

# Brain Connectivity Hyper-Network for MCI Classification

Biao Jie<sup>1,2</sup>, Dinggang Shen<sup>3</sup>, and Daoqiang Zhang<sup>1,\*</sup>

<sup>1</sup> Dept. of Computer Science and Engineering,

Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

<sup>2</sup> Dept. of Computer Science and Technology, Anhui Normal University, Wuhu 241000, China

<sup>3</sup> Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA  
dqzhang@nuaa.edu.cn

**Abstract.** Brain connectivity network has been used for diagnosis and classification of neurodegenerative diseases, such as Alzheimer's disease (AD) as well as its early stage, i.e., mild cognitive impairment (MCI). However, conventional connectivity network is usually constructed based on the pairwise correlation among brain regions and thus ignores the higher-order relationship among them. Such information loss is unexpected because the brain itself is a complex network and the higher-order interaction may contain useful information for classification. Accordingly, in this paper, we propose a new brain connectivity hyper-network based method for MCI classification. Here, the connectivity hyper-network denotes a network where an edge can connect more than two brain regions, which can be naturally represented with a hyper-graph. Specifically, we first construct connectivity hyper-networks from the resting-state fMRI time series using sparse representation modeling. Then, we extract three sets of the brain-region specific features from the connectivity hyper-networks, and exploit a manifold regularized multi-task feature selection method to jointly select the most discriminative features. Finally, we use multi-kernel support vector machine (SVM) for classification. The experimental results demonstrate the efficacy of our proposed method for MCI classification with comparison to the conventional connectivity network based methods.

## 1 Introduction

Alzheimer's disease (AD) is the most common form of dementia in elderly population worldwide. It is predicted that the number of affected people will double in the next 20 years, and 1 in 85 people will be affected by 2050 [1]. Mild cognitive impairment (MCI) is a prodromal stage of AD and has gained a great deal of attention recently due to its high progression rate to AD. Existing studies have shown that MCI subjects progress to clinical AD at an annual rate of approximately 10% to 15%, while the normal controls (NC) develop dementia at an annual rate of 1% to 2% [2]. Thus, accurate diagnosis of MCI is very important for possible early treatment and possible delay of the progression of disease.

---

\* Corresponding author.

In the past decade, modern imaging techniques have provided efficient ways to explore the structural and functional interaction of the human brain, which enables us to better understand the pathological underpinnings of neurological disorder. These interaction patterns, which can be characterized via connectivity networks, have been applied recently to disease diagnosis and classification [3] [4]. However, to the best of our knowledge, existing (functional) connectivity networks are usually constructed by measuring the *pairwise* correlations between paired brain regions, and hence reflect only the second-order relationship among brain regions, ignoring the higher-order relationship among them (i.e., the interaction among more than two brain regions). Such information loss is unexpected because the brain itself is a complex network and the higher-order interaction may contain useful information for classification. It is worth noting that Huang et al. in [5] adopted partial correlation based method to identify brain connectivity from PET data, which measures the correlation between two brain regions after factoring out the influence of other regions.

In this paper, we address this issue by proposing a new brain connectivity hyper-network based method for MCI classification. Here, the connectivity hyper-network is a network where each edge can include more than two nodes, which can be naturally represented with a hyper-graph. Fig. 1 shows the flowchart of our proposed method. Specifically, we first construct the connectivity hyper-networks from the resting-state fMRI time series using sparse representation modeling [6]. Then, we extract three sets of the brain-region specific features (i.e., clustering coefficients) from the constructed connectivity hyper-networks. Furthermore, we exploit a manifold regularized multi-task feature selection (M2TFS) method proposed in [7] to jointly select the most discriminative features from those three sets of extracted features. Finally, we use the multi-kernel support vector machine (SVM) technique [8] for classification. The proposed method is evaluated on a real MCI dataset, demonstrating promising results compared to the conventional connectivity network based methods.

## 2 Materials and Methodology

### 2.1 Subjects

In this study, 12 amnesic MCI patients (6 males and 6 females) and 25 normal controls (9 males and 16 females) were recruited. All the recruited subjects were diagnosed by expert consensus panels. Data acquisition was performed using a 3.0-Tesla GE Signa EXCITE scanner. The resting-state fMRI (R-fMRI) images of each participant were acquired with the following parameters: flip angle= $77^\circ$ , TR/TE=2000/32ms, imaging matrix= $64 \times 64$ , FOV= $256 \times 256 \text{ mm}^2$ , 34 slices, 150 volumes, and voxel thickness=4mm. During scanning, all subjects were instructed to keep their eyes open and stare at a fixation cross in the middle of the screen, which lasted for 5 minutes.

## 2.2 Data Pre-Processing

Data pre-processing was performed using Statistical Parametric Mapping software package (SPM8) (<http://www.fil.ion.ucl.ac.uk/spm/software/SPM8/>). The first 10 R-fMRI images were discarded to ensure magnetization equilibrium. The remaining 140 images were first corrected before they were realigned to the first volume of the remaining images for head motion correction. Regression of nuisance signals, including ventricle and white matter signals, as well as six head-motion profiles were performed. The brain space of R-fMRI scans for each subject was then parcellated into 116 regions of interesting (ROI) based on the Automated Anatomical Labeling (AAL) atlas [9]. The mean R-fMRI time series was then computed for each subject by averaging the R-fMRI time series over all voxels in each ROI.

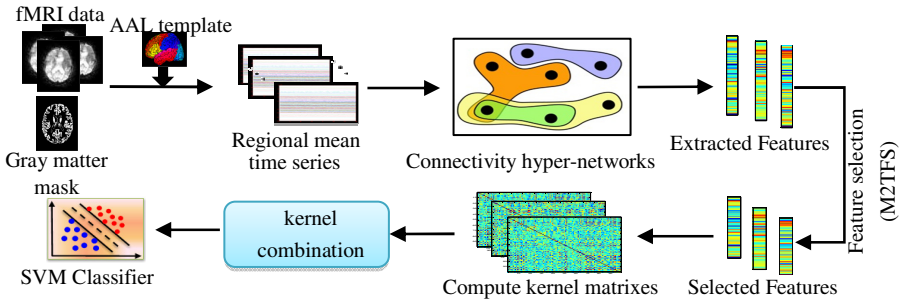


Fig. 1. Flowchart of the proposed method

## 2.3 Hyper-graph

It is known that graph is a powerful tool for representing the pairwise relationships between paired nodes. Actually, besides pairwise relationships, in many applications (e.g., functional connectivity among brain regions), there may exist higher-order relationships, which cannot be represented by the conventional graphs. To overcome this problem, hyper-graph has been proposed to characterize the higher-order relationship among nodes. In general, a hyper-graph is an extended graph where an edge (called hyper-edge in hyper-graph) can connect more than two nodes [10].

Denote a hyper-graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with a node (vertex) set  $\mathcal{V}$  and a hyper-edge set  $\mathcal{E}$ , we can represent  $\mathcal{G}$  using a  $|\mathcal{V}| \times |\mathcal{E}|$  incidence matrix  $H$  with elements  $h(v, e) = 1$  if  $v \in e$  and 0 otherwise. Here,  $v \in \mathcal{V}$  is a node/vertex and  $e \in \mathcal{E}$  is a hyper-edge of  $\mathcal{G}$ . It is worth noting that the conventional graph is a special kind of hyper-graph with each hyper-edge only containing two nodes. Fig. 2 illustrates an example of a hyper-graph. In the literature, hyper-graph has been successfully applied to a variety of problems, such as protein function prediction [11].

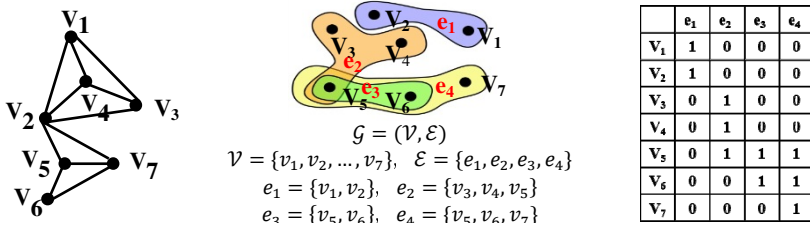
## 2.4 The Construction of Connectivity Hyper-network

In our study, hyper-networks are first constructed from R-fMRI time series using sparse representation modeling [6]. Specifically, denote  $X = [x_1, x_2, \dots, x_M]^T \in R^{M \times d}$

as a training subject with a total of  $M$  ROIs, where  $x_m$  represents the regional mean time series of the  $m$ -th ROI. Then, the regional mean time series of each ROI (i.e.,  $x_m$ ) can be regarded as a response vector, and can be estimated by a linear combination of times series of other  $M - 1$  ROIs, i.e.,

$$x_m = A_m \alpha_m + \varepsilon_m, \quad m = 1, 2, \dots, M \tag{1}$$

where  $A_m = [x_1, x_2, \dots, x_{m-1}, 0, x_{m+1}, \dots, x_M]$  denote a data matrix including all time series except the  $m$ -th ROI (we put 0 in its location),  $\alpha_m$  is the weight vector that quantifies the degree of influence of other ROIs to the  $m$ -th ROI, and  $\varepsilon_m \in R^d$  is a noise term. Note that zero elements in weight vector imply that the corresponding ROIs are insignificant in accurate estimation of a term of time series.



**Fig. 2.** Hyper-graph vs. graph. Left: a conventional graph in which two nodes are connected together by an edge. Middle: a hyper-graph in which each hyper-edge can connect more than two nodes. Right: the incidence matrix for the hyper-graph in the middle.

The sparse solution  $\alpha_m$  can be approximately recovered by solving a standard  $l_1$ -norm regularized optimization problem with the following objective function:

$$\min_{\alpha_m} \|x_m - A_m \alpha_m\|_2 + \lambda \|\alpha_m\|_1 \tag{2}$$

where  $\lambda > 0$  is a regularization parameter controlling the sparsity of the model. Different  $\lambda$  value corresponds to different sparsity solution, and the larger  $\lambda$  value means a sparser model, i.e., more elements in  $\alpha_m$  are zero.

In our method, for each subject, a hyper-network is constructed using sparse representation modeling for each ROI used as a node, and a hyper-edge  $e_m$  includes a centroid ROI (i.e.,  $m$ -th ROI) and those ROIs with the corresponding non-zero elements in the weight vector  $\alpha_m$  computed in Eq. (2). Instead of generating a single hyper-edge for each ROI (node), we generate a group of hyper-edges by varying the  $\lambda$  value in a specified range. Specifically, in our experiment, we vary  $\lambda$  value from 0.1 to 0.9 with an incremental step of 0.1.

### 2.5 Feature Extraction

After constructing the connectivity hyper-networks, another important problem is how to extract meaningful features from hyper-networks. In the conventional connectivity networks, clustering coefficient is one of the most widely used features for quantifying the degree to which nodes in a network tend to cluster together [12]. Recently, the concept of clustering coefficient has been extended from the conventional network to hyper-network [11]. In this study, three different kinds of clustering

coefficients defined in [11] are used to extract features from brain connectivity hyper-networks.

Given a connectivity hyper-network  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , let  $S(v) = \{e_i \in \mathcal{E} : v \in e_i\}$  represent the hyper-edges adjacent to the node  $v$ . Let  $N(v) = \{u \in \mathcal{V} : \exists e \in \mathcal{E}, u, v \in e\}$  be the nodes that are neighbors of the node  $v$ . Then three different clustering coefficients on the node  $v$  can be, respectively, defined as:

$$HCC^1(v) = \frac{2 \sum_{u,t \in N(v)} I(u,t,-v)}{|N(v)|(|N(v)|-1)} \quad (3)$$

$$HCC^2(v) = \frac{2 \sum_{u,t \in N(v)} I'(u,t,v)}{|N(v)|(|N(v)|-1)} \quad (4)$$

$$HCC^3(v) = \frac{2 \sum_{e \in S(v)} (|e|-1) - |N(v)|}{|N(v)|(|S(v)|-1)} \quad (5)$$

where  $I(u,t,-v)=1$  if there exists  $e_i \in \mathcal{E}$  such that  $u, t \in e_i$  but  $v \notin e_i$ , and 0 otherwise;  $I'(u,t,v)=1$  if there exists  $e_i \in \mathcal{E}$  such that  $u, t, v \in e_i$ , and 0 otherwise.

The  $HCC^1$  computes the number of adjacent nodes that have connections not facilitated by node  $v$ , under the hypothesis that these connections are more robust because ‘‘independent’’ evidence is provided. Conversely, the  $HCC^2$  calculates the number of adjacent nodes that have connections facilitated by node  $v$ , considering that those nodes may share a function with each other and node  $v$ . The  $HCC^3$  calculates the amount of overlap amongst adjacent hyper-edges of node  $v$ .

Finally, for each of the three definitions of clustering coefficient defined in Eqs. (3)-(5), we extract a set of features from the brain connectivity hyper-networks, respectively. So, finally we can obtain three sets of features for each subject.

## 2.6 Feature Selection and Classification

Features extracted from connectivity hyper-networks potentially include irrelevant or redundant features for subsequent MCI classification. On the other hand, different types of clustering coefficient features reflect the local clustering properties of the brain connectivity hyper-network in different views. In order to find the intrinsic common subset of features (i.e., from same brain regions) relevant to MCI disease, we adopt the manifold regularized multi-task feature selection (M2TFS) method proposed in [7] to jointly select the most discriminative features, where each task focuses on the classification of each type of features.

Let  $Z^k = [z_1^k, \dots, z_n^k, \dots, z_N^k]^T \in R^{N \times M}$ , represent three sets of features obtained from totally  $N$  training subjects, each with  $M$  ROIs. Here,  $z_n^k = [HCC^k(v_i)]_{i=1:M} \in R^M$  represents the vector of clustering coefficients from the  $n$ -th training subject according to the above definition of  $HCC^k$ . Let  $Y = [y_1, y_2, \dots, y_N]^T \in R^N$  be the response vector for those  $N$  training subjects, where  $y_i$  is the corresponding class label (i.e., MCI patient or normal control (NC)). Then, the M2TFS method optimizes the following objective function [7]:

$$\min_W \frac{1}{2} \sum_{k=1}^3 \|Y - Z^k w^k\|_2^2 + \beta \sum_{k=1}^3 (Z^k w^k)^T L^k (Z^k w^k) + \gamma \|W\|_{2,1} \quad (6)$$

where  $L^k$  represents a combinatorial Laplacian matrix on the  $k$ -th feature set,  $W = [w^1, w^2, w^3] \in R^{M \times 3}$  is the weight matrix, and  $\|W\|_{2,1} = \sum_{j=1}^M \|w_j\|_2$  is the group sparsity regularizer encouraging features to be selected jointly. Here,  $w_j$  is the

$j$ -th row vector of  $W$ . The parameters  $\beta$  and  $\gamma$  are the corresponding regularization coefficients, whose values are determined via inner cross-validation on training data.

Finally, we adopt the multi-kernel SVM technique for classification. Specifically, for each set of clustering coefficient features, a linear kernel is first calculated based on the selected features. Then, the multi-kernel SVM [8] is adopted to combine those kernels for MCI classification.

## 3 Results

### 3.1 Classification Performance

In our experiments, the leave-one-out (LOO) cross-validation was used to evaluate the performances of classifier due to limited sample size. We evaluated the performance of different methods by measuring the classification accuracy, sensitivity (i.e., the proportion of patients that are correctly predicted), specificity (i.e., the proportion of normal controls that are correctly predicted), and area under receiver operating characteristic (ROC) curve (AUC). The proposed method was compared with the Conventional (pairwise Pearson-correlation-based) Network classification method where Clustering Coefficients were extracted as features (denoted as CN\_CC). For extensive comparison, we also compare the results using the single clustering coefficients extracted from our connectivity Hyper-Networks according to the definitions on Eq. (3)-(5) (denoted as HN\_HCC<sup>1</sup>, HN\_HCC<sup>2</sup> and HN\_HCC<sup>3</sup>, respectively). Note that, in all comparison methods, LASSO-based method is used to perform feature selection and a linear SVM classifier is used for classification. Classification results of all methods are summarized in Table 1. Fig. 3 plots the ROC curves for different methods

**Table 1.** Classification performances of different methods

Method	Accuracy	Sensitivity	Specificity	AUC
*CN_CC	62.2	41.7	72.0	0.54
*HN_HCC <sup>1</sup>	75.7	41.7	92.0	0.75
HN_HCC <sup>2</sup>	81.1	75.0	84.0	0.80
HN_HCC <sup>3</sup>	89.2	83.3	92.0	0.93
<b>Proposed</b>	<b>94.6</b>	<b>91.7</b>	<b>96.0</b>	<b>0.96</b>

Note that: \* denotes that there has significant difference (i.e., p-value<0.05) in classification accuracy between proposed method and others with McNemar's test.

As shown in Table 1 and Fig. 3, the proposed method significantly outperforms the other methods. Specifically, the proposed method yields a classification accuracy of 94.6% and an AUC of 0.96, indicating excellent diagnostic power. Also, by comparing HN-HCC<sup>1</sup>, HN-HCC<sup>2</sup> and HN-HCC<sup>3</sup> with CN-CC, we can see that all the three hyper-network based methods consistently outperform the conventional (pairwise Pearson-correlation-based) network classification method, which shows the advantages of connectivity hyper-network over the conventional network in characterizing brain connectivities. Moreover, both Table 1 and Fig. 3 indicate that our proposed

method consistently outperforms three methods using single type of clustering coefficients, which implies that different clustering properties of the connectivity hyper-network contain complementary information and thus can be integrated for further improving the classification performance.

On the other hand, based on the same dataset, we also compared our proposed method with the recent state-of-the-art methods for connectivity-network based MCI classification, including Wee et al.'s method [4] and Jie et al.'s method [3]. The accuracy of 91.9% and AUC of 0.90 were reported in [4], while in [3] the accuracy of

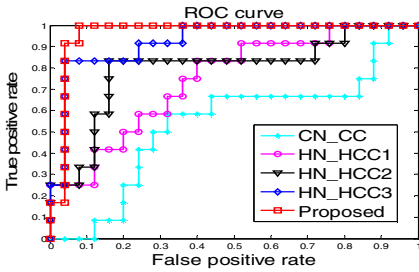


Fig. 4. ROC curves of 5 different methods

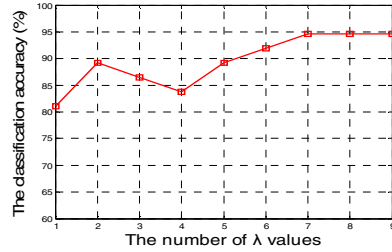
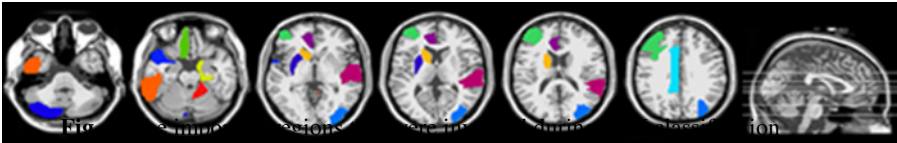


Fig. 4. Classification accuracy w.r.t. the use of different number of λ values.



91.9% and AUC 0.87 were reported. Obviously, those results are inferior to the performance of our proposed method.

Furthermore, to investigate the effect of different number of λ values (in eq. (2)) on classification performance of proposed method, we test 9 groups of λ values, i.e., {0.1}, {0.1, 0.2}, {0.1, 0.2, 0.3}, ..., {0.1, 0.2, ..., 0.9}. Fig. 4 gives the classification results. The result shows the classification accuracy can be improved with increase of number of λ values.

### 3.2 Brain Regions Involved in Classification

In this subsection, we investigated the important features (corresponding to ROIs) selected by our method. Since the selected features are different in each LOO cross-validation fold, we choose those features that were always selected in all folds. Fig. 5 plots the important ROIs detected by our method. These brain regions include frontal gyrus, rectus gyrus, cingulate, parahippocampal gyrus, occipital gyrus, temporal gyrus and temporal pole, which are found to be relevant with MCI pathology [13, 14]. Besides, two regions from cerebellum were also selected. Recent studies show that although the cerebellum might not be directly associated with the origin of AD, it may provide useful information related to its prognosis [15].

## 4 Conclusion

In summary, we have proposed a new (functional) connectivity hyper-network based classification method, which can indicate the higher-order relationship among brain regions. This is completely different from the conventional methods, where only the pairwise relationship can be measured via Pearson correlation. Experimental results on MCI classification indicate that our proposed method can *not only* improve brain disease classification, *but also* facilitate the detection of disease-relevant structures.

**Acknowledgments.** This work was partially supported by JiangsuSF (BK20130034), SRFDP (20123218110009), NUAARF (NE2013105), and NIH grants (B006733, EB008374, EB009634, AG041721).

## References

1. Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H.M.: Forecasting the global burden of Alzheimer's disease. *Alzheimers & Dementia* 3, 186–191 (2007)
2. Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., Ritchie, K., Rosser, M., Thal, L., Winblad, B.: Current concepts in mild cognitive impairment. *Arch. Neurol-Chicago* 58, 1985–1992 (2001)
3. Jie, B., Zhang, D., Gao, W., Wang, Q., Wee, C.Y., Shen, D.: Integration of Network Topological and Connectivity Properties for Neuroimaging Classification. *IEEE Trans. Biomed. Eng.* 61, 576–589 (2014)
4. Wee, C.-Y., Li, Y., Jie, B., Peng, Z.-W., Shen, D.: Identification of MCI Using Optimal Sparse MAR Modeled Effective Connectivity Networks. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part II. LNCS*, vol. 8150, pp. 319–327. Springer, Heidelberg (2013)
5. Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., Chen, K., Reiman, E.: Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *Neuroimage* 50, 935–949 (2010)
6. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust Face Recognition via Sparse Representation. *Ieee T. Pattern Anal.* 31, 210–227 (2009)
7. Jie, B., Zhang, D., Cheng, B., Shen, D.: Manifold regularized multi-task feature selection for multi-modality classification in Alzheimer's disease. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part I. LNCS*, vol. 8149, pp. 275–283. Springer, Heidelberg (2013)
8. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867 (2011)
9. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289 (2002)
10. Zhou, D., Huang, J., Schölkopf, B.: Learning with Hypergraphs: Clustering, Classification, and Embedding. *The Neural Information Processing Systems*, pp. 1601–1608. MIT Press (2006)



11. Gallagher, S.R., Goldberg, D.S.: Clustering Coefficients in Protein Interaction Hypernetworks. In: Gao, J. (ed.) ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics, pp. 552–560 (2013)
12. Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52, 1059–1069 (2010)
13. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q.: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* 32, 2322, e2319–2327 (2011)
14. Van Hoesen, G.W., Augustinack, J.C., Dierking, J., Redman, S.J., Thangavel, R.: The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates. *Ann. N Y Acad. Sci.* 911, 254–274 (2000)
15. Baldacara, L., Borgio, J.G., Moraes, W.A., Lacerda, A.L., Montano, M.B., Tufik, S., Bressan, R.A., Ramos, L.R., Jackowski, A.P.: Cerebellar volume in patients with dementia. *Rev. Bras. Psiquiatr.* 33, 122–129 (2011)