Statistical Validation of Automatic Methods for Hippocampus Segmentation in MR Images of Epileptic Patients

Mohammad-Parsa Hosseini, *Student Member, IEEE,* Mohammad R. Nazem-Zadeh,
Dario Pompili, *Senior Member, IEEE,* Hamid Soltanian-Zadeh, *Senior Member, IEEE*

***Abstract*—Hippocampus segmentation is a key step in the evaluation of mesial Temporal Lobe Epilepsy (mTLE) by MR images. Several automated segmentation methods have been introduced for medical image segmentation. Because of multiple edges, missing boundaries, and shape changing along its longitudinal axis, manual outlining still remains the bench-mark for hippocampus segmentation, which however, is impractical for large datasets due to time constraints. In this study, four automatic methods, namely FreeSurfer, Hammer, Automatic Brain Structure Segmentation (ABSS), and LocalInfo segmentation, are evaluated to find the most accurate and applicable method that resembles the bench-mark of hippocampus. Results from these four methods are compared against those obtained using manual segmentation for T1-weighted images of 157 symptomatic mTLE patients. For performance evaluation of automatic segmentation, Dice coefficient, Hausdorff distance, Precision, and Root Mean Square (RMS) distance are extracted and compared. Among these four automated methods, ABSS generates the most accurate results and the reproducibility is more similar to expert manual outlining by statistical validation. By considering p-value<0.05, the results of performance measurement for ABSS reveal that, Dice is 4%, 13%, and 17% higher, Hausdorff is 23%, 87%, and 70% lower, precision is 5%, -5%, and 12% higher, and RMS is 19%, 62%, and 65% lower compared to LocalInfo, FreeSurfer, and Hammer, respectively.**

Keywords: Segmentation, Hippocampus, Magnetic Resonance Imaging (MRI), Comparison, Epilepsy

I. INTRODUCTION

Hippocampus is one of the most significant structures for epilepsy diagnosis and treatment. Mesial Temporal Lobe Epilepsy (mTLE) is a group of disorders in which patients suffer from recurrent epileptic seizures arising in one or both temporal lobes of the brain. The last resort for long-term seizure freedom for drug-resistant mTLE patients is surgical resection of the epileptogenic hippocampus.

Pathoanatomical and delineate functional changes of hippocampus in mTLE can be evaluated in segmented hippocampus from Magnetic Resonance Imaging (MRI), which is expected to result in a successful surgical outcome. Hippocampus is characterized by multiple edges and missing boundaries. Moreover, the size and shape of hippocampus change along its longitudinal axis. These characteristics make the automatic segmentation an extremely challenging task. In addition, both inter-rater and intra-rater variabilities are prone to manual segmentation, that are absent in automatic methods. Yet, manual segmentation of the hippocampus is the current gold standard assuming proper reproducibility, although it needs trained experts in neuroanatomy and is a time-consuming task requiring multiple hours per subject.

Automatic hippocampus segmentation methods can be categorized to atlas-base methods, energy-minimizing models, information-base algorithms, pattern-recognition models, and various combinations of them. A segmentation based on atlas registration and minimization of an energy function with intensity and prior terms is presented in [1]. Aljabar et al. [2] present a method for multi-atlas segmentation and selection. The effectiveness of their atlas selection is shown by Dice coefficient, and some of the most applicable automatic methods for hippocampus segmentation are compared using statistical validation methods.

FreeSurfer [3] is a software package for automatic analysis of brain structures and is a subcortical atlas-based segmentation method. Volumetric segmentation, inter-subject alignment, segmentation of hippocampal subfields, white matter fascicles segmentation, construction of surface models of cerebral cortex, and some other brain analysis are included in this tool. Nonlinear template matching is used in this tool for the segmentation of brain structures like hippocampus. This software is freely available and open source.

LocalInfo is an automatic segmentation and lateralization algorithm for hippocampus [4]. In this method, right and left hippocampi are segmented using a local information-based multiple atlas method (LocalInfo). Skull stripping, 3-label fuzzy classification and 10-label fuzzy classification, tissue-

This research is supported in part by NIH under grant R01-EB013227.

M. P. Hosseini is with the Department of Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ, USA and the Medical Image Analysis Lab., Radiology Department, Henry Ford Health System, Detroit, MI, USA (phone: +13133200462; fax: +13138744494; e-mail: parsa@cac.rutgers.edu; parsah@rad.hfh.edu).

M. R. Nazem-Zadeh is with the Medical Image Analysis Lab, Radiology Department, Henry Ford Health System, Detroit, MI, USA (e-mail: mohamadn@rad.hfh.edu).

D. Pompili is with the Department of Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ, USA (e-mail: pompili@ece.rutgers.edu).

H. Soltanian-Zadeh is with the Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran and with the Medical Image Analysis Lab., Radiology Department, Henry Ford Health System, Detroit, MI, USA (e-mail: hszadeh@ut.ac.ir, hamids@rad.hfh.edu).

type information extraction and optimization of the shape parameters are the steps of segmentation used in this method. The steps for LocalInfo extraction are Non-rigid registration of MR images with atlases, transformation to the lobe label maps, finding the most similar atlas label maps, affine registration, Principal Component Analysis (PCA) for extraction of principal shapes and mean shapes, respectively.

Different energy-minimizing models guided by internal-shape forces and external-image forces such as discrete contour models, classic snakes, and deformable contour models are used for the automatic segmentation of brain structures [5]. A modified deformable model [6-8] can be used in medical image segmentation. Hammer [9] is an elastic registration method for medical magnetic resonance images of the brain. This method minimizes the energy function for deformable registration and segments the brain structures in an atlas-based approach. A hierarchical procedure for optimization of energy function is used in Hammer and a set of features is applied to derive volumetric features. Moreover, the concept of an attribute vector is used to characterize the brain structures in the vicinity of each voxel. Finally, Geometric Moment Invariants (GMIs) are used for representing the geometric structure of the underlying anatomy. This method includes a morphometric analysis for segmentation of high-resolution images.

Pattern-recognition techniques are used for segmentation tasks. The Automatic Brain Structure Segmentation (ABSS) method is an algorithm based on Artificial Neural Networks (ANNs) [10]. Shape and signed-distance function of the desired structures are represented in different scales using GMIs and ANNs. For each scale, the GMIs as well as voxel intensities and coordinates are used as input parameters, whereas the signed-distance function is considered as output. Finally, ANN outputs of different stages are combined to classify the image voxel in two classes of inside and outside of the structure by another ANN.

In this study, Dice similarity coefficient, Hausdorff distance, Precision, and Root Mean Square (RMS) distance are used as metrics to evaluate the performance of the automatic segmentation methods in comparison with the manual segmentation results. The rest of the paper is organized as follows. In Section II, the material and methods including subjects and imaging protocol, manual and automatic segmentation, and performance measures are explained. In Section III the results of the statistical validation of the four evaluated methods are presented. Finally, the paper is concluded in Section IV.

II. MATERIAL AND METHODS

A. Subjects and Imaging Protocol

An archive review of mTLE patients treated between June 1993 and June 2014 at Henry Ford Hospital, Detroit, MI was used in this study. One hundred fifty-seven symptomatic patients affected by mTLE were selected for this study.

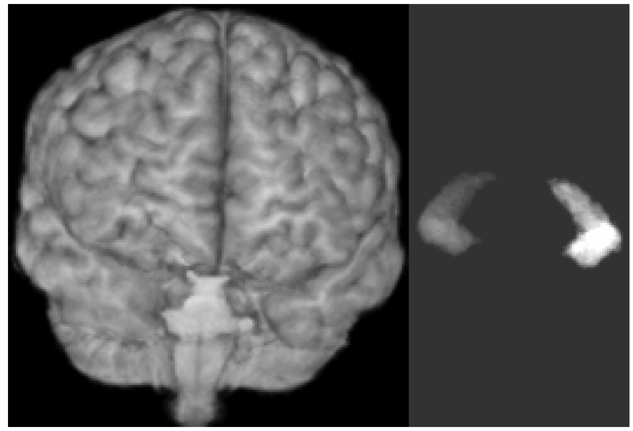

Fig 1. Surface rendering of the cortex surface (left) and segmented hippocampus (right) of T1-weighted MR images of a 52-year-old female who has been affected by mTLE for 19 years. The hippocampus segmentation was performed using the Automatic Brain Structure Segmentation (ABSS) method.

Preoperative MRI images obtained using a 1.5T or a 3.0T MRI system (Signa, GE, Milwaukee, USA) included coronal T1-weighted (using inversion recovery spoiled gradient echo, IRSPGR protocol) and coronal T2-weighted (using fluid attenuated inversion recovery, FLAIR protocol) images. On 1.5T MRI, T1-weighted imaging parameters were: TR/TI/TE=7.6/1.7/500 ms, flip angle=20°, voxel size=$0.781\times0.781\times2.0$ mm^3; whereas the FLAIR imaging parameters were: TR/TI/TE=10002/2200/119 ms, flip angle=90°, voxel size= $0.781\times0.781\times3.0$ mm^3. On 3.0T MRI, T1-weighted imaging parameters were: TR/TI/TE=10.4/4.5/300 ms, flip angle=15◦, voxel size=$0.39\times0.39\times2.00$ mm^3; whereas the FLAIR imaging parameters were: TR/TI/TE= 9002/2250/124 ms, flip angle=90◦, voxel size=$0.39\times0.39\times3.00$ mm^3.

B. Manual and Automatic Segmentation

Manual and Automatic segmentation methods were used to extract hippocampus, where the latter included: FreeSurfer, Hammer, LocalInfo, and ABSS. These fully automatic segmentation methods were applied to all the 157 subjects in order to extract hippocampus volume in the 3D space. For this purpose, first the DICOM images were converted by MRIcro [11] to NIfTI format; then, the automatic segmentation methods were applied to the images. Figure 1 shows the surface rendered cortex and hippocampus of T1-weighted MR images of a 52-year-old female who has been affected by mTLE for 19 years before undergoing surgery and its segmented hippocampus using the ABSS method. For manual segmentation, the Regions Of Interest (ROIs) encompassing the hippocampi were outlined in coronal plane; then, fine-tuning steps were done in sagittal view. The manual outlining is done in sequential coronal T1-weighted MR images; for identifying the hippocampus position, an MRI atlas is used as a reference [12]. For each subject, both the right and left hippocampi were segmented by an expert in the medical image analysis laboratory of the Henry Ford hospital using MRIcro. These were verified by two other investigators. Manual segmentation of hippocampi took approximately 5 hours per subject.

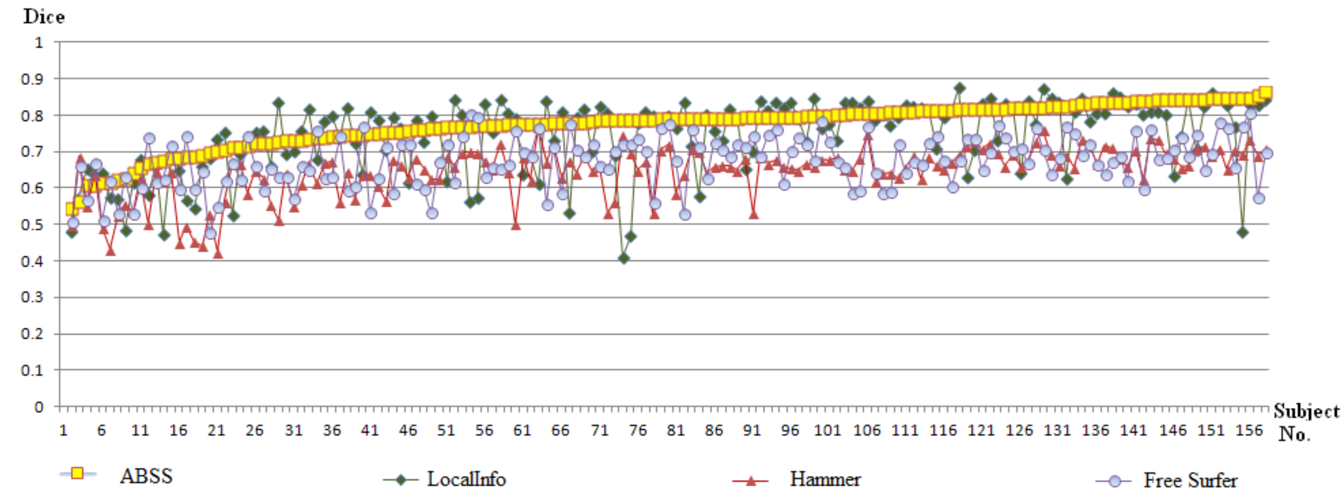

Fig 2: Dice coefficient of the four methods of automatic segmentation (FreeSurfer, Hammer, LocalInfo, ABSS) by case number for all the 157 mTLE patients. The cases have been sorted according to the value of the Dice coefficient obtained for the most accurate method (ABSS).

C. *Performance Measures*

We used the Hausdorff measures as the distance between two compact non-empty subsets of a metric space [13] in order to find the similarity of automatic and manual segmentation results. Hausdorff measure between two closed and bounded subsets A and B of a given metric space M is defined as,

$$H(A,B) = \max\{h(A,B), h(B,A)\}, \quad (1)$$
$$h(A,B) = \max\{d(a,B)\}, \quad (2)$$
$$d(a,B) = \min\{\mu(a,B)\}, \quad (3)$$

where *h(A,B)* is the direct distance between A and B, *d(α,B)* is the distance from a point to the set B, and *μ(α ,B)* is a point distance in the metric space M. The smaller $H(A,B)$, the more similar the distance between A and B.

The similarity between automatic and manual segmentation results can be assessed using overlap measures. One of the most popular methods, which we used for comparing each of the four automatic segmentation method against the gold standard, is based on the Dice Coefficient, defined as,

$$Dice\ Coefficient = \frac{2\,|A \cap B|}{|A|+|B|} \quad (4)$$

where A and B represent the regions being compared. Dice coefficient ranges from 0 to 1, where 1 means complete overlap. The volumes are measured by voxel counts. In addition, the following *Similarity* measure is related to Dice coefficient,

$$Similarity = \frac{True\ Positive}{True\ Positive + False\ Positive + False\ Negative} \quad (5)$$

$$Dice\ Coefficient = \frac{2\ Similarity}{1 + Similarity} \quad (6)$$

Positive predictive value or precision is defined as the number of true positives pixels for segmentation divided by both numbers of true positives and false positives for pixel segmentation.

$$Precision = \frac{True\ Positive}{True\ Positive\ + False\ Positive} \quad (7)$$

Root Mean Square (RMS) distance is used as a statistical measure to show the magnitude of a varying quantity of objects and is defined as,

$$Root\ Mean\ Square = \sqrt{0.5\,(x^2 + y^2)} \quad (8)$$

where x and y are the horizontal and vertical distances between the result of automatic segmentation and the bench-mark. RMS is used to show if the quantity of segmentation is varying, so the smaller its value, the higher the similarity.

III. Results

TABLE I summarizes the results in terms of mean and standard error of Dice coefficient, Hausdorff distance, Precision, and RMS for the four automatic segmentation methods considered in this study applied to the T1-weighted images of the mTLE patients. The Dice coefficient of the four evaluated methods for automatic segmentation by the case number for all subjects is shown in Fig. 2. The cases have been sorted according to the value of Dice obtained for the ABSS method, which provided the best accuracy among all the four methods.

TABLE I: Mean and standard error of different measures.

	Automatic Segmentation Method			
	ABSS	**LocalInfo**	**FreeSurfer**	**Hammer**
Dice	0.78± 0.01	0.74± 0.01	0.67± 0.01	0.65± 0.01
Hausdorff	3.09± 0.20	3.79± 0.23	5.77± 0.20	5.24± 0.23
Precision	0.81± 0.01	0.77± 0.01	0.85± 0.01	0.71± 0.01
RMS	1.26± 0.06	1.50± 0.08	2.04± 0.06	2.08± 0.07

The Dice coefficient for ABSS is 4% (p-value$<2\times10^{-3}$), 13% (p-value$<5\times10^{-33}$), and 17% (p-value$<2\times10^{-47}$) higher compared to LocalInfo, FreeSurfer, and Hammer, respectively, which shows that the segmentation performed using the ABSS method has more overlap with the gold standard than the others. The Hausdorff distance for ABSS is 23% (p-value$<3\times10^{-2}$), 87% (p-value$<7\times10^{-19}$), and 70% (p-value$<2\times10^{-11}$) lower compared to LocalInfo, FreeSurfer, and Hammer, respectively, which also suggests that the ABSS automatic segmentation method is more similar to the gold standard. The Precision for ABSS is 5% (p-value$<3\times10^{-10}$), -5% (p-value$<3\times10^{-8}$), and 12% (p-value$<2\times10^{-21}$) higher compared to LocalInfo, FreeSurfer, and Hammer, respectively. Note that the precision obtained using the FreeSurfer is the highest, which is an interesting result that is discussed in the conclusion section. The RMS distance for ABSS is 19% (p-value$<2\times10^{-2}$), 62% (p-value$<6\times10^{-18}$), and 65% (p-value$<5\times10^{-16}$) lower compared to LocalInfo, FreeSurfer, and Hammer, respectively, which shows that the ABSS has less varying quantity than the other competing methods.

IV. CONCLUSION

Several automatic segmentation techniques have been proposed in the literature; however, most of them have been tested only in nonepileptic subjects. To the best of our knowledge, there are only few reports of automatic hippocampal segmentation in the case of patients affected by Mesial Temporal Lobe Epilepsy (mTLE). In some studies very limited numbers of epileptic subjects are used. Our group used 46 epileptic patients in a previously publish study. In this study, we assessed and validated the most applicable automatic segmentation methods on 157 epileptic subjects. The results show that the Automatic Brain Structure Segmentation (ABSS) method is the most accurate automatic segmentation method for mTLE among the four evaluated methods. Specifically, LocalInfo, Hammer, and FreeSurfer are less accurate methods, respectively, according to the Dice coefficient, Hausdorff distance, and Root Mean Square Distance. Precision measure shows that FreeSurfer is more precise than LocalInfo and Hammer while using other measures this superiority is not confirmed. This result is logical and predictable because the FreeSurfer segments larger regions than the other methods and for this reason true positive and, ultimately, precision increase. Consequently, precision is not as meaningful as other measures in the evaluation of hippocampus segmentation results.

ACKNOWLEDGMENT

We would like to thank Alireza Akhondi-Asl and Mostafa Jabarouti-Moghadam for providing their software and helping with the image segmentation. Also, we thank Oltion Mesi, Abdelrahman Hassane, Harini Vijay, and Kourosh Jafari-Khozani for helping in manual image segmentation and skull stripping. Finally, we thank Kost V. Elliserich for referring epileptic cases to us for this study.

REFERENCES

[1] F. van der Lijn, T. den Heijer, M. Breteler, and W. J. Niessen, “Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts,” Neuroimage, vol. 43, no. 4, pp. 708-720, 2008.

[2] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, “Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy,” Neuroimage, vol. 46, no. 3, pp. 726-738, 2009.

[3] B. Fischl, “FreeSurfer,” Neuroimage, vol. 62, no. 2, pp. 774-781, 2012.

[4] A. Akhondi-Asl, K. Jafari-Khouzani, K. Elisevich, and H. Soltanian-Zadeh, “Hippocampal volumetry for lateralization of temporal lobe epilepsy: automated versus manual methods,” Neuroimage, vol. 54, pp. S218-S226, 2011.

[5] L. Shen, A. J. Saykin, S. Kim, H. A. Firpi, J. D. West, S. L. Risacher, B. C. McDonald, T. L. McHugh, H. A. Wishart, and L. A. Flashman, “Comparison of manual and automated determination of hippocampal volumes in MCI and early AD,” Brain imaging and behavior, vol. 4, no. 1, pp. 86-95, 2010.

[6] M.-P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor, “Three Cuts Method for Identification of COPD,” Acta Medica Iranica, vol. 51, no. 11, pp. 771-778, 2013.

[7] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor, “Detection and Severity Scoring of Chronic Obstructive Pulmonary Disease Using Volumetric Analysis of Lung CT Images,” Iranian Journal of Radiology, vol. 9, no. 1, pp. 22, 2012.

[8] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor, “A new scheme for evaluation of air-trapping in CT images,” IEEE, 2010.

[9] D. Shen, and C. Davatzikos, “HAMMER: hierarchical attribute matching mechanism for elastic registration,” Medical Imaging, IEEE Transactions on, vol. 21, no. 11, pp. 1421-1439, 2002.

[10] M. J. Moghaddam, and H. Soltanian-Zadeh, "Automatic segmentation of brain structures using geometric moment invariants and artificial neural networks." pp. 326-337.

[11] https://www.MRIcro.com.

[12] K. Jafari-Khouzani, K. Elisevich, S. Patel, B. Smith, and H. Soltanian-Zadeh, “FLAIR signal and texture analysis for lateralizing mesial temporal lobe epilepsy,” Neuroimage, vol. 49, no. 2, pp. 1559-1571, 2010.

[13] A. Andreev, and N. Kirov, “Hausdorff distances for searching in binary text images,” Serdica Journal of Computing, vol. 3, no. 1, pp. 23-46, 2009.