3844

Automatic Segmentation of the Hippocampus in T1-Weighted MRI with Multi-Atlas Label Fusion Using Open Source Software: Evaluation in 1.5 and 3.0T ADNI MRI

J. W. Suh[1], H. Wang[1], S. Das[1], B. Avants[1], and P. A. Yushkevich[1]
[1]PICSL, Radiology, University of Pennsylvania, Philadelphia, PA, United States

Introduction

Hippocampal volume is arguably the most widely accepted MRI-based Alzheimer's disease (AD) biomarker. We provide and evaluate a highly reliable, open-source, validated turnkey software solution for automatic measurement of hippocampal volume and atrophy in T1 MRI data. The results are comparable to the best in a large clinical dataset. Such a tool will enable rapid analysis of imaging data for disease progression and treatment effect evaluation in clinical applications.

Methods

For automatic hippocampus volume measurement, we use a multi-atlas based segmentation with label fusion. For the experiment of hippocampus segmentation, we use the 1.5T and 3T MRI (MPRAGE) data from the Alzheimer's Disease Neuroimaging Initiatative (ADNI)[6] database. In our study, we only use data from mild cognitive impairment (MCI) patient (n=82 for 3T, n=72 for 1.5T) and controls (n=57 for 3T, n=88 for 1.5T). For 1.5T data, we use 20 mixed MCI/control atlases as a single atlas subset while we use 10 randomly selected atlas subsets (n=20) for 3T data for the cross-validation purpose. The end-user open source software toolkit for automatic hippocampus segmentation is composed of four components: rigid registration, symmetric normalized (SyN) deformable registration, multi-atlas similarity-weighted voting, and learning based segmentation correction. For the rigid transform, FSL FLIRT tool [1] is employed with 6 degrees of freedom. Deformable registration is performed using the Symmetric Normalization (SyN) algorithm [2] implemented in the ANTS software package. From the rigid and deformable registration, images are normalized to an atlas, and the hippocampus segmentation in the atlas is mapped back to the images, producing the segmentation of the latter. Using multiple atlases for the image registration, great improvement can be achieved by assigning a spatially-varying weight to each atlas, based on the similarity of the target image and the atlas after normalization. The segmentation results are corrected by a statistical learning with an AdaBoost classifier [3] that discriminates between voxels correctly and incorrectly labeled. The AdaBoost method is trained to recognize the mistakes made by the segmentation method and correct them.

Results

We have evaluated the software in both 3.0 and 1.5 Tesla MRI from ADNI. In both cases, we get remarkably similar results for agreement between automatic segmentation results and manual segmentations (See below, LH: Left Hippocampus).

ADNI 1.5T MRI Evaluation

Group	LH Dice	RH Dice
Control	.886	.886
MCI	.881	.891
Both	.884	.888

Table 1: Avg. Dice overlap between auto and manual segmentation.

Group	LH ICC	RH ICC
Control	.943	.927
MCI	.930	.968
Both	.946	.955

Table 2: Interclass correlation coefficient (ICC) for auto & manual H. volumes.

Group	LH	RH
Auto	.606	.613
Manual	.710	.730

Table 3: Statistical power of manual vs. auto volume comparison (for n=20).

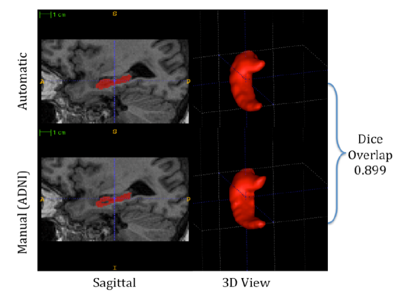

Figure 1: ADNI 1.5T MRI segmentation example

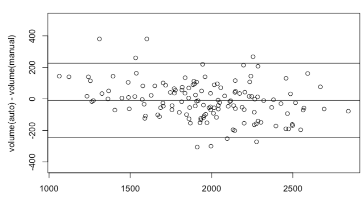

Figure 2: Agreement of automatic and manual LH volume measurements.

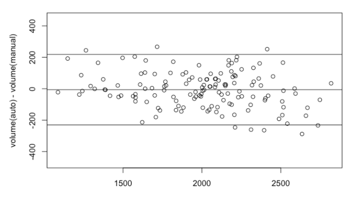

Figure 3: Agreement of automatic and manual RH volume measurements.

The evaluation shows very consistent high manual/automatic segmentation overlap in both datasets (Table 1 and 4). It also shows that ICC is high in both datasets (Table 2 and 5.) Typically, ICC=0.9 is considered sufficient for inter-observer reliability in manual hippocampus segmentation studies. Volume measurement results show consistent agreements between automatic and manual hippocampal volumes.

ADNI 3.0T MRI Evaluation

Group	LH Dice	RH Dice
Control	.910	.906
MCI	.897	.890
Both	.903	.898

Table 4: Avg. Dice overlap between auto and manual segmentation.

Group	LH ICC	RH ICC
Control	.925	.925
MCI	.928	.926
Both	.947	.944

Table 5: Interclass correlation coefficient (ICC) for auto & manual H. volumes.

Group	LH	RH
Auto	.950	.898
Manual	.961	.962

Table 6: Statistical power of manual vs. auto volume comparison (for n=20).

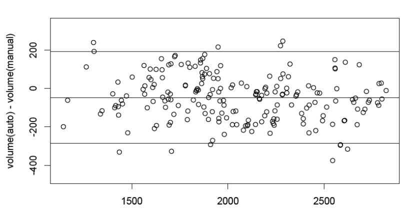

Figure 4: Agreement of automatic and manual LH volume measurements.

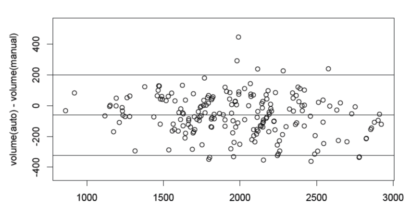

Figure 5: Agreement of automatic and manual RH volume measurements.

Discussion

As open-source software with clinical trials, our method accomplished very remarkable accuracy in hippocampus segmentation for agreement between automatic segmentation results and manual segmentations (Dice overlap ratio in the range 0.88 to 0.90). This is very competitive with the results reported in the literature. Especially, It uses only 20 atlases to produce comparable results of others produced by using 80 [4] or 110 atlases ([5] used 55 atlases with flipped mirror images, which are effectively 110 atlases).

Acknowledgement

This work is supported by NIH grants K25 AG027785, R01 AG037376 and grant 10295 from the Penn-Pfizer Alliance.

References:

[1] Smith et el., NeuroImage 23:208-219, 2004
[2] Avants et el., Med. Image Anal. 12:26-41, 2008
[3] Freund et el., Comput. Learning Theory, 904, 23-37,1995
[4] Collins et el., NeuroImage, 52:1355-1366, 2010
[5] Leung et el. NeuroImage, 51:1345-1359, 2010
[6] Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI)