Detecting Alzheimer's disease by morphological MRI using hippocampal grading and cortical thickness

Simon F. Eskildsen[1], Pierrick Coupé[3], Vladimir Fonov[2], D. Louis Collins[2] and the Alzheimer's Disease Neuroimaging Initiative

[1]Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark. [2]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada. [3]Laboratoire Bordelais de Recherche en Informatique, Unité Mixte de Recherche CNRS (UMR 5800), PICTURA Group, Bordeaux, France.

Abstract. Structural MRI is an important imaging biomarker in Alzheimer's disease as the cerebral atrophy has been shown to closely correlate with cognitive symptoms. Recognizing this, numerous methods have been developed for quantifying the disease related atrophy from MRI over the past decades. Special effort has been dedicated to separate AD related modifications from normal aging for the purpose of early detection and prediction. Several groups have reported promising results using automatic methods; however, it is very difficult to compare these methods due to varying cohorts and different validation frameworks. To address this issue, the public challenge on Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia) was proposed. The challenge calls for accurate classification of 354 MRI scans collected among AD patients, subjects with mild cognitive impairment and cognitively normal control. The true diagnosis is hidden from the participating groups, thus making the validation truly objective. This paper describes our proposed method to automatically classify the challenge data along with a validation on 30 scans with known diagnosis also provided for the challenge.

1 Introduction

Neuronal injury is an important component of the pathophysiological process involved in Alzheimer's disease (AD). The cerebral atrophy resulting from the progressive neurodegeneration can be measured using magnetic resonance imaging (MRI) and is currently among the most important biomarkers of AD [1]. Optimizing such MRI-based biomarkers for detection and prediction of AD may have a significant impact on early diagnosis of patients as well as being valuable tools when designing therapeutic studies of individuals at risk of AD to prevent or alter the progression of the disease.

Hippocampal atrophy has long been recognized as an early feature of the degenerative process involved in AD [2]. Reductions in hippocampal volume appear to be correlated to early memory decline [3]. While sensitive to first stage of AD, hippocampal degeneration is involved in other dementias, such as vascular dementia [4], and is known to be part of non-pathological brain aging [5]. Thus, volumetric meas-

urements of the hippocampus (HC) might be limited in their ability to predict the progression of AD [6-9]. Evidence suggests that the nature of degeneration in the HC and surrounding structures, such as the entorhinal cortex (ERC) and parahippocampal gyrus, is different in AD compared to other dementias and different from the changes occurring during normal aging [10]. We recently obtained results that support this finding since we showed that AD detection could be improved by considering the structural composition of the HC and its surrounding structures in the medial temporal lobe [8]. These results were obtained using a novel concept of measuring structural similarities, comparing the anatomy of a test subject to a library of AD patients and cognitive normal (CN) subjects.

Studies have shown that, apart from hippocampal and medial temporal lobe (MTL) atrophy, AD has a characteristic neocortical atrophy pattern [11, 12]. Cortical thinning of temporal and parietal lobe regions, the posterior cingulate and the precuneus seem to be involved at early stages of the disease [13]. In the advanced stages of the disease, atrophy spreads to almost the entire cortex sparing only the sensory-motor and visual cortex [14]. Recently, we showed [15] that if cortical thickness is measured in a consistent manner, patterns of cortical thinning can predict conversion to AD among mild cognitive impaired (MCI) subjects with higher accuracy.

In the current challenge the aim is to classify morphological MRI scans into the classes: AD, MCI and CN. For this purpose we propose to combine measurements of structural pathological patterns, measured by analyzing morphological alterations in key structures of the MTL with degenerative patterns of the neocortex, measured by cortical thickness.

2 Methods

2.1 Image data

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu/) and from the public challenge on Computer-Aided Diagnosis of Dementia based on structural MRI data (CADDementia, http://caddementia.grand-challenge.org/). The ADNI database contains 1.5T and 3.0T T1w MRI scans for AD, MCI, and CN at several time points. For this study we used cohorts with 1.5T scans from the ADNI1 study and separate cohorts with 3.0T images from the ADNI2 study (see Table 1 for cohort statistics). The CADDementia data consist exclusively of 3.0T T1w MRI scans from three different sites in Europe. It should be noted that scanner equipment and protocols are not harmonized between the sites, which gives rise to additional variation in the data. In total 384 scans are provided of which 30 scans have known labels (Table 1). The remaining 354 scans are used for evaluating the proposed classification methods in a completely blinded fashion. The authors do not have access to the diagnostic labels behind the test scans.

Table 1. Demographics of the datasets used in the analyses.

Dataset	N (females)	Age±sd
ADNI1 AD	181 (88)	75.3±7.5
ADNI1 MCI	381 (139)	74.8±7.4
ADNI1 CN	222 (105)	75.9±5.0
ADNI2 AD	48 (16)	75.6±8.8
ADNI2 MCI	183 (69)	71.7±7.6
ADNI2 CN	73 (36)	75.6±6.2
CAD AD	9 (6)	66.1±5.2
CAD MCI	9 (4)	68.0±8.5
CAD CN	12 (3)	62.3±6.3
CAD test set	354 (141)	65.1±7.8

2.2 Image preprocessing

All images were processed using a fully automatic pipeline [16]. Images were denoised [17] using a Rician-adapted noise estimation [18], bias field corrected [19], and registered to MNI space using a 12 parameter affine transformation [20]. To enable robust registrations we used as registration target a population-specific template derived from the ADNI1 database constructed using a series of linear and non-linear registrations as described in [21]. The custom template was created from 50 AD patients and 50 CN subjects randomly selected. This template better reflects the anatomy of ADNI data compared to the conventional ICBM template build from young healthy adults. Image intensities were normalized to match the intensity profile of the template [22], and finally the images were skull stripped using BEaST [23].

2.3 Hippocampus and entorhinal cortex

Structural features of the hippocampal complex were estimated using SNIPE (Scoring by Nonlocal Image Patch Estimator) method [8, 24]. In this technique, the local structural information surrounding each voxel (i.e., 3D patch) of a test subject is compared to those in a training library of MRI datasets from ADNI1 and ADNI2 AD and CN subjects with segmentation of the considered structures (HC and ERC). In short, a small patch of MRI data around each voxel (e.g., 7x7x7 voxel patches) from the test subject is compared to the training library with the goal to find similar patches. The patch similarity is used to compute a weight for the match and used to determine two type of information. First, the weights are used to perform segmentation of both considered structures using label fusion strategy [25]. Second, the weights are used to obtain grading values reflecting the proximity of the current patch to both training populations (e.g., CN and AD) (see [24] for more details). As proposed in [24], the 50 closest subjects were first selected from each training population (i.e., 50

AD[1] and 50 CN) using SSD over an initialization mask. Then, the grading maps and the segmentations of the considered structures were obtained simultaneously using SNIPE. The ADNI1 and ADNI2 data were graded in a leave-one-out fashion, while the CADDementia data were graded using the entire ADNI1 and ADNI2 library. Finally, the 8 SNIPE-based features extracted were the average grading value over the left and right HC and ERC and the volume of the same four structures. Volumes were calculated in normalized space (MNI space).

2.4 Neocortex

Cortical thickness was calculated using FACE (Fast Accurate Cortex Extraction) [26-28] and mapped to the cortical surface of the population-specific average non-linear anatomical template [21] using an iterative, feature-based algorithm [29]. MCI and CN subjects were used to generate a statistical map of group differences in cortical thickness. From this t-map, cortical thickness features were derived with the procedure described in [15] using the proportion of the cortical surface with the 15% largest t-values corresponding to a threshold of t=4.3 and t=1.0 for ADNI1 and ADNI2 respectively (see Figure 1). In brief, candidate ROIs were calculated using a multi-seed constrained surface based region growing algorithm initialized at local maxima of the 15% t-map. In [15] we found a cortical area of 10-15% to be suitable for searching for candidate ROIs. We used MCI/CN contrast to generate cortical features as opposed to the more conventional AD/CN contrast. This was to enable sensitivity to the three-way classification problem, acknowledging that the MCI classification is the most difficult task. A total of 90 cortical thickness ROIs were identified this way using ADNI1 data and 87 using ADNI2 (see Figure 2). Neocortical features comprised the mean cortical thickness within each of these ROIs measured in subject native space.

2.5 Classification

We used multinomial regression with lasso and L1/L2 elasticnet regularization for the classification. We applied the GLMNET matlab implementation for the purpose (http://www.stanford.edu/~hastie/glmnet_matlab/). During experiments we used ADNI1 or ADNI2 as training datasets. The proposed classification framework was based on ensemble learning approach [30]. In order to create the ensemble we used an iterative approach. For each iteration the following steps were performed:

1. **Age correction based on the CN ADNI training population**
   As in [31]**,** for each feature we estimated the age-related effect on the CN population using linear regression. The features for all the considered groups (ADNI and CAD groups) were then corrected using the estimated linear regression coefficients.

[1] For ADNI2 only 48 AD images were used in the segmentation and grading process.

2. Under-sampling and Over-sampling of the ADNI training dataset

In presence of imbalanced population sizes, the evaluation of the performance of classification algorithms might be biased [32]. Therefore, in this study, we used two different strategies to compensate for this bias. We first randomly under-sample majority classes (i.e., CN and MCI) to obtain similar number of subjects for the 3 classes. Then we used SMOTE (Synthetic Minority Over-sampling Technique) to increase the number of samples in each classes. SMOTE creates new synthetic samples by performing interpolation of the nearby neighbors in the feature space

3. Grid search for optimal classifier parameters

In order to estimate the optimal parameters of the classifier, we performed a grid search using the CAD training dataset. When the accuracy of the classification was balanced between the 3 classes of CAD training dataset and higher than a given threshold, the model was applied on the CAD test dataset and the prediction (i.e., the 3 *a posteriori* probabilities for all the subjects) was stored in the model ensemble. Step 2 and 3 are iteratively performed until 25 good models were found.

4. Ensemble learning classification

Finally, at the end of the iterative process, all the selected models in the ensemble were fused using a non-weighted strategy. We estimated the mean of the 3 *a posteriori* probabilities for all the subjects over all the selected models. The maximum fused *a posteriori* probabilities were finally used to estimate the final label of the CAD test dataset.

This procedure is used in the four following scenarios:

1. Using ADNI1 as training dataset and SNIPE and FACE features (98 features in total)
2. Using ADNI1 as training dataset and SNIPE features (8 features)
3. Using ADNI2 as training dataset and SNIPE and FACE features (95 features in total)
4. Using ADNI2 as training dataset and SNIPE features (8 features)

Finally, the four scenarios were combined by using the grand mean of all a posterior probabilities from the scenarios. As in step four above, the maximum posterior probability was used to label the images.

Using the CADDementia training data (n=30), the correct classification rates of each of the three classes (AD/MCI/CN) were calculated for all five scenarios.

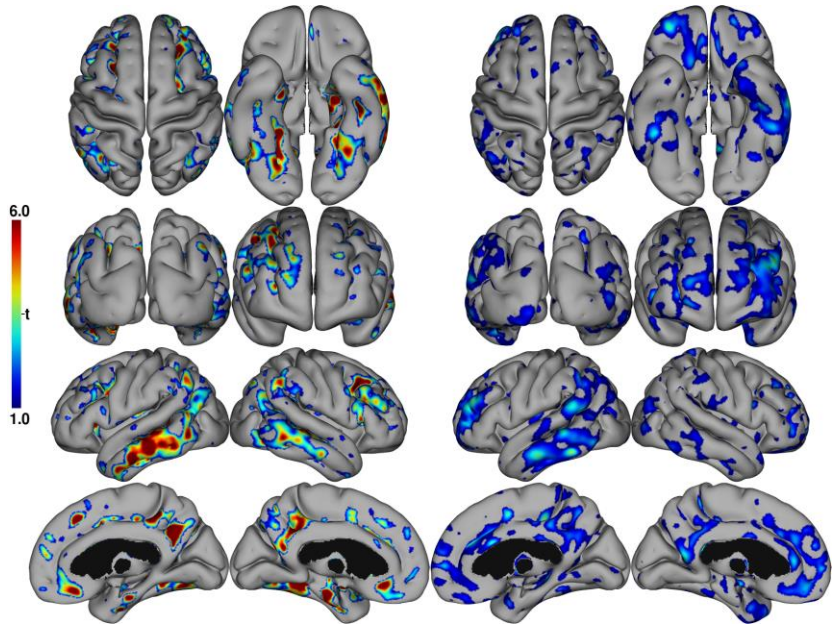

Fig. 1. Thresholded t-maps for ADNI1 MCI/CN contrast (left) and ADNI2 MCI/CN contrast (right). Notice the difference in statistical strength due to sample size differences.

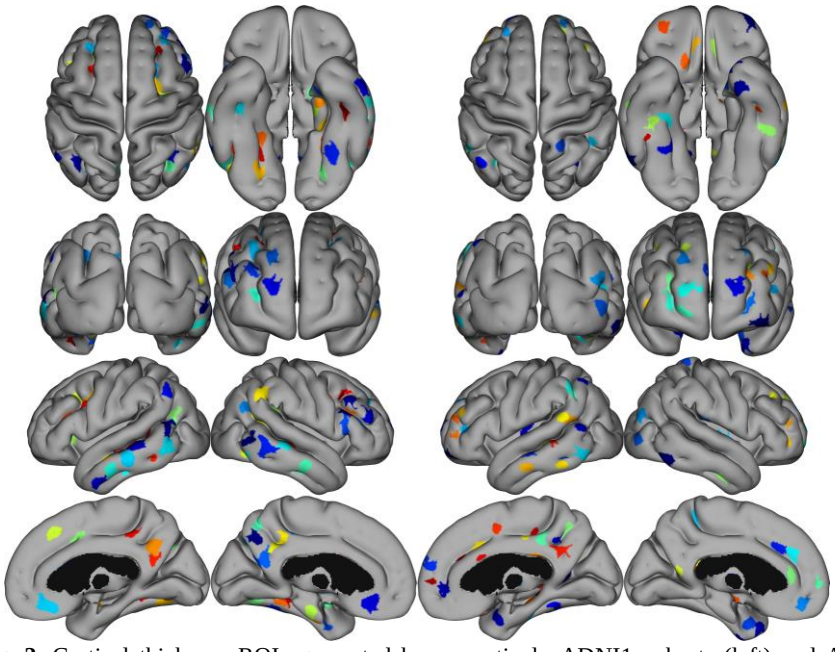

Fig. 2. Cortical thickness ROIs generated by respectively ADNI1 cohorts (left) and ADNI2 cohorts (right).

3 Results

3.1 Computational time

The classification process is fully automatic. Using a single core (Intel Core i7 @3.40Ghz) per subject the total computational time was approximately 55 minutes distributed on preprocessing (30 min), FACE (15 min), and SNIPE (10 min). Applying the classifier after it has been trained takes only a few seconds.

3.2 CADDementia training set

The final accuracy of the ensemble classifiers obtained on the CADDementia training set are presented in Table 2.

Table 2. Classification accuracies of the five classifiers when applied on CADDementia training data. All numbers are in percentage (%).

	Classification accuracy			
Classifier	CN	MCI	AD	Overall
SNIPE/FACE ADNI1	75.0	66.7	66.7	70.0
SNIPE ADNI1	75.0	77.8	77.8	76.7
SNIPE/FACE ADNI2	58.3	66.7	77.8	66.7
SNIPE ADNI2	66.7	66.7	77.8	70.0
Combined	75.0	66.7	77.8	73.3

4 Discussion

The results on the CADDementia training data may be inflated due to the grid search for optimal parameters. Nevertheless the results indicate the ranking of the different classifiers. It seems that training on ADNI1 data provides better results than training on ADNI2 data even though ADNI2 data should better represent the CADDementia data using only 3T images. The difference is most likely due to more available training data in the ADNI1 cohorts leading to well-defined classes. Thus the morphological variation of the three populations (AD/MCI/CN) is better represented in the larger sample. Adding cortical thickness features to the SNIPE features does not seem to improve results. In fact, it seems to impair the classifiers, possibly due to adding noise. Perhaps fewer and more carefully selected ROIs in the neocortex would have better complemented the MTL features and thus added discriminative information. Combining the four classifiers does not increase accuracy.

The AD group has consistently the highest (or tied with the highest) classification accuracy across all five classifiers. As expected the main difficulty in the three-way classification problem is to separate CNs and MCIs. A different strategy could have been to primarily focus on this problem by combining binary classifiers with the idea of “one class against all” applied in succession.

Looking at the demographics there is a significant age difference between ADNI and CADDementia data. We addressed this issue by adjusting the features for linear age effects using the ADNI CN subjects. However, this may not sufficiently compensate for the age differences between the cohorts. Due to aging effects, such as demyelination of the brain tissues, the MRI signal is not independent of subject age. Thus features based on image intensity and texture, such as the SNIPE features, are affected by these properties of the MR signal. Cortical thickness estimates are also affected, however, most likely to a lesser extent. Similarly, differences in field strength between training and testing data will affect the classifier performance.

The sample size of the CADDementia training data is too small to be representative of the large morphological variation typically found in demented as well as healthy brains. Thus training on a separate dataset, in our case ADNI data, is absolutely necessary to obtain a well-balanced and unbiased classifier. When the underlying labels of the CADDementia testing data is revealed, we will know whether our strategy using ADNI data was the right choice.

Acknowledgements

The research was supported by the MINDLab UNIK initiative at Aarhus University, funded by the Danish Ministry of Science, Technology and Innovation, grant agreement number 09-065250. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the Investments for the future Programme IdEx Bordeaux (ANR-10-IDEX-03-02), Cluster of excellence CPU and TRAIL (HR-DTI ANR-10-LABX-57).We acknowledge funding from Canadian Institutes of Health Research (MOP-111169) and les Fonds de Recherche Santé Québec.

References

1. Jack, C.R., Jr., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Lowe, V., Kantarci, K., Bernstein, M.A., Senjem, M.L., Gunter, J.L., Boeve, B.F., Trojanowski, J.Q., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Knopman, D.S.: Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. Archives of neurology 69, 856-867 (2012)
2. Ball, M.J., Fisman, M., Hachinski, V., Blume, W., Fox, A., Kral, V.A., Kirshen, A.J., Fox, H., Merskey, H.: A new definition of Alzheimer's disease: a hippocampal dementia. Lancet 1, 14-16 (1985)
3. De Leon, M., George, A., Stylopoulos, L., Smith, G., Miller, D.: EARLY MARKER FOR ALZHEIMER'S DISEASE: THE ATROPHIC HIPPOCAMPUS. The Lancet 334, 672-673 (1989)
4. Gainotti, G., Acciarri, A., Bizzarro, A., Marra, C., Masullo, C., Misciagna, S., Tartaglione, T., Valenza, A., Colosimo, C.: The role of brain infarcts and hippocampal atrophy in subcortical ischaemic vascular dementia. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology 25, 192-197 (2004)

5. Driscoll, I., Hamilton, D.A., Petropoulos, H., Yeo, R.A., Brooks, W.M., Baumgartner, R.N., Sutherland, R.J.: The aging hippocampus: cognitive, biochemical and structural findings. Cereb Cortex 13, 1344-1351 (2003)
6. Chupin, M., Gerardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehericy, S., Benali, H., Garnero, L., Colliot, O.: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19, 579-587 (2009)
7. Clerx, L., van Rossum, I.A., Burns, L., Knol, D.L., Scheltens, P., Verhey, F., Aalten, P., Lapuerta, P., van de Pol, L., van Schijndel, R., de Jong, R., Barkhof, F., Wolz, R., Rueckert, D., Bocchetta, M., Tsolaki, M., Nobili, F., Wahlund, L.-O., Minthon, L., Frölich, L., Hampel, H., Soininen, H., Visser, P.J.: Measurements of medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. Neurobiol Aging (2013)
8. Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Pruessner, J.C., Allard, M., Collins, D.L.: Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. NeuroImage: Clinical 1, 141-152 (2012)
9. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lotjonen, J.: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLoS One 6, e25446 (2011)
10. Devanand, D.P., Bansal, R., Liu, J., Hao, X., Pradhaban, G., Peterson, B.S.: MRI hippocampal and entorhinal cortex mapping in predicting conversion to Alzheimer's disease. NeuroImage 60, 1622-1629 (2012)
11. Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B.T., Morris, J.C., Fischl, B., Buckner, R.L.: The Cortical Signature of Alzheimer's Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and is Detectable in Asymptomatic Amyloid-Positive Individuals. Cereb Cortex 19, 497-510 (2009)
12. McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M.: Alzheimer Disease: Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment1. Radiology 251, 195-205 (2009)
13. Reiman, E.M., Jagust, W.J.: Brain imaging in the study of Alzheimer's disease. NeuroImage 61, 505-516 (2012)
14. Eskildsen, S.F., Fonov, V., Coupe, P., Collins, D.L.: Visualizing stages of cortical atrophy in progressive MCI from the ADNI cohort. In: Alzheimer's Association International Conference on Alzheimer's Disease. Alzheimer's & Dementia, (Year)
15. Eskildsen, S.F., Coupe, P., Garcia-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L.: Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. NeuroImage 65C, 511-521 (2013)
16. Aubert-Broche, B., Fonov, V.S., Garcia-Lorenzo, D., Mouiha, A., Guizard, N., Coupe, P., Eskildsen, S.F., Collins, D.L.: A new method for structural volume analysis of longitudinal brain MRI data and its application in studying the growth trajectories of anatomical brain structures in childhood. NeuroImage 82C, 393-402 (2013)

17. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Trans Med Imaging 27, 425-441 (2008)
18. Coupe, P., Manjon, J.V., Gedamu, E., Arnold, D., Robles, M., Collins, D.L.: Robust Rician noise estimation for MR images. Med Image Anal 14, 483-493 (2010)
19. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17, 87-97 (1998)
20. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18, 192-205 (1994)
21. Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L.: Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54, 313-327 (2011)
22. Nyul, L.G., Udupa, J.K.: Standardizing the MR image intensity scales: making MR intensities have tissue-specific meaning. pp. 496-504. SPIE, (Year)
23. Eskildsen, S.F., Coupe, P., Fonov, V., Manjon, J.V., Leung, K.K., Guizard, N., Wassef, S.N., Ostergaard, L.R., Collins, D.L.: BEaST: brain extraction based on nonlocal segmentation technique. NeuroImage 59, 2362-2373 (2012)
24. Coupe, P., Eskildsen, S.F., Manjon, J.V., Fonov, V.S., Collins, D.L.: Simultaneous segmentation and grading of anatomical structures for patient's classification: application to Alzheimer's disease. NeuroImage 59, 3736-3747 (2012)
25. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage 54, 940-954 (2011)
26. Eskildsen, S.F., Ostergaard, L.R.: Quantitative comparison of two cortical surface extraction methods using MRI phantoms. Med Image Comput Comput Assist Interv 10, 409-416 (2007)
27. Eskildsen, S.F., Ostergaard, L.R.: Active surface approach for extraction of the human cerebral cortex from MRI. Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention 9, 823-830 (2006)
28. Eskildsen, S.F., Uldahl, M., Ostergaard, L.R.: Extraction of the cerebral cortical boundaries from MRI for measurement of cortical thickness. pp. 1400-1410. (Year)
29. Eskildsen, S.F., Ostergaard, L.R.: Evaluation of Five Algorithms for Mapping Brain Cortical Surfaces. Proceedings of the 2008 XXI Brazilian Symposium on Computer Graphics and Image Processing, pp. 137-144. IEEE Computer Society (2008)
30. Dietterich, T.: Ensemble Methods in Machine Learning. Multiple Classifier Systems, vol. 1857, pp. 1-15. Springer Berlin Heidelberg (2000)
31. Koikkalainen, J., Polonen, H., Mattila, J., van Gils, M., Soininen, H., Lotjonen, J.: Improved classification of Alzheimer's disease data via removal of nuisance variability. PLoS One 7, e31112 (2012)
32. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321-357 (2002)