. Almirall.

[12] LONI. (2011). Retrieved November 5, 2011, from ADNI (Alzheimer's Disease Neuroimaging Initiative): adni.loni.ucla.edu

[13] Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. (1991). Comparing functional (PET) images: the assessment of significant change. *J. Cereb. Blood Flow Metab., 11*, 690-699.

[14] Nanni, L., & Lumini, A. (2008). Wavelet decomposition tree selection for palm and face authentication. *Pattern Recognition Letters, 29* (3), 343-353

[15] Chan, C., & Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligen Systems and Technology, 2* (3), 2:27:1-27:27.

[16] David Jaramillo, Ignacio Rojas, Olga Valenzuela, Ignacio Garcia, Alberto Prieto: Advanced systems in medical decision-making using intelligent computing. Application to magnetic resonance imaging. IJCNN 2012: 1-8

[17] S.S. Keerthi, C.J. Lin, "Asymptotic behavior of Support Vector Machines with Gaussian Kernel," Neural Computation 15(7) (2003) 1667-1689.

[18] B. Schoelkopf, A. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press (2002).

[19] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 42(2), (2012) 513-529

[20] Zhang D, Wang Y, Zhou L, Yuan H, Shen D; Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011 Apr 1;55(3):856-67

[21] Pablo A. Estévez, Michel Tesmer, Claudio A. Perez, and Jacek M. Zurada. 2009. Normalized mutual information feature selection. Trans. Neur. Netw. 20, 2 (February 2009), 189-201.