Multimodal image data fusion for Alzheimer's Disease diagnosis by sparse representation

Andrés ORTIZ [a,1], Daniel FAJARDO [a] Juan M. GÓRRIZ [b] Javier RAMÍREZ [b] and Francisco J. MARTÍNEZ-MURCIA [b]

[a] *Communications Engineering Department, Universidad de Málaga, Málaga - Spain*
[b] *Dpt. Signal Theory, Networking and Communications. Universidad de Granada, Granada - Spain*

Abstract. Alzheimer's Diasese (AD) diagnosis can be carried out by analysing functional or structural changes in the brain. Functional changes associated to neurological disorders can be figured out by positron emission tomography (PET) as it allows to study the activation of certain areas of the brain during specific task development. On the other hand, neurological disorders can also be discovered by analysing structural changes in the brain which are usually assessed by Magnetic Resonance Imaging (MRI). In fact, computer-aided diagnosis tools (CAD) that have been recently devised for the diagnosis of neurological disorders use functional or structural data. However, functional and structural data can be fused out in order to improve the accuracy and to diminish the false positive rate in CAD tools. In this paper we present a method for the diagnosis of AD which fuses multimodal image (PET and MRI) data by combining Sparse Representation Classifiers (SRC). The method presented in this work shows accuracy values up to 95% and clearly outperforms the classification outcomes obtained using single-modality images.

Keywords. Alzheimer's Disease, Information Fusion, Multimodal image, Sparse Representation Classifiers

Introduction

Alzheimer's disease (AD) diagnosis still remains a challenge, specially in the early stages of the disease when only mild cognitive symptoms are present. As the disease advances, brain functions become affected and it is more difficult to contain the neurodegeneration process. Moreover, the cause of AD is not well-known and available drugs only help to slow down the advance of the disease. Thus, early diagnosis results crucial to treat the disease effectively and may help to develop new drugs.

Image analysis is a common technique for AD diagnosis, as current medical imaging systems provide a vast amount of information about the subject under study. Image techniques can be classified into two groups. The first group include functional image tech-

[1]Corresponding Author: Andrés Ortiz, Communications Engineering Department, Universidad de Málaga, 29071 - Málaga, Spain; E-mail: aortiz@ic.uma.es

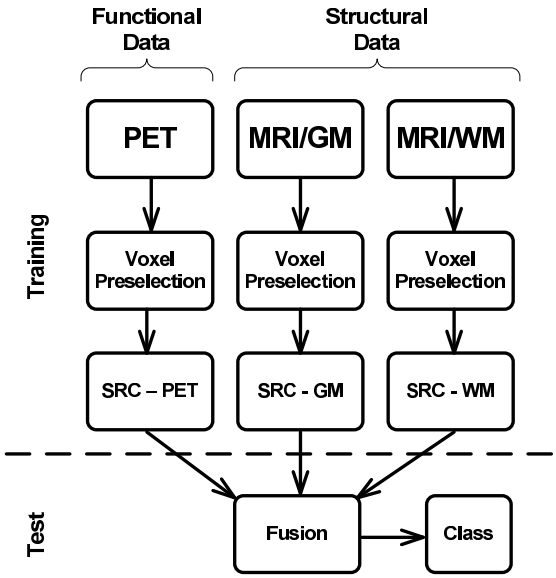


Figure 1. Block diagram of the proposed method. Functional and Structural data is fused by combining different SRC classifiers

niques, which aim to capture the activation levels of different parts of the brain when the patient is developing a specific task. Most CAD tools based on functional imaging can use Single Emission Computerized Tomography (SPECT) [1,2,3] or Positron Emission Tomography (PET) [8,4]. Other methods dealing with objective AD diagnosis are based on Brain Magnetic Resonance Images (MRI) [5,6,7]. In all cases, CAD systems aim to exploit the information contained in the images to learn patterns associated to cerebral neurodegeneration. However, structural changes in the early stages of AD are similar to the ones that appear due to the ageing process.

Previous works have provided good classification results (about 90%) using PET [8,4] or MRI images [5,6,7]. Nevertheless, functional and structural information can be jointly used to improve the classification performance. In this paper, we use a database containing multimodal PET/MRI image data from 138 patients to fuse functional and structural information by means of Sparse Representation Classifiers (SRC), outperforming previous results that have been obtained using single-mode images. Figure 1 shows the block diagram of the proposed method, where different SRC classifiers are trained using single-modality image data. Then, these classifiers are combined as explained in Section 2.1. After this introduction, the rest of the paper is organized as follows. Section 1 shows details on the database used in this work, as well as preprocessing and voxel preselection stages. Section 2 introduces the SRC classifiers and shows the technique used to combine them by means of the representation residuals. Then, experimental results are shown in Section 3 and finally, conclusions are drawn in Section 4.

1. Database

The database used in this work contains multimodal PET/MRI image data from 138 subjects, comprising 68 Controls (CN) and 70 AD patients from the ADNI database [9]. This repository was created to study the advance of the Alzheimer disease, collecting a vast amount of MRI and Positron Emission Tomography (PET) images as well as blood

biomarkers and cerebrospinal fluid analyses. The main goal of this database is to provide a way to the early diagnose of the Alzheimer disease. Patient's demographics are shown in Table 1.

Table 1. Patient Demographics

Evaluation	Sex (M/F)	Mean Age $\pm$ Std	Mean MMSE $\pm$ Std
NC	43/25	75.81 ± 4.93	29.06 ± 1.08
AD	46/24	75.33 ± 7.17	22.84 ± 2.91

1.1. Image Preprocessing

Since PET and MRI images have been used in this work, a different preprocessing techniques was used for each image modality. PET images were first spatially normalized according to a PET template using SPM [11]. Then, images were normalized in intensity in order to be able to compare them. This has been done as indicated in [8,10], where the mean value of the 0.1% voxels with the highest intensity levels is selected as normalization value. Moreover, voxels whose activation is below 50% have been removed and considered as background, as low activation voxels do not provide relevant information for classification. MRI images have to be treated in a different manner. In this case, MRI have been spatially normalized according to the VBM-T1 template and segmented into White Matter (WM) and Grey Matter (GM) tissues using the VBM toolbox for SPM [12]. The segmentation process provide information about GM and WM tissue distributions, which can be used to differentiate AD from CN patients [13,14,15]. Segmentation through SPM/VBM provides values in the range $[0,1]$ indicating the membership probability to a specific tissue.

1.2. Voxel preselection

Voxel preselection has been applied to each image modality separately, in order to remove low significance voxels and to reduce the dimension of the input space of the classifier in an early stage. This aims to build the SRC dictionary using the most informative voxels. This feature preselection was performed by means of Welch's t-test hypothesis testing separately for each image type. However, depending on the image modality, the value at each voxel position refers to a different magnitude, i.e. voxel values represent activation levels in PET images and membership probabilities in segmented tissues obtained from MRI.
Welch's t-test allows to test the difference in the means of two populations when the variances are unequal, and can be calculated using the following expression

$$I^t = \frac{I^{\mu}_{CN} - I^{\mu}_{AD}}{\sqrt{\frac{I^{\mu}}{N_{CN}} + \frac{I^{\mu}}{N_{CN}}}} \tag{1}$$

where I^{μ}_{CN} and I^{μ}_{AD} are the mean images for CN and AD respectively, I^{σ}_{CN} and I^{σ}_{AD} are the variance images for CN and AD, respectively and N_{CN}, N_{AD} are the number of CN and AD images, respectively. Mean images I^{μ}_{CN} and I^{μ}_{AD} are computed as

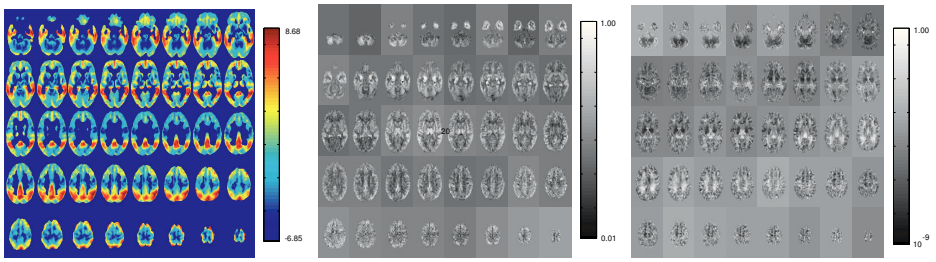
$$I^{\mu}_{CN} = \frac{1}{N_{CN}} \sum_{j=1}^{N_{CN}} I_j \quad I^{\mu}_{AD} = \frac{1}{N_{AD}} \sum_{j=1}^{N_{AD}} I_j \tag{2}$$

and variance images I^{σ}_{CN} and I^{σ}_{AD} are computed as

$$I^{\sigma}_{CN} = \sqrt{\frac{1}{N_{CN}} \sum_{j=1}^{N_{CN}} \left(I_j - I^{\mu}_{CN}\right)^2} \quad I^{\sigma}_{AD} = \sqrt{\frac{1}{N_{AD}} \sum_{j=1}^{N_{AD}} \left(I_j - I^{\mu}_{AD}\right)^2} \tag{3}$$

I^t represent the image composed by the t-value provided by the Welch's t-test for each image voxel, which is a significance measurement on the means difference. Hence, greater t-values correspond to lower p-values that allows to reject the null hyphotesis (which argues for equal means). In practice, only voxels with p-value below 0.05 (5% significance level) has been selected to build the SRC classifier.

In Figure 2 Welch's t-values for all voxels in the images are shown as different colours. Thus, depending on the threshold choosen for the *p-value*, a different number of voxels will be selected for classification. In other words, lower *p-values* will select less voxels for classification.



(a) (b) (c)

Figure 2. Welch's t-values (a) PET, (b) GM and (c) WM images.

It is worth mentioning that PET and MRI data can be jointly analysed as both modalities are available for the same patients. Moreover, patients age and Mini Mental State Examination scores (MMSE) are in the same range as shown in Table 1.

2. Sparse Representation Classifiers (SRC)

Sparse representation has been applied to different classification problems such as face recognition providing results comparable or even better than the ones provided by more complex classifiers such as Support Vector Machines (SVM) [16].

Sparse representation theory shows that sparse signals can be exactly reconstructed from a small number of linear measurements. However, SRCs use the training samples as measurements under the assumption that a sample of a specific class should lie in the subspace spanned by the training samples belonging to that class. SRC algorithm can be summarized as follows. Let $\mathbf{D} = [I_1, ..., I_n] \in \mathbb{R}^{m \times n}$ be the set of training samples organised by columns (i.e. n training samples of dimension m). Thus, a test sample $\mathbf{y} \in \mathbb{R}$ can be expressed as a linear combination of all the training samples as $\mathbf{y} = \mathbf{Dx}$. The sparsest solution of this equation can be found by solving the optimization problem

$$\mathbf{x_p} = min \|\mathbf{x}\|_0 \qquad \text{s.t.} \qquad \mathbf{y} = \mathbf{Dx} \tag{4}$$

However, this optimization problem cannot be solved in polynomial time and it is even difficult to approximate. Fortunately, if the solution is sparse enough, the solution provided by l_0 norm optimization is equivalent to the provided by the l_1 optimization problem, which is can be solved in polynomial time by standard linear programming methods [17]. Alternatively, it is possible to obtain an approximated solution by solving

$$\mathbf{x_p} = \min \|\mathbf{Dx} - \mathbf{y}\|^2 + \lambda \|\mathbf{x}\|_1 \tag{5}$$

where $\lambda \geq 0$ is a regularization parameter related to the sparsity of the solution.
As shown above, it is supposed that an image could be expressed as linear combination of all the training images belonging to the same class. Nevertheless, in real world applications such as image classification, not all the components from other classes will be zero. Thus, in order to improve the robustness and to diminish the representation error, residuals associated to each class are computed as

$$r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{D}\delta_\mathbf{i}(\mathbf{x_p})\|_2 \tag{6}$$

where $\delta_\mathbf{i}(\mathbf{x_p})$ is a function that selects the coefficients associated to class i by multiplying by zero the entries corresponding to other classes. Then, the class predicted for the test sample $\mathbf{y}$ by the SRC is computed as

$$class(\mathbf{y}) = \underset{i}{\operatorname{argmin}}\, r_i(\mathbf{y}) \tag{7}$$

2.1. Multimodal image data fusion by combining SRCs

Information from PET and MRI images can be fused in order to improve the classification performance by combining specialized SRCs. This way, different SRCs specialized in classifying single-modality images can be combined by means of their residuals. Indeed, class specific residuals can be taken as confidence measurements for predicted classes. Moreover, the prediction should be more accurate for greater residual differences. On the other hand, when residuals corresponding to the different classes are similar, the prediction will not be reliable. In our case, we build three single-modality

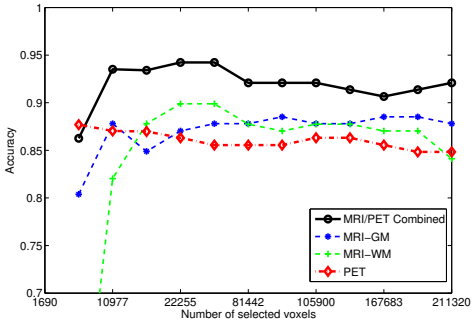


Figure 3. Classification accuracy obtained for the proposed method, fusing information from GM, WM and PET images. Figure also shows the accuracy values when classifying only with GM, WM or PET data.

classifiers (PET, MRI-GM and MRI-WM), and the residual of the test sample $\mathbf{y}$ obtained for the class i of the k classifier is defined as $r_i^k(\mathbf{y})$. Then, residuals corresponding to class i can be averaged in order to diminish the influence of classifiers providing similar residuals for different classes by using the equation

$$R_i(\mathbf{y}) = \frac{\sum_{k=1}^{3} r_i^k(\mathbf{y})}{3} \tag{8}$$

and class label of the test sample y can be computed as

$$class(\mathbf{y}) = \underset{i}{\operatorname{argmin}} R_i(\mathbf{y}) \tag{9}$$

3. Experimental Results

In this section, results from the classification experiments performed using the proposed algorithm are presented. Moreover, experiments using different values of p which provides different number of selected voxels by means of Welch's t-test are shown. Thus, Figure 3 shows classification accuracy values obtained using the proposed technique. Moreover, classification accuracy using the SRC technique independently for WM, GM and PET images are also shown. The experiments performed show a clear improvement when WM, GM and PET data is fused through SRC combination. Regarding the evaluation method, *k-fold* cross-validation technique with k=10 was used to assess the proposed method. Only training samples were used to compute preselected voxels and to build SRC dictionaries in order to avoid biasing. This way, accuracy values up to 94% are obtained by SCR combination while independent SRCs are not able to provide accuracy values above 90%.

In Figure 3, different number of voxels were selected by means of the threshold in the *p-value* used in each experiment. Number of selected voxels in 3 refers to the total number of voxels including WM, GM and PET selected voxels.

4. Conclusions

In this paper a method for fusing information from MRI and PET images is proposed. The presented approach consist on training independent SRCs for segmented MRI (classifiers based on GM and WM tissues are built) and PET images. Data fusion is performed by means of the SRC residuals, avoiding low-significance classification results due to similar residuals. Moreover, experiments using multimodal image data from the ADNI database has been performed, showing clear improvements from the baselines that use only GM, WM or PET data. The results obtained show classification accuracy values up to 94%, meaning at least 4% of improvement from the baseline.

Acknowledments

This work was partly supported by the MICINN under the TEC2012-34306 project and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Projects P09-TIC-4530 and P11-TIC-7103 and Universidad de Málaga, Campus de Excelencia Andalucía Tech.

References

[1] J. Ramírez, R. Chaves, J. M. Górriz, M. López, I. Álvarez, D. Salas-Gonzalez, F. Segovia, Computer Aided Diagnosis of the Alzheimer's Disease Combining SPECT-Based Feature Selection and Random Forest Classifiers, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC 2009), Orlando, USA, 2009.

[2] J. M. Górriz, F. Segovia, J. Ramírez, A. Lassl, D. Salas-Gonzalez, GMM-based SPECT Image Classification for the Diagnosis of Alzheimer's Disease, Applied Soft Computing, Vol. 11, No 2, pp. 2313-2325, 2011.

[3] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gmez-Ro, the Alzheimer's Disease Neuroimaging Initiative, Principal Component Analysis-Based Techniques and Supervised Classification Schemes for the Early Detection of the Alzheimer's Disease, Neurocomputing, Vol. 74, No. 8, pp. 1260-1271, 2011.

[4] F. Segovia, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, I. Álvarez, M. López, R. Chaves, the Alzheimer's Disease Neuroimaging Initiative, A comparative study of feature extraction methods for the diagnosis of Alzheimer's Disease using the ADNI database, Neurocomputing, Vol. 75, No. 1, pp. 64-71, 2012.

[5] R. Cuingnet R, E. Gerardin, J. Tessieras, G. Auzias, S. Lehricy, et al. Automatic classification eof patients with Alzheimers disease from structural MRI: a comparison of ten methods using the adni database. Neuroimage Vol. 56, No. 2, pp: 766781. 2010.

[6] M. Termenon, M. Graña. A Two Stage Sequential Ensemble Applied to the Classification of Alzheimer's Disease Based on MRI Features. Neural Processing Letters, Vol. 35, No. 1, pp:1-12. 2012.

[7] D. Chyzhyk, M. Graña, A. Savio, J. Maiora. Hybrid dendritic computing with kernel-lica applied to Alzheimers disease detection in MRI. Neurocomputing Vol. 75, No. 1, pp: 7277. 2012

[8] I.A. Illán, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, M.M. López, F. Segovia, R. Chaves, M. Gómez-Rio, C.G. Puntonet and the Alzheimers Disease Neuroimaging Initiative, 18F-FDG PET Imaging for Computer Aided Alzheimer's Diagnosis, Information Sciences, Vol. 181, No. 4, pp. 903-916, 2011.

[9] Alzheimers Disease Neuroimaging Initiative. Available: http://adni.loni.ucla.edu/. Accessed 2014 Mar 10.

[10] P. Padilla, M. López, J.M. Górriz, J. Ramírez, D. Salas-Gonzalez, I. A. Illán, NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease, IEEE Transactions on Medical Imaging, Vol. 31, No. 2, pp. 207-216, 2012.

[11] Statistical Parametric Mapping (SPM). shburner, J. and T.F.M Group, Functional Imaging Laboratory. Institute of Neurology, 12, Queen Square, Lonon WC1N 3BG, UK.

[12] Structural Brain Mapping Group Department of Psychiatry. Available: http://dbm.neuro.uni-jena.de/vbm8/VBM8-Manual.pdf. Accessed 2014 March 10.

[13] A. Ortiz, J. M. Górriz, J. Ramírez, F.J. Martínez-Murcia, LVQ-SVM Based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's Disease, Pattern Recognition Letters, Vol. 34, No. 14, pp. 1725-1733, 2013.

[14] M. Liu, D. Zhang, D. Shen, for the Alzheimers Disease Neuroimaging Initiative. Ensemble sparse classification of alzheimers disease. Neuroimage Vol. 60, pp. 1106-1116. 2012

[15] A. Ortiz, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, Improving MR Brain Image Segmentation using Self-Organizing Maps and Entropy-Gradient Clustering, Information Sciences, Vol. 262, No. 20, pp. 117-136, 2014.

[16] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma. Robust Face Recognition via Sparse Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 31, No. 2, pp. 210-227-2009.

[17] D. Chen, D. Donoho, M.Saunders. Atomic decomposition by basis pursuit. SIAM rev. Vol. 43, No. 1, pp. 129-159. 2001.