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. Suh1, H. Wang1, S. Das1, B. Avants1, and P. A. Yushkevich1
1PICSL, 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 scanning 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 Initiative (ADNI) 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

<table>
<thead>
<tr>
<th>Group</th>
<th>LH Dice</th>
<th>RH Dice</th>
<th>Group</th>
<th>LH Dice</th>
<th>RH Dice</th>
<th>Group</th>
<th>LH Dice</th>
<th>RH Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>.866</td>
<td>.866</td>
<td>MCI</td>
<td>.881</td>
<td>.891</td>
<td>Both</td>
<td>.884</td>
<td>.888</td>
</tr>
</tbody>
</table>

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

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

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

ADNI 3.0T MRI Evaluation

<table>
<thead>
<tr>
<th>Group</th>
<th>LH Dice</th>
<th>RH Dice</th>
<th>Group</th>
<th>LH Dice</th>
<th>RH Dice</th>
<th>Group</th>
<th>LH Dice</th>
<th>RH Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>.910</td>
<td>.906</td>
<td>MCI</td>
<td>.897</td>
<td>.890</td>
<td>Both</td>
<td>.903</td>
<td>.898</td>
</tr>
</tbody>
</table>

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

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

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

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

Figure 1: ADNI 1.5T MRI segmentation example

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.

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: