# SNIPE: A New Method to Identify Imaging Biomarker for Early Detection of Alzheimer's Disease

Pierrick Coupé, Simon F. Eskildsen, José V. Manjón, Vladimir Fonov, Jens C. Pruessner, Michèle Allard, Louis Collins

HAL Id: hal-00739288

https://hal.science/hal-00739288

Submitted on 7 Oct 2012

# SNIPE: A New Method to Identify Imaging Biomarker for Early Detection of Alzheimer's Disease

Pierrick Coupé [1,2], Simon F. Eskildsen [2,3], José V. Manjón[4], Vladimir Fonov[2], Jens C. Pruessner [6], Michèle Allard[5], D. Louis Collins[2] and the Alzheimer's Disease Neuroimaging Initiative*

[1] Laboratoire Bordelais de Recherche en Informatique,
Unité Mixte de Recherche CNRS (UMR 5800), Bordeaux, France
[2] McConnell Brain Imaging Centre, Montreal Neurological Institute,
McGill University, Montreal, Canada, Montreal, Canada H3A 2B4
[3] Center of Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark
[4] Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones
Avanzadas (ITACA), Universidad Politécnica de Valencia, Spain
[5] Institut de Neurosciences cognitives et intégratives d'Aquitaine,
CNRS (UMR 5287) - Bordeaux, France
[6] Departments of Psychiatry, Neurology and Neurosurgery,
McGill University, Montreal, Canada

**Abstract.** While the automatic detection of AD has been widely studied, the problem of the prediction of AD (or its early detection) has been less investigated. This might be explained by the fact that the prediction problem is clearly more challenging since the anatomical changes are more subtle. However, from a clinical point of view the prediction of AD is the key question since it is in that moment when treatment is possible. The potential use of structural MRI as imaging biomarker for Alzheimer's disease (AD) for early detection has become generally accepted, especially the use of atrophy of entorhinal cortex (EC) and hippocampus (HC). Therefore, in this study, we analyze the capabilities of the recently proposed method, SNIPE (Scoring by Nonlocal Image Patch Estimator), for the early detection of AD to analyze EC and HC atrophy over the entire ADNI database (834 subjects). During validation, the detection (AD vs. CN) and the prediction (pMCI vs. sMCI) efficiency of SNIPE were studied. The obtained results showed that SNIPE obtained competitive or better results than HC volume, cortical thickness and TBM. Moreover, results indicated that MRI grading-based biomarkers are more relevant than volume-based biomarkers. Finally, the success rate obtained by SNIPE was 90% for detection (AD vs. CN) and 74% for prediction (pMCI vs. sMCI).

## 1    Introduction

Clinical trials for Alzheimer's disease (AD) have been lately targeting disease-modifying therapies [1] stressing the need for identifying the disease in its prodromal stage when the pathological injury is not too severe. Finding biomarkers that could lead to this detection is therefore a major issue for current international research. Structures in the medial temporal lobe are more and more studied because of their

strong involvement in the pathogenesis of Alzheimer's disease (AD). The histopathology investigations of Braak and Braak [2] suggests that AD begins with the formation of neurofibrillary tangles in the medial temporal lobe, particularly the entorhinal cortex (EC), a structure of the parahippocampal cortex, and is then followed by the hippocampus (HC), and from there on expands to other structures across the neocortex. When the evolution of AD can be studied by post-mortem studies, the question arises if neuroimaging techniques might be able to detect these subtle changes *in-vivo* before the onset of more downstream symptoms.

In addition to the accepted use of CSF and PET biomarkers, the potential use of structural MRI as early imaging biomarker for AD detection has taken more importance in the literature [3]. Especially the use of atrophy of the EC and HC as early imaging biomarkers is considered a promising way to follow AD progression [3] since decrease in cognitive performances on episodic memory tests, the cognitive hallmark of AD [4-6], are associated with temporal lobe atrophy. However, the automatic extraction of these medial temporal lobe structures is challenging especially in case of EC [7]. Moreover, the inter-subject variability of brain anatomy tends to limit AD detection using only volumetric approaches [8, 9]. Recently, new nonlocal patch-based frameworks have been proposed on these two aspects: a robust approach to automatically segment HC and EC [10] and the characterization of structure atrophy using a scoring method [9].

The scoring of the considered structure is achieved by estimating the nonlocal similarity of the subject structures under study with different training populations. Thanks to a nonlocal framework, the Scoring by Nonlocal Image Patch Estimator (SNIPE) handles the inter-subject variability by enabling a one-to-many mapping between the subject's anatomy and the anatomies of many training templates. Moreover, enabled by the patch-based comparison principle, SNIPE detects subtle changes caused by the disease as already shown in [9]. In a previous study, the high success rate accuracy of SNIPE for AD detection (i.e., AD patients vs. cognitively normal (CN) subjects) has been demonstrated on a subset of the ADNI database (i.e., 100 subjects).

From a clinical point a view, AD prediction (i.e., progressive mild cognitive impairment (MCI) vs. stable MCI) is a more crucial question than diagnosis, but this question is clearly more clinically challenging since the anatomical changes that need to detect are more subtle. Recently, this problem has been studied using image analysis such as HC volume, Cortical Thickness measurement (CTH), Voxel Based Morphometry (VBM) and Tensor Based Morphometry (TBM) [8, 11-18]. Comparison of these imaging biomarkers can be found in [8, 13]. According to these comparisons, the accuracy of AD prediction of the compared methods (e.g., HC volume, CTH, VBM or TBM) is inferior to 66% [8]. To the best of our knowledge, the highest accuracy obtained on the entire ADNI database has been obtained by combining four methods resulting in 68% of accuracy for pMCI vs. sMCI [8].

In this study, we propose to investigate the capabilities of SNIPE for early detection of AD on the entire ADNI database (834 scans). Moreover, we propose to compare our results with the different methods compared by Wolz et al. in [8] as these represent some of the best results published to date using the whole ADNI database.

## 2    Methods

### 2.1    MRI scans

The data analyzed in this paper was obtained from the Alzheimer's Disease Neuro-imaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clini-cians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

To facilitate comparison with previous work, the 1.5T baseline scans used here are the same as those used in [8], where the 834 available baseline ADNI scans were divided into 4 populations for CN, progressive MCI (pMCI), stable MCI (sMCI) and AD. An MCI subject is considered as progressive if he or she converted to AD at the date of July 2011. The four groups contained 231 CN, 238 sMCI, 167 pMCI and 198 AD. Demographic details of the used dataset can be found in Table 1.

**Table 1:** Demographic details of the four groups analyzed: patients with AD (AD), cognitively normal subjects (CN), patients with progressive mild cognitive impairment (pMCI) and patients with stable mild cognitive impairment.

|      | Population size | % males | Age ± std | MMSE ± std |
|------|-----------------|---------|-----------|------------|
| CN   | 231             | 52%     | 76.0 ± 5.0 | 29.1 ± 0.9 |
| sMCI | 238             | 67%     | 74.9 ± 7.7 | 27.2 ± 2.5 |
| pMCI | 167             | 60%     | 74.5 ± 7.2 | 26.4 ± 2.0 |
| AD   | 198             | 50%     | 75.6 ± 7.7 | 22.8 ± 2.9 |

### 2.2    Image pre-processing

Before applying SNIPE, all the images were preprocessed through a fully automatic image processing pipeline. This pipeline was composed of the following steps: esti-mation of the standard deviation of noise with [19], denoising based on an optimized nonlocal means filter [20], correction of inhomogeneities using N3 [21], registration to the stereotaxic space based on a linear transform to the ICBM152 template (1x1x1 mm³ voxel size) [22], linear intensity normalization of each subject on template inten-sity, brain extraction using BEaST [23], image cropping around the structures of in-terest (see Fig. 1), cross-normalization of the MRI intensity between the subjects us-ing the method proposed in Nyul and Udupa within the estimated brain mask [24].

## 2.3    Scoring by Nonlocal Image Patch Estimator (SNIPE)

SNIPE is an image-based metric that uses a non-local, patch-based method to compare the local neighborhood (i.e., the 3D *patch*) around each voxel from a segmented structure of a test volume to similar patches from MRI data in a template library. This technique is not unlike patch-based segmentation procedures [6], but instead of using a voting scheme to determine the voxel label for segmentation, the patch-based similarity is used as weights in a function that yields a metric that describes the similarity of the test voxel to a particular patient group. In this case, the SNIPE metric score will indicate if the voxel in question is more like one group (e.g., NC) or another (e.g., AD). Details of the method are given in the following.
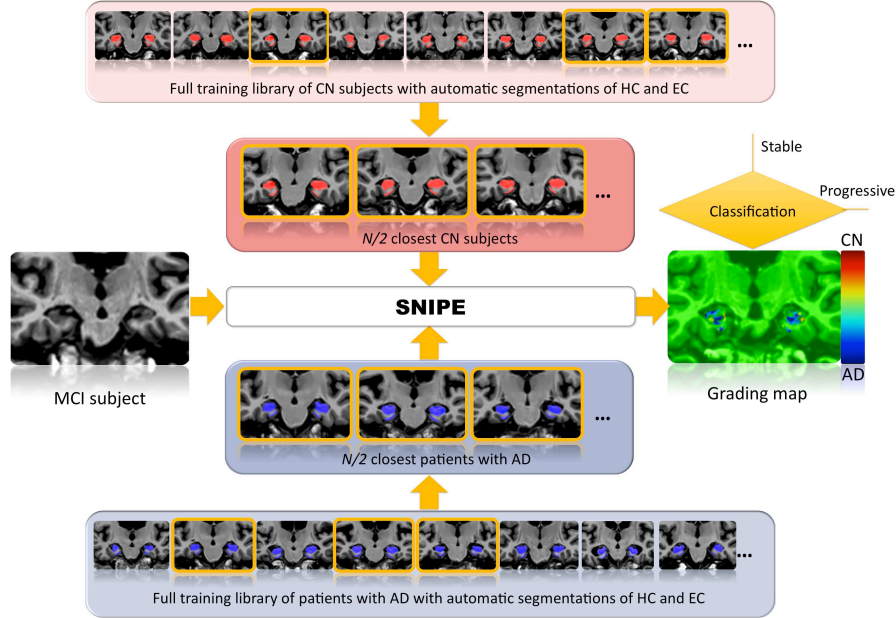
**Label propagation for library generation**



Figure 1: **Example of SNIPE workflow for an MCI subject**. After initial label propagation step, the resulting template training libraries are used by SNIPE to estimate the grading maps of the entire ADNI database (AD, pMCI, sMCI and CN).

For efficiency in template library generation, SNIPE propagates structure labels from a small number of manually segmented templates to all members of the training library. As in our previous work [5], the hippocampus (HC) and entorhinal cortex (EC) were manually labeled on 10 CN and 10 AD subjects using protocol described in [25]. These manual segmentations are then used in a patch-based segmentation procedure [6] to label the HC and EC in the 231 CN and the 198 AD datasets consti-

tuting the template library. Once completed, the template library is used to both segment and grade the HC and EC of new subjects.

### Structure segmentation and grading

With the template library built, the SNIPE algorithm can be used to segment the HC and EC and estimate the grading map for each subject within each ADNI database group (AD, pMCI, sMCI and CN) using the following procedure (see Fig. 1):

- Templates selection: The closest N/2 subjects from each training population (i.e., AD and CN) are selected based on an SSD metric evaluated over a standard, predefined initialization mask. For AD and CN subjects, we carefully removed the subject under study from the training library.
- Scoring and segmentation of the subject under study: For each voxel (included in the initialization mask) of the subject under study (progressive MCI in the example provided in Fig. 1), we compared its surrounding patch with all the patches from the N training templates selected from the AD and CN populations. This way, we simultaneously obtained a grading map and a segmentation for HC and EC.
- Feature extraction: The segmentations were used to compute the structure volumes, and the average grading value was estimated over the HC and EC segmentations. Both biomarkers were used as features during the classification step.

### Feature classification

The classification between different groups is based on a linear discriminant analysis (LDA). In previous work [9], we showed that slightly better classification accuracy could be obtained for AD vs. CN using quadratic discriminant analysis (QDA). However, in order to enable a more direct comparison with the work of Wolz et al. [8], we used LDA in this study. Moreover, in [9], it has been demonstrated that better classification accuracy could be obtained using subjects' age as feature in addition to volume or grade. Therefore, age was included as an additional feature for all experiments presented here for grade and volume biomarkers. The correlation between imaging biomarkers and the subjects' age will be analyzed. In addition, in [9], it has been shown that better classification was obtained for HC and HC-EC complex. Thus we will use these two structures for image-based biomarkers during methods comparisons. Finally, we used a repeated leave-N-out cross validation procedure (100 x LNOCV) in a similar manner to that presented in [8]. In each LNOCV experiment, 95% of the datasets were used as training set and the remaining 5% as testing set, randomly chosen. To reduce the variance of the results, this procedure was repeated 100 times and the mean classification rate was used as the final result.

### 2.4    Implementation details

All parameters proposed in [5] are used here, except the patch size for EC and the number of used training templates N.  In more recent work, it has been shown in [26] that a patch of 5x5x5 voxels can be enough for EC segmentation and thus has been

used for better computational efficiency instead of a larger patch size. Therefore, we used this patch size for the EC and patches of 7x7x7 voxels for HC as suggested in [9]. Moreover, in [9], it was suggested to select 60% of the entire library during template selection (i.e. 30 AD and 30 CN of the 50 subjects available). In this study, we use only around 25% of the entire library ($N_{AD}$=50 and $N_{CN}$=50) for computational reasons as well.
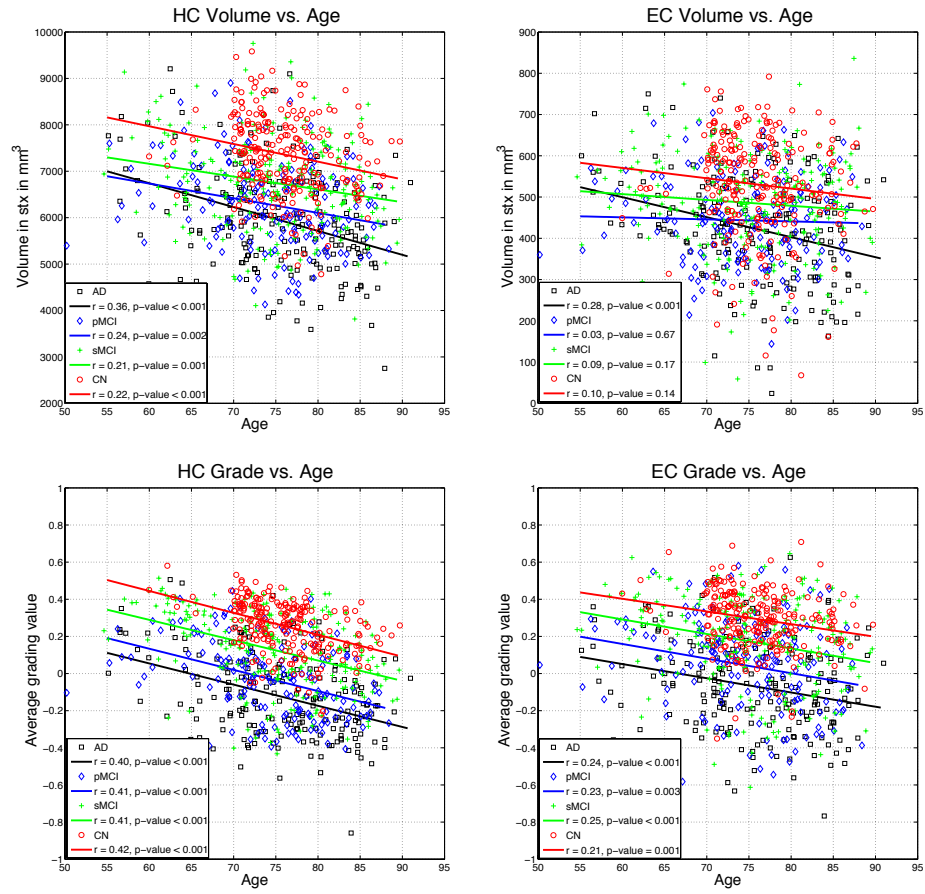
# 3    Results



Figure 2: Volumes and grades of HC and EC structures for the studied populations according to the age of the subjects. Linear regressions are displayed for better visualization of global tendencies. Pearson's coefficients and p-values of the regressions are provided in the legend.

### 3.1 SNIPE volumetric study

The top of Fig. 2 shows the volumes obtained by SNIPE for HC and EC. These volumes are plotted according to the age of the subjects for the four studied populations. First, we can observe a significant reduction of the volumes with age for HC while for EC this reduction is not statistically significant as assessed by p-values and Pearson's coefficients. Moreover, for HC a stronger reduction can be noted for the AD population. This can be explained by the addition of the atrophy related to the age with the atrophy related to the pathology. The evolution of EC volumes along the age is more difficult to interpret. The low Pearson's coefficient r and the high p-values of the linear regressions indicate a non-significant linear correlation of EC volumes with age except for AD. Compared to HC volumes, this might be due to higher inter-subject variability and more frequent errors in the segmentation as discussed in [9]. Therefore, for EC, the pathology-related patterns seem partially hidden by inter-subject variability.

### 3.2 SNIPE grading study

The bottom of Fig. 2 presents the average grading values obtained by SNIPE for HC and EC. For the structures studied, the grading values are significantly correlated with age (all the p-values are < 0.005) and decrease with age. Moreover, compared to the volumetric study, the correlation coefficients obtained with grade are larger. As expected, CN subjects have the highest grading values and AD patients the lowest. More interestingly, the same observation holds for sMCI compared to pMCI. Finally, as we will show later during the classification experiment by comparing volume and grade biomarkers, the higher correlation with age enables a better distinction of anatomical differences due to age-related modifications and pathology-related alterations, and the lower intra-population variance enables a better distinction of anatomical differences due to inter-subject variability and pathology-related alterations.

### 3.3 Patient's classification

Table 2 shows the success rates obtained by the proposed imaging biomarkers (i.e., HC volume, HC score, HC-EC volume, HC-EC score) for AD vs. CN, pMCI vs. CN, and pMCI vs. sMCI.

First, the SNIPE results show that grading-based biomarkers outperform volume-based biomarkers. Moreover, the HC-grade and HC-EC complex grade provided similar results. As expected, classification accuracies decrease when populations with more similar pathological status are used. Thus, the lowest accuracy was obtained for the pMCI vs. sMCI experiment.

***For AD vs. CN***, SNIPE obtained similar results to the combination of 4 methods as proposed in [8] (90% here compared to 89% in [4]). However, SNIPE obtained better results than HC volume, manifold-based learning, CTH and TBM; although TBM obtained results close to those of SNIPE. However, it should be noted that SNIPE

does not require nonlinear registration contrary to TBM and thus is less computationally expensive (i.e., around 5 minutes of processing).

*For pMCI vs. CN,* SNIPE obtained better results than all the methods compared in [8] as well as their combination (88% compared to 84%). Moreover, these results are close to the results obtained by SNIPE for AD vs. CN. This seems to indicate that the pMCI subjects can be as reliably classified as the AD population indicating that the SNIPE technique with the inclusion of HC and EC grade is better able to distinguish pMCI from CN than the multi-method used in [4].

*For pMCI vs. sMCI,* SNIPE obtained clearly better results than all the methods compared in [8] (74% compared to 68%). These worst results compared to pMCI vs CN, could be explained by the heterogeneity of sMCI group including a mix of individuals including some who will convert to AD as well as others who will not. In any case, these results highlight the potential of SNIPE for AD prediction by enabling the detection of anatomical changes caused by AD at the early stages of the pathology.

**Table 2: Imaging Biomarker Comparison**. Results obtained for AD vs. CN, pMCI vs. CN and pMCI vs. SMCI using 100 x LNOCV for several imaging biomarkers.

| | AD *vs.* CN | pMCI *vs.* CN | pMCI *vs.* sMCI |
|---|---|---|---|
| **SNIPE** | | | |
| • HC Volume | 82 | 78 | 66 |
| • HC Grade | **90** | 87 | **74** |
| • HC-EC Volume | 81 | 77 | 67 |
| • HC-EC Grade | 90 | **88** | 73 |
| **Multi-Method [8]** | | | |
| • HC Volume | 81 | 76 | 65 |
| • Manifold learning | 85 | 78 | 65 |
| • Cortical thickness | 81 | 77 | 56 |
| • TBM | 87 | 79 | 64 |
| • All | 89 | 84 | 68 |

Finally, it is noted that HC volume-based classification obtained with patch-based label fusion [10] yielded results similar to those based on multi-atlas label fusion [27].

## 4    Conclusion

In this study, we have shown that SNIPE-based biomarkers are as good as, or in some cases better than, the HC volume, manifold learning, cortical thickness, and TBM methods compared by Wolz et al. [8]. Moreover, we demonstrated a better classification rate using grading approaches than volumetric methods. Finally, the competitive results obtained on pMCI vs. sMCI highlight the potential use of SNIPE for early detection of AD. Although the obtained prediction rate (74%) is not yet suitable

for clinical use, the recent progresses of the MRI-based biomarkers [8, 9] on this classification problem are encouraging. Finally, SNIPE-based biomarkers might be combined with other efficient biomarkers to improve results as proposed in [8].

# References

1. Cummings, J.L., Benson, D.F.: Dementia of the Alzheimer type. An inventory of diagnostic clinical features. J Am Geriatr Soc 34, 12-19 (1986)
2. Braak, H., Braak, E.: Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 82, 239-259 (1991)
3. Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6, 67-77 (2010)
4. Schmidt-Wilcke, T., Poljansky, S., Hierlmeier, S., Hausner, J., Ibach, B.: Memory performance correlates with gray matter density in the ento-/perirhinal cortex and posterior hippocampus in patients with mild cognitive impairment and healthy controls--a voxel based morphometry study. Neuroimage 47, 1914-1920 (2009)
5. Remy, F., Mirrashed, F., Campbell, B., Richter, W.: Verbal episodic memory impairment in Alzheimer's disease: a combined structural and functional MRI study. Neuroimage 25, 253-266 (2005)
6. Di Paola, M., Macaluso, E., Carlesimo, G.A., Tomaiuolo, F., Worsley, K.J., Fadda, L., Caltagirone, C.: Episodic memory impairment in patients with Alzheimer's disease is correlated with entorhinal cortex atrophy. A voxel-based morphometry study. J Neurol 254, 774-781 (2007)
7. Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y.Y., Jagust, W.J., Yaffe, K., Kramer, J.H., Reed, B., Norman, D., Chui, H.C., Weiner, M.W.: Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. J Neurol Neurosurg Psychiatry 71, 441-447 (2001)
8. 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)
9. 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)
10. Coupe, P., Manjon, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. Neuroimage 54, 940-954 (2011)
11. Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y.: Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage 59, 2217-2230 (2012)
12. 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)

13.    Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage 56, 766-781 (2011)

14.    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)

15.    Koikkalainen, J., Lotjonen, J., Thurfjell, L., Rueckert, D., Waldemar, G., Soininen, H.: Multi-template tensor-based morphometry: application to analysis of Alzheimer's disease. Neuroimage 56, 1134-1144 (2011)

16.    Misra, C., Fan, Y., Davatzikos, C.: Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage 44, 1415-1422 (2009)

17.    Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Demonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P.: Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain : a journal of neurology 132, 2036-2047 (2009)

18.    Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M.W., Lovestone, S., Spenger, C., Wahlund, L.O.: AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. NeuroImage 58, 818-828 (2011)

19.    Coupe, P., Manjon, J.V., Gedamu, E., Arnold, D., Robles, M., Collins, D.L.: Robust Rician noise estimation for MR images. Medical image analysis 14, 483-493 (2010)

20.    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)

21.    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)

22.    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)

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.    Nyul, L.G., Udupa, J.K.: Standardizing the MR image intensity scales: making MR intensities have tissue specific meaning. P Soc Photo-Opt Ins 1, 496-504 (2000)

25.    Pruessner, J.C., Kohler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., Kabani, N., Collins, D.L., Evans, A.C.: Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. Cereb Cortex 12, 1342-1353 (2002)

26.    Hu, S., Coupé, P., Pruessner, J., Collins, D.L.: Validation of appearance-model based segmentation with patch-based refinement on medial temporal lobe structures. In: MICCAI Workshop on Multi-Atlas Labeling and Statistical Fusion, pp. 28-37.  (Year)

27.    Lotjonen, J., Wolz, R., Koikkalainen, J., Julkunen, V., Thurfjell, L., Lundqvist, R., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. Neuroimage 56, 185-196 (2011)