Learning Longitudinal MRI Patterns by SICE and Deep Learning: Assessing the Alzheimer's Disease Progression

Andrés Ortiz[1(✉)], Jorge Munilla[1], Francisco J. Martínez-Murcia[2], Juan M. Górriz[2], Javier Ramírez[2], and for the Alzheimer's Disease Neuroimaging Initiative

[1] Communications Engineering Department, University of Málaga, 29004 Málaga, Spain
aortiz@ic.uma.es
[2] Department of Signal Theory, Communications and Networking, University of Granada, 18060 Granada, Spain

Abstract. Automatic knowledge extraction from medical images constitutes a key point in the construction of computer aided diagnosis tools (CAD). This takes a special relevance in the case of neurodegenerative diseases such as the Alzheimer's disease (AD), where an early diagnosis makes the treatments easier and more effective. Moreover, the study of the evolution of the illness results crucial to differentiate the neurodegenerative process associated to the disease from the natural degeneration due to the ageing process. In this paper we present a method to construct longitudinal models from subjects using a series of MRI images. Specifically, the method presented here aims to model Gray matter (GM) variation at different brain areas of a subject across subsequent examinations, being possible to relate those regions which degenerate jointly. Hence, it allows determining variation patterns that differentiate controls from AD patients. Additionally, White matter (WM) density is also incorporated to the longitudinal model to complement the information provided by GM. The results obtained demonstrated the effectiveness of the method in the extraction of these patterns, that can be used to classify between Controls (CN) and AD subjects with 94% of accuracy, outperforming other previous methods.

1 Introduction

Alzheimer's Diasese (AD) is the most common cause of dementia among older people. Nowadays, 44 million people worldwide and it is expected AD affects

60 million people worldwide over the next 50 years. AD progressively causes the loss of nerve cells, whose symptoms usually start with mild memory problems, turning into severe brain damage in several years. The early diagnosis of AD results crucial, since there is no cure and currently developed drugs can only help to temporarily slow down the progression of the disease [3]. This way, the use of current functional or structural image systems have provided a way to explore new insights of the disease and to improve the diagnosis accuracy. In fact, many previous works used functional images such as Single Emission Computerized Tomography (SPECT) [8,14,22] or Positron Emission Tomography (PET) [1,23] to reveal functional differences between controls and AD patients and then to evaluate the loss of brain functions as the disease progress. On the other hand, since AD also causes structural changes in the brain, these can be figured out by means of Magnetic Resonance Image (MRI) analysis [5,6,16,17]. These works use GM or WM images on whole brain volume to classify controls and AD patients [5,6] or to compute Regions of Interest (ROI), searching for common patterns in Controls (CN) and AD subjects. On the other hand, the construction of neurodegeneration models to study the progression of the disease plays an important role for a better understanding of the neurodegeneration process. The main goal of this work, unlike previous works, is to model the degeneration process in different brain areas for individual subjects from a series of MRI images, rather than searching for static patterns in CN subjects or AD patients. This is addressed by modelling the regional grey matter density by means of the Sparse Inverse Covariance (SIC) and building a per-subject model that shows the relationship between areas in which the GM density covaries. This has important implications from the functional point of view, as positive covariance in GM evidences the existence of a fiber tract connection between these two areas [4,21]. These models, as explained later, can be further used to extract patterns that can be used to discriminate between CN and AD. Although most part of the information regarding the neurodegeneration process associated to AD is included in GM patterns, we also included WM density covariation to the model, as it provides complementary information that leverages the classification accuracy. On the other hand, we opted to use a deep-learning (DL) classifier instead of statistical classifiers such as SVM, since DL has demonstrated its effectiveness in extracting relevant information from noisy patterns [18]. In particular, we use a stacked denoising autoencoder to train a deep neural network which is fine-tuned by backpropagation.

The rest of the paper is organized as follows. Section 2 describes the database used in this work. Section 3 introduces the methods used in this work, including image preprocessing, density computation and the methods used for longitudinal modelling (and feature extraction) and classification in Subsects. 3.1, 3.2, 3.3, and 3.4, respectively. Section 4 shows details on the conducted experiments and the results obtained using longitudinal MRI data from the ADNI database. Finally, the main conclusions are drawn in Sect. 5.

2   Database

The database used in this work contains longitudinal MRI image data from 138 subjects, comprising 68 Controls (CN), and 70 AD patients from the ADNI database [2], comprising one evaluation each 6 months per subject in a period of three years. Thus, 6 evaluations are usually available for each subject, although in a few subjects, only 4 or 5 images were available in the database. This repository, which was created to study AD and provide a means for its early diagnosis, collects a vast amount of MRI and PET images as well as blood biomarkers and cerebrospinal fluid analyses. In this work, however, only MRI data have been used. Patients' demographics are shown in Table 1.

Table 1. Patient demographics. Mini-Mental State Examination Scores (MMSE) are indicated for diagnosis reference

Evaluation	Sex (M/F)	Mean age ± std (1st visit)	Mean MMSE ± std (1st visit)
NC	61/61	75.74 ± 5.12	29.22 ± 0.02
AD	47/49	75.19 ± 7.49	23.14 ± 1.94

3   Methods

3.1   Image Preprocessing

MRI images from the ADNI database have been spatially normalized according to the Voxel-based morphometry (VBM) T1 template and segmented into White Matter (WM) and Grey Matter (GM) tissues using the VBM toolbox for Statistical Parametric Mapping (SPM) software [7,24]. This ensures each image voxel corresponds to the same anatomical position. After image registration, all the images from the ADNI database were resized to $121 \times 145 \times 121$ voxels with voxel-sizes of $1.5\,\mathrm{mm}$ (sagittal) $\times 1.5\,\mathrm{mm}$ (coronal) $\times 1.5\,\mathrm{mm}$ (axial). MRIs are further segmented to obtain information about GM and WM tissue distributions, which can be used to differentiate AD from CN patients [13,16,17]. This process is guided by means of tissue probability maps of grey matter, white matter or cerebro-spinal fluid. A nonlinear deformation field is estimated that best overlays the tissue probability maps on the individual images. The tissue probability maps provided by the International Consortium for Brain Mapping (ICBM) are derived from 452 T1-weighted scans, which were aligned with an atlas space, corrected for scan inhomogeneities, and classified into grey matter, white matter and cerebro-spinal fluid. Segmentation through SPM/VBM provides values in the range $[0, 1]$ which denote the membership probability to a specific tissue.

3.2   Density Computation

Features used in this work are based on brain volumes at specific regions delineated by the Automatic Anatomical Labelling atlas (AAL). Although this brain atlas delimitates 116 regions, we just used 42 of them which are considered the most closely related to AD [11,19,25], discarding cerebellum and vermis regions. The volume corresponding to the i-region can be computed using the following expression,

$$\text{Vol}_i = \frac{\#voxels_i > thr}{1000} * \text{voxel size} \tag{1}$$

where thr denotes the probability threshold that indicates whether a voxel belongs to a specific tissue and voxel size is indicated in mm^3. Similarly, tissue density for each region can be computed as follows,

$$D_i = \frac{\#voxels_i > thr}{Vol_i} \tag{2}$$

Since SPM segmentation provides the membership probability values to a specific tissue for each voxel, the threshold thr, in Eqs. 1 and 2, indicates how the partial volume effect is taken into account. Thus, the lower the thr the less the relative importance of one tissue over the other, and for $thr=1$ no partial volume effect is taken into account. In this work, we selected $thr \geq 0.3$, meaning that voxels with a value ≥ 0.3 in the GM probability map are considered belonging to GM.

3.3   Sparse Inverse Covariance Estimation (SICE)

The characterization of the tissue covariation between brain areas is usually addressed by correlation analysis. Correlation between two variables is, however, a necessary but not sufficient condition for a causal relationship. In other words, a correlation between two variables captures pairwise information but does not guarantee that the occurrence of one variable causes the other, as the correlation could be originated by third-party effects. Partial correlation can be used then to effectively characterize the interaction of two brain areas varying together while factoring out the influence of the rest of the regions [20]. When applied to AD, this mathematical tool can be used to identify patterns associated to cerebral neurodegeneration by identifying conditional independence between regions given the rest as constant. Partial correlations, in turn, coincide with the off-diagonal entries of the inverse covariance matrix, also known as precision matrix. Partial correlations are thus usually computed using the Maximum Likelihood Estimation (MLE) of the inverse covariance matrix. MLE, however, is not recommended when the sample size is not considerably higher than the number of variables; e.g. the number of patients is not higher than the number of regions of interest. In those cases, and taking into account the inherent sparseness of the brain network [27], sparse computation can be employed [9]. Sparse Inverse Covariance Estimation (SICE), also known as known as Gaussian

graphical model or graphical LASSO (Least Absolute Shrinkage and Selection Operator), uses a regularization parameter that controls the number of zero entries. Next, we explain this with more detail.

Let $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ denote p-dimension vectors, corresponding to n samples measured at p selected ROIs, which follow a multivariate Gaussian distribution where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean and $\Sigma \in \mathbb{R}^{p \times p}$ is the covariance, then the empirical covariance is:

$$S = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x_i} - \mu)(\mathbf{x_i} - \mu)^T. \tag{3}$$

Let also $\Theta = \Sigma^{-1}$ be the inverse covariance (or precision) matrix, the maximum log likelihood estimation (or MLE) of Θ under a multivariate Gaussian model can be obtained as follows:

$$\widehat{\Theta} = \underset{\Theta \succ 0}{argmax} \left(\log(\det \Theta) - tr(S\Theta) \right), \tag{4}$$

where $tr(S\Theta)$ is the trace of $(S\Theta)$. If S is not singular, by deriving with regards to Θ and setting it to zero, we would get, as expected, that the estimate of the inverse covariance is $\widehat{\Theta} = S^{-1}$. However, because $p > n$ the empirical estimate of S becomes singular and a regularization must be applied so that a shrunken estimate of Θ can be obtained through a maximization of the penalized log likelihood function. In particular, an estimate of the inverse covariance matrix $\widehat{\boldsymbol{\Theta}}$ of the brain regions is computed by solving the following optimization problem using the algorithm proposed in [12]:

$$\widehat{\boldsymbol{\Theta}} = \underset{\Theta \succ 0}{argmax} \left(\log(\det \Theta) - tr(S\Theta) - \lambda ||\Theta||_1 \right), \tag{5}$$

where $|| \cdot ||_1$ denotes the "entrywise" l_1-norm regularization [11], which corresponds to the sum of absolute values of all the entries in a matrix, and $\lambda > 0$ is the pre-selected regularization parameter. The larger the value of λ the more sparse are the estimates for Θ provided by SICE. Conversely, when λ is small the constraint has little effect and SICE becomes the conventional MLE.

SICE reports conditional independence between two variables (given the other variables in the multivariate Gaussian distribution) and therefore, once it is computed, neurodegenerative patterns can be developed. Two brain regions are connected if and only if they are not conditionally independent.

3.4   Deep Learning-Based Classifier

Deep learning (DL) classifiers have demonstrated their effectiveness under noisy patterns. Moreover, they discover the underlying structure of the data and model it, allowing the computation of representative and/or discriminative features from the original patterns [10,18]. Thus, we used a deep-learning based classifier to learn patterns from the covariance values computed using the SICE method described above. Specifically, we used a stacked denoising autoencoder (SDA) to

train a deep neural network in a stepwise manner. Stacked autoencoders compute representative features in their most inner layer while trying to reconstruct the original sample. Note also that, as this is the layer with fewer neurons, it usually becomes the bottleneck in the neural structure. Subsequently, these most inner layer from each autoencoders are used as hidden layers of a deep neural network. Positive saturating linear transfer functions are used at the encoder part and Linear transfer function at the decoder part of the autoencoders. The learning rate for the scaled conjugate gradient descent training algorithm was selected to 0.005. These learning parameters make the network to converge in less than 500 iterations. Additionally, the training phase is stopped when no improvements in the loss function are obtained during 100 iterations to avoid overfitting.

At the top most, a single unit softmax layer is included to implement the classifier. This method provides the unsupervised pre-training stage of the deep neural structure, which is further fine-tuned by backpropagation using not the samples (as in the autoencoder phase) but the labels and a learning rate of 0.001. This way, the first stage using autoencoders is focused in minimizing the representation error provided by the features computed in the hidden layer, while the second stage tries to minimize the classification error. Figure 1 shows the neural structure used in this work, where the first stage consists of three autoencoders (with 1000, 500 and 30 neurons in the hidden layer) and the deep neural network contains 5 layers (1 visible layer, 3 hidden layers and 1 softmax layer).
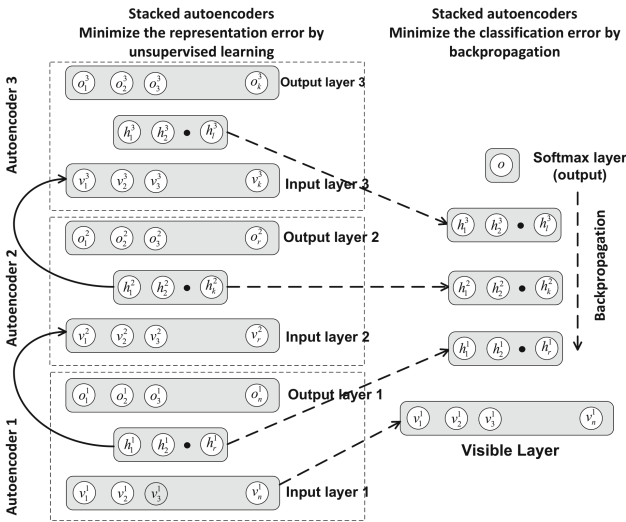

Fig. 1. Deep learning architecture used for classification. Unsupervised pre-training is addressed by three autoencoders using 1000, 500 and 30 neurons, respectively in the hidden layer. The multilayer perceptron is composed of autoencoder hidden layers and a softmax layer. Fine-tuning is performed by backpropagation.

In order to improve the robustness of the classifier and its noise immunity, inputs at the first autoencoder are corrupted by adding white gaussian noise [26] with a certain power level to keep a specific Signal-to-Noise ratio (SNR). This way, the autoencoder will be trained to reconstruct each input from noisy versions. In this work, we experimentally determined a $SNR = 30$ dB as it provides the best results, which means that the input units at the first autoencoder are corrupted with gaussian noise while keeping a SNR of 30 dB for each sample.

4   Experimental Results

For experimental results the deep learning architecture described in the previous section was applied to GM and WM data in the database as shown in Fig. 2. GM and WM voxels from 42 regions closely related to AD according to the literature [11] were used to compute the regional GM and WM densities for all the available examinations (visits) of each subject in the database. This results in a $N \times 42$ matrix for each subject, where N is the number of available examinations. These matrices are then used to compute the corresponding SICE; i.e. an estimate of the corresponding inverse covariance between regions.
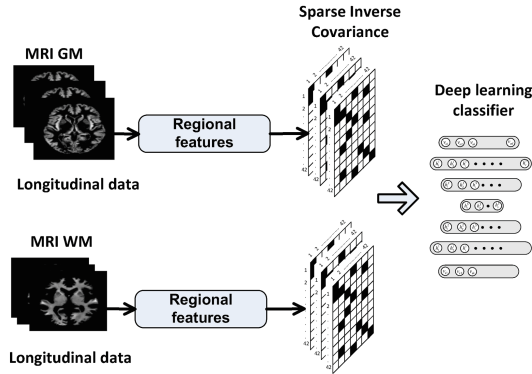

Fig. 2. Deep learning architecture applied to MRI-GM and MRI-WM data through SICE. Regional features refer to regional densities computed according to Eq. 2 and as indicated in [19]

Besides the discriminative information contained in SICE data (as we will see later), these also allows us to carry out certain exploratory analysis by identifying covarying regions. Figure 3 shows these data graphically for $\lambda = 0.05$, when $thr = 0.3$ is applied (see Subsect. 3.2). As shown in Subsect. 3.3, larger λ values provide sparser SICE matrices (i.e. more zero entries) and only stronger relationships between regions are kept. On the other hand, small λ values provide less sparse SICE matrices, capturing weak between-regions relationships that can introduce non-relevant information to the classifier. Hence, value $\lambda = 0.05$ was selected by

experimentation as it is a good trade-off between sparsity and computational burden, at the time it provided the best classification results. In the same way, the probability threshold to determine whether a voxel is considered to belong to GM or WM has also been determined by experimentation to obtain the best classification results. Thus, these values of the parameters have been used for the rest of this work. Finally, we also note that to feed the deep learning architecture only the lower triangle elements are employed, since SICE matrix is symmetric (diagonal elements are also discarded).
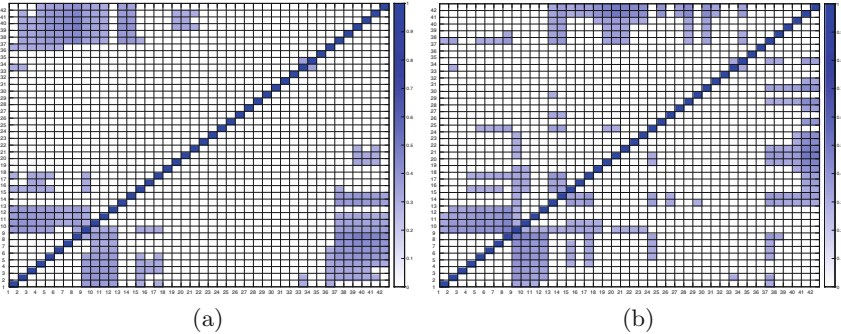

(a)   (b)

Fig. 3. Representation of the mean SICE ($\lambda = 0.05$) for CN (a) and AD (b) subjects thresholded by 0.3

4.1   Disease Patterns and Diagnostic Relevance of Brain Regions

Information regarding regions which covary without any third-party influence can be extracted from the SICE model. This shows regions whose GM/WM densities vary jointly during the examinations performed in the three-years period. This way, variations due to the ageing process as well as variations due to a possible neurodegenerative process are captured by the inverse covariance. Hence, analysing statistically significant differences between CN and AD SICEs by means of the Wilcoxon test, we found different between-region relationships. In Fig. 4, we show these relationships using BrainNet Viewer [15] where edges connect the 20 most significant covarying regions according to the rank provided by the Wilcoxon test. As shown in this figure, the parahipoccampal and hippocampal regions are related to the rectus and frontal orbital regions, indicating that the GM density in these regions varies jointly in AD.

4.2   Classification

Covariance values computed by SICE determine a per-subject, compact longitudinal model. In fact, we used these values as features to train the deep neural structure described above to classify subjects. The overall method here presented
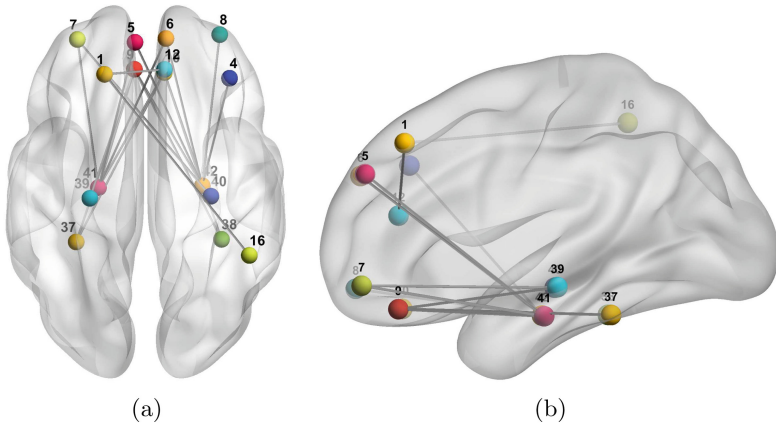
(a)   (b)

Fig. 4. Axial and sagittal views of the surface model of the WM of a representative brain, generated using brainviewer [15] with the 42 ROIs selected and their association (indicated with a line between points). An edge indicates covariation between regions. 1 = Left Precentral Gyrus; 4 = Right superior frontal gyrus (dorsolateral); 5 = Left superior frontal gyrus (dorsolateral); 6 = Right superior frontal gyrus (orbital); 7 = Left middle frontal gyris (lateral); 8 = Right middle frontal gyris (lateral); 9 = Left middle frontal gyrus (orbital), 10 = Right middle frontal gyrus (orbital); 12 = Right opercular part of inferior frontal gyrus; 16 = Right Orbital part of inferior frontal gyrus; 37 = Left Hippocampus; 38 = Right Hippocampus; 39 = Left parahippocampal gyrus; 40 = Right parahippocampal gyrus; 41 = Left amygdala; 42 = Right amygdala
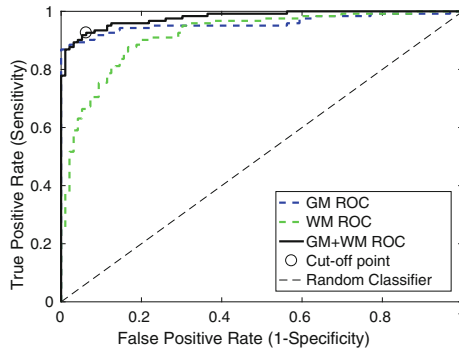

Fig. 5. ROC curve for SICE features computed from GM and WM. Classification performance using GM and WM features at the same time is also shown.

was assessed by k-fold cross-validation (specifically, k = 10), which ensures that training and testing subsets do no share any sample and estimates the generalization error. Hence, testing samples are never used during the training stage and double-dipping is avoided.

The ROC curve in Fig. 5 summarizes the classification performance using GM, WM and information from both tissues simultaneously. As expected, this

figure shows that most part of information linked to AD is contained in GM, while WM provides some extra knowledge not contained in GM that slightly leverages the classification performance.

The proposed classification method provides an Area under ROC curves (AUC) of 0.94 and 0.90 and for CN/AD classification using GM and WM, respectively. The combination of GM and WM information improves the AUC up to 0.98.

Additionally, experiments using different classification approaches were carry out. The results of these experiments, summarized in Table 2, show the superiority of the deep learning approach based on Stacked Denoising Autoencoders.

Table 2. Results obtained for different classification approaches using the same descriptors. Standard deviations obtained through k-fold (k = 10) cross-validation is indicated.

Method	Accuracy	Sensitivity	Specificity	AUC
Decision tree	0.90 ± 0.11	0.86 ± 0.10	0.97 ± 0.04	0.90
Linear SVM	0.90 ± 0.09	0.88 ± 0.09	0.93 ± 0.07	0.92
Linear LP-SVM	0.85 ± 0.09	0.84 ± 0.10	0.87 ± 0.12	0.85
DL-SDA	$\mathbf{0.94 \pm 0.07}$	$\mathbf{0.90 \pm 0.08}$	$\mathbf{0.96 \pm 0.07}$	**0.98**

5   Conclusions and Future Work

This paper proposes a method to compute longitudinal models from tissue densities at different brain regions defined by the AAL atlas. In the conducted experiments these densities were used as features to assess their capability to distinguish between CN and AD subjects from longitudinal MRI data. Thus, the exploratory analysis performed using the SICE model on density data obtains discriminant regions corresponding to those found in medical literature, such as the hippocampus in both hemispheres. In addition to these, relationships between hippocampal areas and other regions such as the inferior temporal gyrus have also been revealed. Finally, classification experiments (using k-fold cross-validation) were performed using SICE features computed from GM and WM density values, showing an accuracy of up to 94%, and AUC of up to 0.98.

As future work we plan to include other biomarkers jointly with GM and WM densities in order to study possible interdependencies in CN and AD subjects. The development of new methods to corrupt the inputs, specially for MRI data, becomes also particularly relevant to improve the robustness of the DL classifiers. Lastly, the proposed method can be extended to MCI subjects, which is an interesting task that could reveal neurodegeneration patterns moving towards the early diagnosis of AD.

References

1. Alvarez, I., Górriz, J., Ramírez, J., Salas-González, D., Lopez, M., Segovia, F., Chaves, R., Gomez-Rio, M., García-Puntonet, C.: 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. Inf. Sci. **184**(4), 903–916 (2011)
2. Alzheimer's Disease Neuroimaging Initiative. http://adni.loni.ucla.edu/. Accessed May 2017
3. Alzheimer's Disease Society: Factsheet: Drug Treatments for Alzheimer's Disease, February 2017. https://www.alzheimers.org.uk. Accessed May 2017
4. Chen, Z.J., He, Y., Rosa-Neto, P., Germann, J., Evans, A.C.: Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. Cereb. Cortex **18**(10), 2374–2381 (2008)
5. Chyzhyk, D., Graña, M., Savio, A., Maiora, J.: Hybrid dendritic computing with kernel-lica applied to Alzheimer's disease detection in MRI. Neurocomputing **75**(1), 72–77 (2012)
6. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M., Chupin, M., Benali, H., Colliot, O., 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**(2), 766–781 (2010)
7. Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, Cambridge (2007)
8. Górriz, J., Segovia, F., Ramírez, J., Lassl, A., Salas-González, D.: Gmm based spect image classification for the diagnosis of Alzheimer's disease. Appl. Soft Comput. **11**, 2313–2325 (2011)
9. Hilgetag, C., Kötter, R., Stephan, K., Sporns, O.: Computational methods for the analysis of brain connectivity. In: Ascoli, G.A. (ed.) Computational Neuroanatomy, pp. 295–335. Humana Press, New York (2002)
10. Hinton, G.: Where do features come from? Cogn. Sci. **38**(6), 1078–1101 (2014)
11. Huang, S., Li, J., Sun, L., Jun, L., Wu, T., Chen, K., Fleisher, A., Reiman, E., Jieping, Y.: Learning brain connectivity of Alzheimer's disease from neuroimaging data. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (eds.) Advances in Neural Information Processing Systems 22, pp. 808–816. Curran Associates Inc., Red Hook (2009)
12. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections. Arizona State University (2009). http://www.public.asu.edu/jye02/Software/SLEP
13. Liu, M., Zhang, D., Shen, D.: Ensemble sparse classification of Alzheimer's disease. NeuroImage **60**(2), 1106–1116 (2012)
14. López, M., Ramírez, J., Górriz, J., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R., Padilla, P., Gómez-Río, M.: Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease. Neurocomputing **74**(8), 1260–1271 (2011). Selected Papers from the 3rd International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC 2009)

15. Mingrui, X., Jinhui, W., Yong, H.: BrainNet viewer: a network visualization tool for human brain connectomics. PLoS ONE **8**(7), e68910 (2013)

16. Ortiz, A., Górriz, J.M., Ramírez, J., Martinez-Murcia, F.J., Alzheimer's Disease Neuroimaging Initiative: Automatic roi selection in structural brain MRI using som 3D projection. PLOS ONE **9**(4), e93851 (2014)

17. Ortiz, A., Górriz, J.M., Ramírez, J., Martínez-Murcia, F.J.: Lvq-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease. Pattern Recogn. Lett. **34**(14), 1725–1733 (2013)

18. Ortiz, A., Munilla, J., Górriz, J.M., Ramírez, J.: Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. Int. J. Neural Syst. **26**(07), 1650025 (2016)

19. Ortiz, A., Munilla, J., Illán, I.Á., Górriz, J.M., Ramírez, J., Alzheimer's Disease Neuroimaging Initiative: Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's disease diagnosis. Front. Comput. Neurosci. **9**, 132 (2015)

20. Pourahmadi, M.: High-Dimensional Covariance Estimation, 1st edn. Wiley, Hoboken (2013)

21. Raamana, P.R., Weiner, M.W., Wang, L., Beg, M.F.: Thickness network features for prognostic applications in dementia. Neurobiol. Aging **36**(1), S91–S102 (2015)

22. Ramirez, J., Chaves, R., Gorriz, J.M., Lopez, M., Alvarez, I.A., Salas-Gonzalez, D., Segovia, F., Padilla, P.: Computer aided diagnosis of the Alzheimer's disease combining spect-based feature selection and random forest classifiers. In: Proceedings of IEEE Nuclear Science Symposium Conference Record (NSS/MIC), pp. 2738–2742 (2009)

23. Segovia, F., Górriz, J.M., Ramírez, J., Álvarez, I., Jiménez-Hoyuela, J.M., Ortega, S.J.: Improved Parkinsonism diagnosis using a partial least squares based approach. Med. Phys. **39**(7), 4395–4403 (2012)

24. Structural Brain Mapping Group: Department of Psychiatry. http://dbm.neuro.uni-jena.de/vbm8/VBM8-Manual.pdf. Accessed Oct 2014

25. Sun, L., Patel, R., Liu, J., Chen, K., Wu, T., Li, J., Reiman, E., Ye, J.: Mining brain region connectivity for Alzheimer's disease study via sparse inverse covariance estimation. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1335–1344. ACM, New York (2009)

26. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**, 3371–3408 (2010)

27. Zalesky, A., Fornito, A., Bullmore, E.: Network-based statistic: identifying differences in brain networks. NeuroImage **53**(4), 1197–1207 (2010)