Iranian Journal of Medical Physics

ijmp.mums.ac.ir

# Evaluation and Comparison of Automatic Brain Segmentation Methods Based On the Gold Standard Method

Seyed Amir Zamanpour[1], Zohreh Ganji[1], Bahare Bigham[1], Fariba Zemorshidi[2], Hoda Zare[1, 3*]

1. Department of Medical Physics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
2. Department of Neurology, Ghaem Hospital, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
3. Medical Physics Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

ARTICLE INFO

*Article type:*
Original Paper

*Article history:*
Received: Jun 11, 2022
Accepted: July 17, 2022

*Keywords:*
MRI
Brain
Segmentation
FSL
Freesurfer
SPM

## ABSTRACT

***Introduction:*** Accurate segmentation of brain tissue in magnetic resonance imaging (MRI) is an important step in the analysis of brain images. There are automated methods used to segmentation the brain and minimize the disadvantages of manual segmentation, including time consuming and misinterpretations. These procedures usually involve a combination of skull removal, bias field correction, and segmentation. Therefore, segmented tissue quality assessment segmentation of gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) is required for the analysis of neuroimages.

***Material and Methods:*** This paper presents the performance evaluation of three automatic methods brain segmentation, fluid and white matter suppression [FSL, Freesurfer (FreeSurfer is an open source package for the analysis and visualization of structural, functional, and diffusion neuroimaging data from cross-sectional and longitudinal studies) and SPM12 (Statistical Parametric Mapping)]. Segmentation with SPM12 was performed on three tissue probability maps: i) threshold 0.5, ii) threshold 0.7 and iii) threshold 0.9. In order to compare and evaluate the automatic methods, the reference standard method, i.e., manual segmentation, was performed by three radiologists.

***Results:*** Comparison of GM, WM and CSF segmentation in MR images was performed using similarities between manual and automatic segmentation. The similarity between the segmented tissues was calculated using diagnostic criteria.

***Conclusion:*** Several studies have examined the classification of GM, WM, and CSF using software packages. In these studies, different results have been obtained depending on the type of method and images used and the type of segmented tissues. In this study, the evaluation of the segmentation of these packages with reference standard method is performed. The results can help users in selecting an appropriate segmentation tool for neuroimages analysis.

► *Please cite this article as:*
Zamanpour SA, Ganji Z, Bigham B, Zemorshidi F, Zare H. Evaluation and Comparison of Automatic Brain Segmentation Methods Based On the Gold Standard Method. Iran J Med Phys 2023; 20: 233-245. 10.22038/IJMP.2022.66025.2134.

## Introduction

Magnetic Resonance Imaging (MRI) technology has been widely used in research to measure detailed anatomical structures of the human brain. Because of high contrasts between soft tissues, MRI can make a useful distinction between different types of tissues, including white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) [1].

Obtaining quantitative features of each region from brain tissue such as tissue volume and shape requires segmentation. Segmentation of the brain in MR images into different classes, especially GM, WM, and CSF, is a crucial first step in clinical research. If the manual segmentation of brain tissues is performed by an expert radiologist or a specialized clinician, gives the best and most reliable results but certainly time-consuming due to many voxels in the brain MR image and prone to subjective errors. So the automatic segmentation tools has become very popular. However, the accuracy of the automatic segmentation is of great significance in brain image analysis, and the medical computing community continues to use manual segmentation to train the algorithms and to solve automatic segmentation problems, such as lack of reliable result [2-11]. Automated segmentation methods using MRI are usually based on basic image processing of pixel intensities and/or texture features (for example, relationships between pixels groups) [12]. These strategies basically incorporate intensity and edge-based strategies (counting Gaussian distribution models [13], Markov random field models [14], clustering approaches [15]), shape prior based strategies (incorporate atlas-based models [16], and deformable approaches [17]), and machine learning-based methods (counting SVM [18], KNN [19], random forest [20] and deep neural networks [21-24]). The performance of segmentation techniques depends in part on the validation criteria used. Generally, the validation is based on evaluation by experts. Because

---

*Corresponding Author:* Tel: +98-5138002321; Email: Zareh@mums.ac.ir

assessment by a single expert may be biased, some studies employ multiple experts using evaluation methods [25]. To increase the reliability and accuracy of automatic segment results, we need to evaluate automated segment softwares using manual segmentation as a reference (reference standard) [26].

Currently, there are several software packages used in neuroimaging analysis (in particular, image segmentation). One of the open-source software used for manual segmentation is ITK-SNAP, and the most widely used software tools for automated brain segmentation are the FSL packages [27-29], Freesurfer [30], and Statistical Parametric Mapping [31].

In this paper, we evaluated the automatic segmentation results of three common software packages to manual segmentation for the three tissues of, GM, WM, and CSF.

## Materials and Methods

### *Image Acquisition*

To evaluate the segmentation methods, MRI data of subjects were downloaded from the Alzheimer's disease Neuroimaging Initiative (ADNI) database [32]. The data set contains 19 T1-weighted MR brain images.

The MRI data of subjects were generated from ADNI2 project with the following parameters:

T1-weighted images were acquired on 3 Tesla scanner (GE Medical Systems), Acquisition Type=3D; Pulse Sequence=GR; TI=400 ms; TR=6.98 ms; TE=2.85 ms; flip angle=11°; Slice Thickness=1.2; Matrix X=256.0 pixels; Matrix Y=256.0 pixels; Matrix Z=196.0.

### *Segmentation Methods*

### *Manual Segmentation*

The manual segmentation was performed by drawing outlines on 2D cross-sections of a 3D image in 6 slices of the brain by "polygon" and "paintbrush" tools of ITK-SNAP v3.8.0 software (Penn Image Computing and Science Laboratory- PICSL -, and Scientific Computing and Imaging Institute - SCI, USA) [33]. Three expert radiologists implemented ITK-SNAP software to perform ground truth manual segmentation of the brain into GM, WM, and CSF structures and three labels were assigned using different colors in all subjects. For an accurate segmentation, a paintbrush size of 1 was selected.

### *Automated Segmentation*

### *FSL software*

One of the foremost broadly utilized program for fMRI, MRI, and DTI analyses is FMRIB's Program Library (FSL) [34], developed by members of the Oxford Centre for Functional MRI of the Brain (Oxford University). In this work, FSL version 6 is utilized for skull stripping and automatic segmentation of brain tissue. At the primary step, skull removal was carried out utilizing FSL's brain extraction tool (BET) to remove the extracerebral tissues. At the second, the FAST (FMRIB's Automated Segmentation Tool) utilized to segment the brain T1 images into three tissues of GM, WM, and CSF and performing a bias field and partial-volume effect correction.

### *Freesurfer software*

Freesurfer is a set of powerful and precise tools that arrange an automated analysis of essential features of the human brain. This software was developed by the Laboratory for Computational Neuroimaging (Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital, Boston USA) [35-38]. In this study Freesurfer v6.0.0 was implemented on T1 images for automatic brain segmentation, skull stripping and bias field correction. Table 1 describes all image processing steps in Freesurfer software.

Table 1. Image processing steps in freesurfer software

| Number | Task |
|---|---|
| 1 | Motion Correction and Conform |
| 2 | NU (Non-Uniform intensity normalization) |
| 3 | Talairach transform computation |
| 4 | Intensity Normalization 1 |
| 5 | Skull Strip |
| 6 | EM Register (linear volumetric registration) |
| 7 | CA Intensity Normalization |
| 8 | CA Non-linear Volumetric Registration |
| 9 | Remove Neck |
| 10 | LTA with Skull |
| 11 | CA Label (Volumetric Labeling, ie Aseg) and Statistics |
| 12 | Intensity Normalization 2 (start here for control points) |
| 13 | White matter segmentation |
| 14 | Edit WM With ASeg |
| 15 | Fill (start here for wm edits) |
| 16 | Tessellation (begins per-hemisphere operations) |
| 17 | Smooth1 |
| 18 | Inflate1 |
| 19 | QSphere |
| 20 | Automatic Topology Fixer |
| 21 | Final Surfs (start here for brain edits for pial surf) |
| 22 | Smooth2 |
| 23 | Inflate2 |
| 24 | Spherical Mapping |
| 25 | Spherical Registration |
| 26 | Spherical Registration, Contralateral hemisphere |
| 27 | Map average curvature to subject |
| 28 | Cortical Parcellation -Desikan_Killiany and Christophe (Labeling) |
| 29 | Cortical Parcellation Statistics |
| 30 | Cortical Ribbon Mask |
| 31 | Cortical Parcellation mapping to Aseg |

### *SPM software*

SPM is a free and open-source MATLAB-based (MathWorks Inc.) software developed by the Welcome Department of Cognitive Neurological Health members. The toolbox "LST: lesion segmentation tool" of SPM12 (University College London, London, UK) [39] is used for automatic segmentation. The procedure used by default in SPM is formed on an integrated segmentation model. This method involves tissue segmentation, registration and bias-field correction, all in one model. The main idea of this method is to model the intensity of the image based on tissue probability maps and from T1 weighted images [40]. In SPM, the segmentation process is performed using a modified Gaussian mixture model and the input MR image automatically segments into 3 distinct clusters WM, GM, and CSF [41]. In this study, all segmentations were performed by SPM12 using the default atlas and three tissue probability maps.

### *Statistical Analysis*

Comparison of manual and automatic segmentation were performed using various metrics as described below:

Dice and Jaccard coefficients [42, 43] were used as similarity indicators for quantitative analysis of the proposed method. These metrics indicate the overlap between two binary images and their values between 0 (without overlap) and 100 (full agreement).

Dice coefficient, could be a statistical tool which measures the similarity between two sets of data. This index has become arguably the foremost broadly utilized tool within the validation of image segmentation algorithms created with Artificial intelligence, but it may be a much more general concept which can be connected to sets of data for a variety of applications including NLP (Natural language processing) [44]. The equation for this concept is:

$$D = \frac{2|A \cap G|}{|A| + |G|} \tag{1}$$

- where A and B are two sets
- a set with vertical bars either side refers to the cardinality of the set, i.e. the number of elements in that set, e.g. |A| means the number of elements in set A
- ∩ is used to represent the intersection of two sets, and means the elements that are common to both sets

Jaccard Index is a statistical tool that helps assess the the overlap in two sets of data. This tool is similar to the Dice coefficient, but mathematically different and typically used for different applications. This tool is suitable for evaluating of object detection, where the detected area is understood in terms of pixels or voxels, and for various other tasks in AI. When this statistic is used for object detection in the context of artificial intelligence, it is sometimes referred to as the IOU (intersection over union) metric [45]. The equation for this tool is:

$$J = \frac{|A \cap G|}{|A \cup G|} \tag{2}$$

The ability to detect the actual tissue in the segmented mask is known as sensitivity. The higher sensitivity means lower missed correct voxels [46]. Its equation is:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

- where TP is true positive (actual tissue) and FN is false negative (false background)

The specificity means the ability to remove non-desired voxels. The lower missed non-desired voxels lead to the higher specificity [47]. The equation for this index is:

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

- where TN is true negative (actual background) and FP is false positive (false tissue)

Accuracy is how close or far off a certain series of measurements (observations or readings) is from its true value. Accuracy indicates the percentage of pixels in an image that are correctly classified [48]. Accuracy is also used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. In other words, the accuracy is the proportion of correct predictions (both true positives and true negatives) out of the total number of cases studied [48]. As such, it compares of pre- and post-test probability estimates. The binary accuracy quantization formula is:

$$Accuracy = \frac{TP+TN}{TN+FP+TP+FN} \tag{5}$$

- where TP = true positive; FP = false positive; TN = true negative; FN = false negative

To enable easy and fast access to the segmentation codes, GUI exe file was developed in MATLAB software in Figure 1 and the overall procedure is shown in Figure 2.

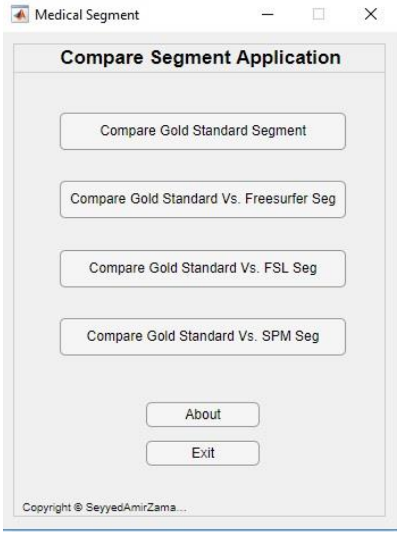

Figure 1. MATLAB GUI for MR images segmentation

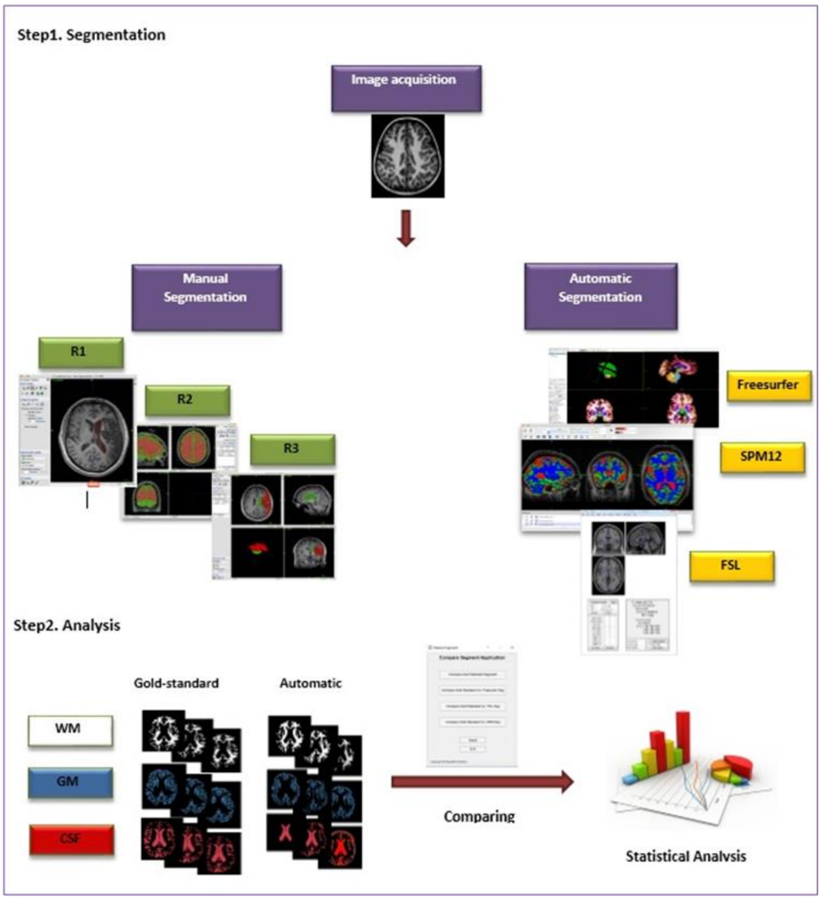

Figure 2. Overall procedure

## Results

To evaluate the accuracy of three brain tissue segmentation tools: FreeSurfer, FSL, and SPM12. The segmentation results of these three software were compared with the manual segmentation results.

The resulting WM, GM and CSF tissues from segmenting the MR image using the brain tissue segmentation tools: FSL, Freesurfer and SPM12, and the manual segmentation by three radiologists are shown in Figure 3.

### *Comparison based on segmented WM data*

Figure 4 shows the WM segmentation results by the three methods and the manual segmentation.

Table 2 shows the mean similarity based on Dice and Jaccard coefficients for segmented WM, GM and CSF from 20 MR images. The table also provides the average sensitivity, specificity, and accuracy.

### *Comparison based on segmented GM data*

Figure 5 shows the GM tissue extracted from the MR image using the segmentation tools and the manual segmentation.

Table 3 presents the results of the quantitative evaluation of GM segmentation.

### *Comparison based on segmented CSF data*

Figure 6 shows the CSF segmentation results from MR images. Table 4 shows the results of the quantitative analysis of CSF segmentation.

Figures 7 and 8 show the similarity of the segmented brain in terms of Dice and Jaccard metrics for different tissues in MR image. Figures 9 to 11 show the performance of the segmentation methods in terms of sensitivity, specificity and accuracy measurement for different tissues.

As shown in the figures, Freesurfer performed better for WM and GM segmentation, and SPM software performed better for CSF segmentation.

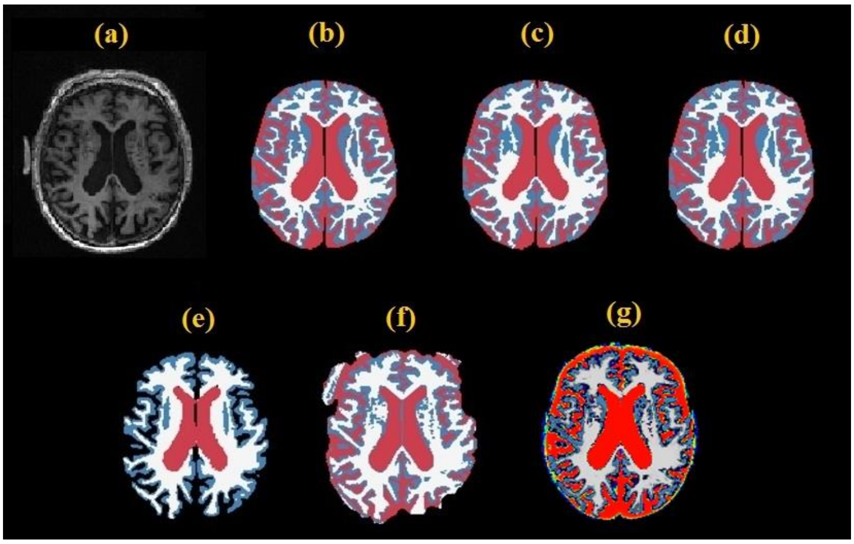

Figure 3. Example of tissue brain segmentation. The original image (a), the manual segmentation by three radiologists: R1 (b), R2 (c) and R3 (d), the automated segmentation using three methods: Freesurfer (e), FSL (f), and SPM12 (g)

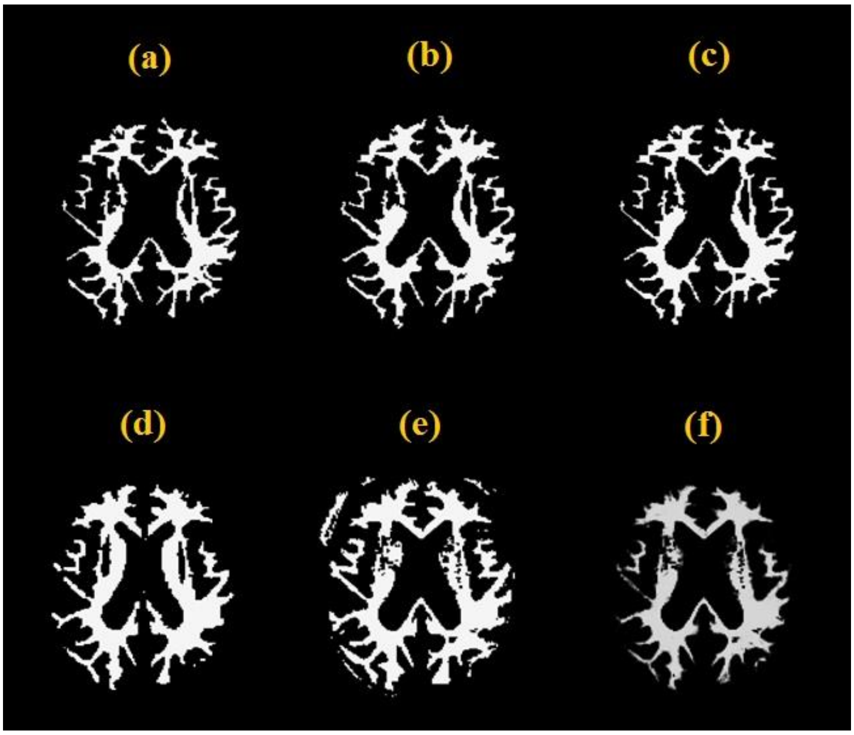

Figure 4. Example of WM segmentation. The manual segmentation by three radiologists: R1 (a), R2 (b) and R3 (c), the automated segmentation using three methods: Freesurfer (d), FSL (e), and SPM12 (f).

Table 2. Comparison of WM segmentation metrics of FSL, Freesurfer, and SPM and manual segmentation

| Segmentation method | | Radiologist | Statistic | Jaccard | Dice | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|---|---|---|---|
| FSL | | R1 | Mean | 69.6558 | 80.73607 | 81.00756 | 98.54299 | 97.54569 |
| | | | Std | 17.02813 | 13.99299 | 17.21833 | 1.438971 | 1.295754 |
| | | R2 | Mean | 71.54158 | 82.31591 | 82.94677 | 98.56809 | 97.71761 |
| | | | Std | 15.48887 | 12.357 | 15.69236 | 1.376885 | 1.24321 |
| | | R3 | Mean | 72.75194 | 83.11547 | 83.6062 | 98.66777 | 97.84555 |
| | | | Std | 15.75794 | 12.42814 | 15.56114 | 1.417886 | 1.249092 |
| Freesurfer | | R1 | Mean | 79.98281 | 88.15266 | 89.45147 | 98.97078 | 98.37872 |
| | | | Std | 13.6521 | 9.777102 | 10.86599 | 0.672408 | 1.064087 |
| | | R2 | Mean | 85.97701 | 92.15896 | 94.18231 | 99.20172 | 98.92012 |
| | | | Std | 9.183529 | 6.186748 | 6.557874 | 0.543151 | 0.618996 |
| | | R3 | Mean | 89.34611 | 94.03216 | 95.94818 | 99.40025 | 99.22714 |
| | | | Std | 10.00601 | 6.558683 | 6.249165 | 0.577739 | 0.654607 |
| SPM | Thresholding 0.5 | R1 | Mean | 67.42601 | 79.4098 | 72.57112 | 99.33036 | 97.55218 |
| | | | Std | 15.44266 | 12.53003 | 14.82178 | 0.519959 | 0.9688 |
| | | R2 | Mean | 69.55013 | 81.14119 | 74.62605 | 99.36471 | 97.73826 |
| | | | Std | 14.15654 | 10.94621 | 13.4697 | 0.512719 | 0.928344 |
| | | R3 | Mean | 70.50572 | 81.82824 | 75.12459 | 99.44114 | 97.82546 |
| | | | Std | 14.07384 | 10.75329 | 13.15261 | 0.530023 | 0.91037 |
| | Thresholding 0.7 | R1 | Mean | 64.7283 | 77.23776 | 68.8465 | 99.43087 | 97.38741 |
| | | | Std | 16.40975 | 13.79707 | 16.05116 | 0.46112 | 0.989032 |
| | | R2 | Mean | 66.75252 | 78.93578 | 70.76985 | 99.46469 | 97.57139 |
| | | | Std | 15.36906 | 12.41896 | 14.93586 | 0.46746 | 0.959137 |
| | | R3 | Mean | 67.43196 | 79.45065 | 71.09535 | 99.52491 | 97.63034 |
| | | | Std | 15.26822 | 12.23678 | 14.66376 | 0.479902 | 0.943304 |
| | Thresholding 0.9 | R1 | Mean | 58.93741 | 72.27872 | 61.68754 | 99.56702 | 97.01031 |
| | | | Std | 18.27863 | 16.67966 | 18.18156 | 0.381342 | 1.051066 |
| | | R2 | Mean | 60.51424 | 73.71467 | 63.23498 | 99.58645 | 97.16902 |
| | | | Std | 17.6101 | 15.63621 | 17.49101 | 0.397436 | 1.0354 |
| | | R3 | Mean | 60.92234 | 74.06126 | 63.38822 | 99.63185 | 97.20035 |
| | | | Std | 17.50591 | 15.48061 | 17.23984 | 0.407532 | 1.024814 |

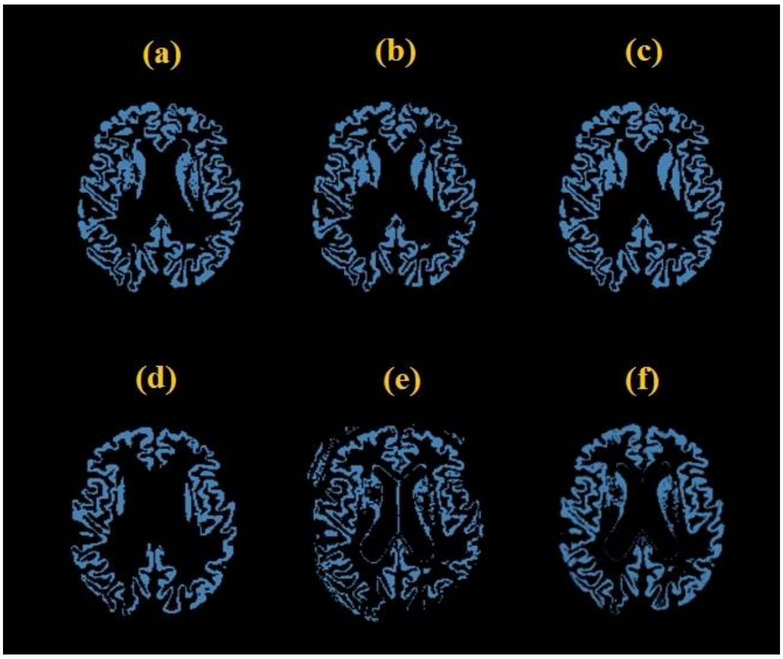

Figure 5. Example of GM segmentation. The manual segmentation by three radiologists: R1 (a), R2 (b), and R3 (c), the automated segmentation using three methods: Freesurfer (d), FSL (e), and SPM12 (f).

Table 3. Comparison of GM segmentation metrics of FSL, Freesurfer and SPM and manual segmentation

| Segmentation method | | Radiologist | Statistic | Jaccard | Dice | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|---|---|---|---|
| FSL | | R1 | Mean | 55.74637 | 70.91427 | 68.71167 | 97.91646 | 95.61813 |
| | | | Std | 10.67474 | 9.947647 | 13.78396 | 1.065085 | 1.592119 |
| | | R2 | Mean | 56.49182 | 71.6904 | 71.2621 | 97.77324 | 95.84129 |
| | | | Std | 9.264718 | 8.687463 | 12.24858 | 1.101369 | 1.497153 |
| | | R3 | Mean | 58.81038 | 73.53421 | 73.19417 | 97.91665 | 96.1199 |
| | | | Std | 9.645052 | 8.838576 | 12.64348 | 1.06556 | 1.490721 |
| Freesurfer | | R1 | Mean | 68.584 | 80.64821 | 80.0395 | 98.37884 | 96.88477 |
| | | | Std | 13.19078 | 9.264757 | 11.50245 | 1.00745 | 1.741527 |
| | | R2 | Mean | 75.98958 | 86.05034 | 88.07515 | 98.60223 | 97.81716 |
| | | | Std | 8.697224 | 6.283569 | 6.789785 | 0.877766 | 1.133978 |
| | | R3 | Mean | 82.55493 | 90.04418 | 92.24255 | 98.92592 | 98.42794 |
| | | | Std | 10.41171 | 7.101927 | 7.632416 | 0.853384 | 1.207239 |
| SPM | Thresholding 0.5 | R1 | Mean | 62.85285 | 77.00154 | 84.56481 | 96.85924 | 95.91368 |
| | | | Std | 6.26298 | 4.946716 | 8.697549 | 1.393693 | 1.392899 |
| | | R2 | Mean | 63.63599 | 77.68763 | 88.12492 | 96.69116 | 96.08219 |
| | | | Std | 4.414194 | 3.364795 | 5.565701 | 1.423083 | 1.382632 |
| | | R3 | Mean | 65.60914 | 79.14059 | 89.89778 | 96.81646 | 96.32661 |
| | | | Std | 4.578023 | 3.403054 | 5.922893 | 1.410104 | 1.352514 |
| | Thresholding 0.7 | R1 | Mean | 61.99949 | 76.28605 | 79.94985 | 97.36853 | 96.02247 |
| | | | Std | 7.239839 | 5.808703 | 10.08375 | 1.238466 | 1.348335 |
| | | R2 | Mean | 63.67106 | 77.67985 | 83.91649 | 97.26252 | 96.30863 |
| | | | Std | 5.158871 | 3.969583 | 7.217308 | 1.252168 | 1.267617 |
| | | R3 | Mean | 65.68977 | 79.16839 | 85.64454 | 97.3838 | 96.54659 |
| | | | Std | 5.253859 | 3.954718 | 7.536769 | 1.24352 | 1.238074 |
| | Thresholding 0.9 | R1 | Mean | 57.03541 | 72.07664 | 69.11076 | 98.08373 | 95.83783 |
| | | | Std | 10.11084 | 8.867493 | 13.20961 | 0.996802 | 1.382573 |
| | | R2 | Mean | 59.61422 | 74.35359 | 73.17393 | 98.04841 | 96.26078 |
| | | | Std | 8.011962 | 6.959771 | 11.10347 | 1.003961 | 1.183293 |
| | | R3 | Mean | 61.8363 | 76.04695 | 74.9709 | 98.17883 | 96.51457 |
| | | | Std | 8.484042 | 7.190838 | 11.58667 | 0.980553 | 1.155411 |

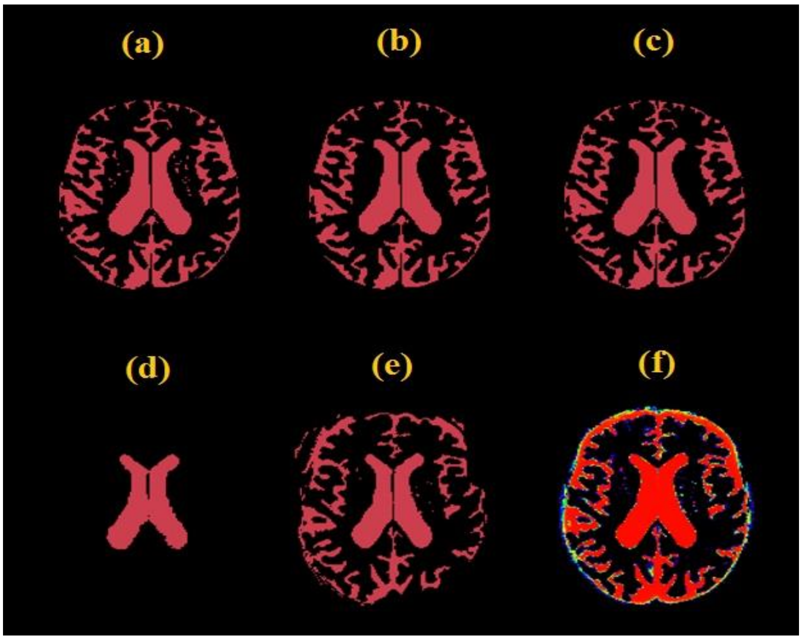

Figure 6. Example of CSF segmentation. The manual segmentation by three radiologists: R1 (a), R2 (b), and R3 (c), the automated segmentation using three methods: Freesurfer (d), FSL (e), and SPM12 (f)

Table 4. Comparison of CSF segmentation metrics of FSL, Freesurfer and SPM and manual segmentation

| Segmentation method | | Radiologist | Statistic | Jaccard | Dice | Sensitivity | Specificity | Accuracy |
|---|---|---|---|---|---|---|---|---|
| FSL | | R1 | Mean | 50.0069 | 65.6336 | 60.3501 | 98.8482 | 96.7721 |
| | | | Std | 12.41752 | 12.81171 | 14.72502 | 0.615634 | 0.90462 |
| | | R2 | Mean | 50.02659 | 65.7913 | 57.61453 | 99.04669 | 96.55922 |
| | | | Std | 11.54664 | 11.91468 | 13.25455 | 0.54863 | 0.981143 |
| | | R3 | Mean | 52.00207 | 67.47785 | 59.55206 | 99.10553 | 96.76505 |
| | | | Std | 12.0162 | 12.19746 | 13.7459 | 0.554026 | 0.990232 |
| Freesurfer | | R1 | Mean | 11.35384 | 18.22935 | 11.54445 | 99.91823 | 95.10053 |
| | | | Std | 12.68276 | 19.24673 | 12.87491 | 0.171028 | 1.342731 |
| | | R2 | Mean | 10.44725 | 17.0716 | 10.5531 | 99.93241 | 94.52765 |
| | | | Std | 11.5349 | 17.90752 | 11.62404 | 0.166433 | 1.388738 |
| | | R3 | Mean | 10.88002 | 17.64334 | 10.95779 | 99.94265 | 94.64422 |
| | | | Std | 12.03045 | 18.50282 | 12.07964 | 0.16533 | 1.399666 |
| SPM | Thresholding 0.5 | R1 | Mean | 58.91496 | 73.44195 | 79.62271 | 98.0298 | 96.99108 |
| | | | Std | 11.8456 | 9.573757 | 14.37381 | 1.060765 | 0.971219 |
| | | R2 | Mean | 59.73744 | 73.98816 | 75.87836 | 98.30459 | 96.93187 |
| | | | Std | 12.69466 | 10.28521 | 14.59359 | 0.900564 | 1.027294 |
| | | R3 | Mean | 61.02265 | 74.97436 | 77.42321 | 98.33206 | 97.07722 |
| | | | Std | 12.91709 | 10.35465 | 14.4423 | 0.927343 | 1.025664 |
| | Thresholding 0.7 | R1 | Mean | 58.68849 | 73.13035 | 74.25042 | 98.55598 | 97.19218 |
| | | | Std | 12.77488 | 10.54099 | 16.34549 | 0.916176 | 0.930316 |
| | | R2 | Mean | 58.92161 | 73.19515 | 70.47204 | 98.79669 | 97.06713 |
| | | | Std | 13.61711 | 11.33221 | 15.99604 | 0.771842 | 1.013906 |
| | | R3 | Mean | 60.25214 | 74.22505 | 71.91474 | 98.82566 | 97.21178 |
| | | | Std | 13.86997 | 11.39367 | 16.01185 | 0.790879 | 1.008484 |
| | Thresholding 0.9 | R1 | Mean | 54.93454 | 69.76866 | 63.73864 | 99.19204 | 97.21809 |
| | | | Std | 14.21049 | 12.68458 | 18.18987 | 0.662016 | 0.944241 |
| | | R2 | Mean | 53.93664 | 68.79449 | 59.8737 | 99.36749 | 96.95811 |
| | | | Std | 14.87703 | 13.48679 | 17.34354 | 0.534868 | 1.041973 |
| | | R3 | Mean | 55.37343 | 69.98359 | 61.2309 | 99.39974 | 97.11465 |
| | | | Std | 15.12353 | 13.52182 | 17.43519 | 0.551705 | 1.028414 |

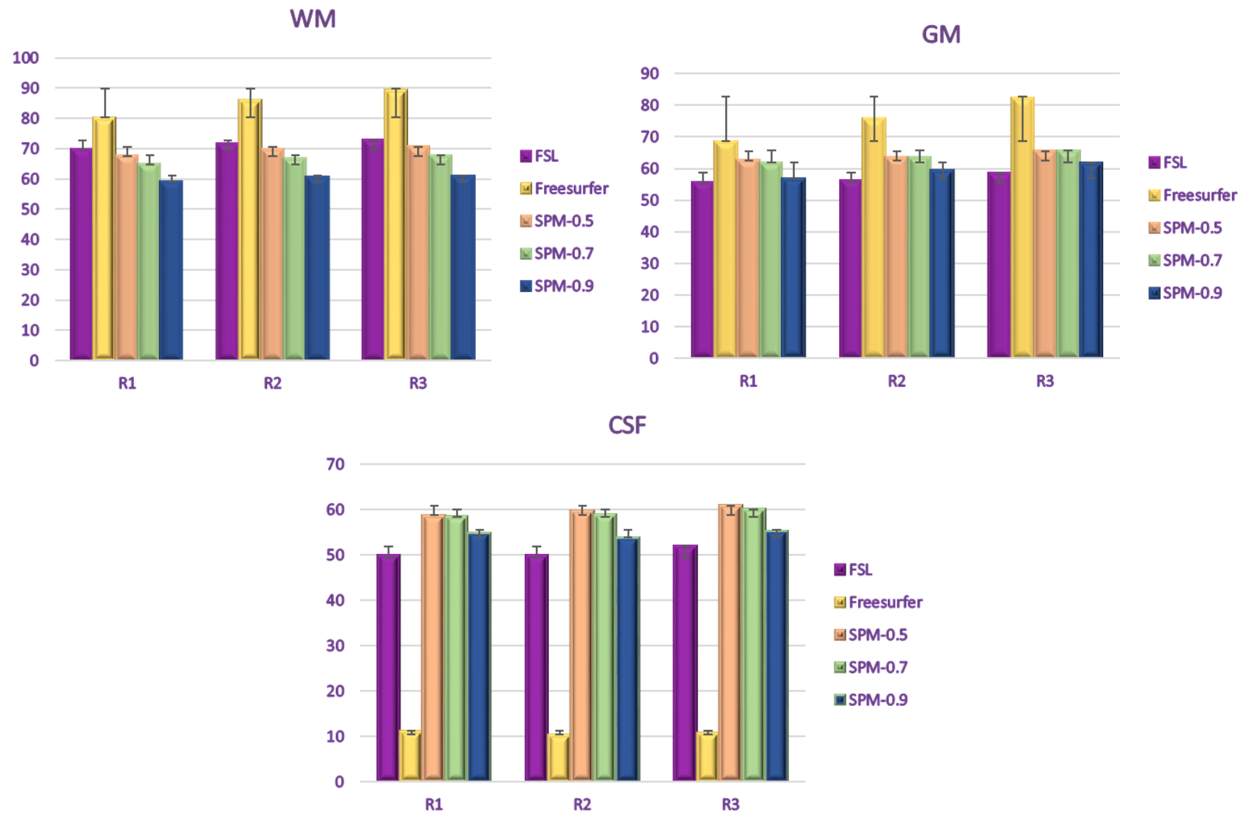

Figure 7. Dice similarity metric for segmented Brain from different tissues

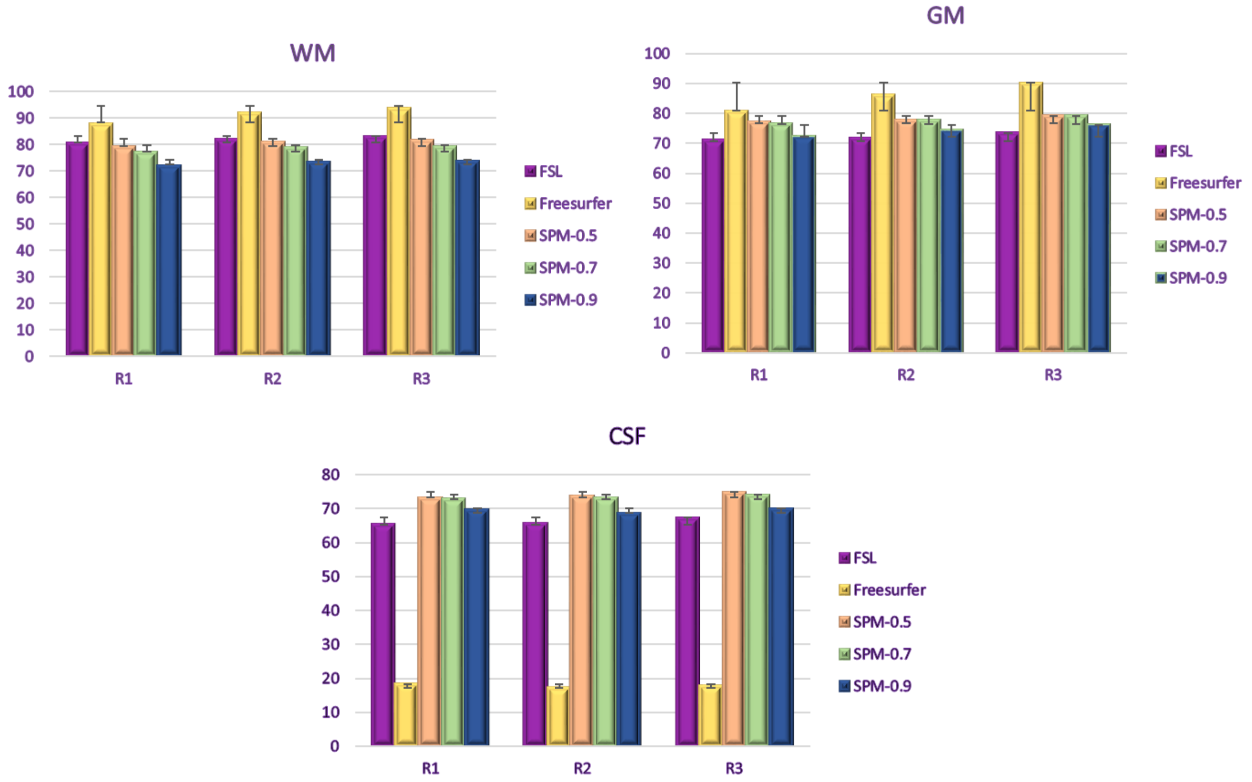

Figure 8. Jaccard similarity metric for segmented Brain from different tissues

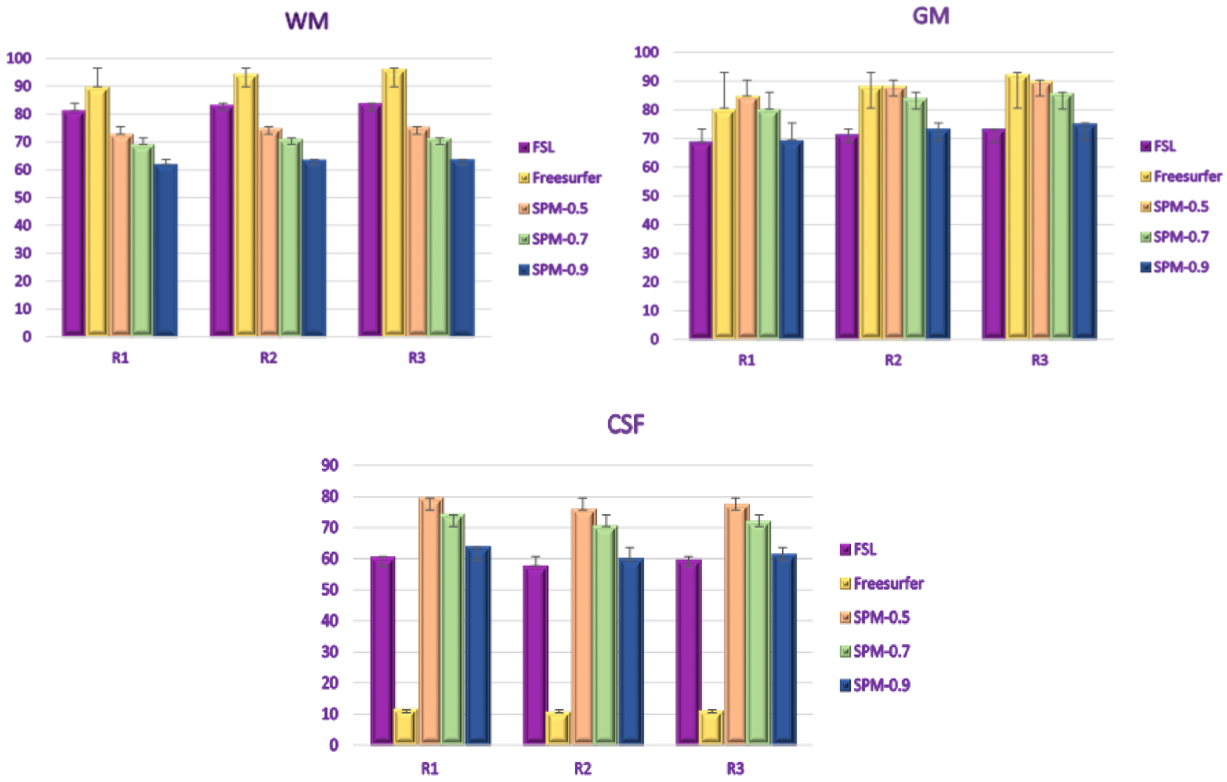

Figure 9. Sensitivity metric for segmented Brain from different tissues

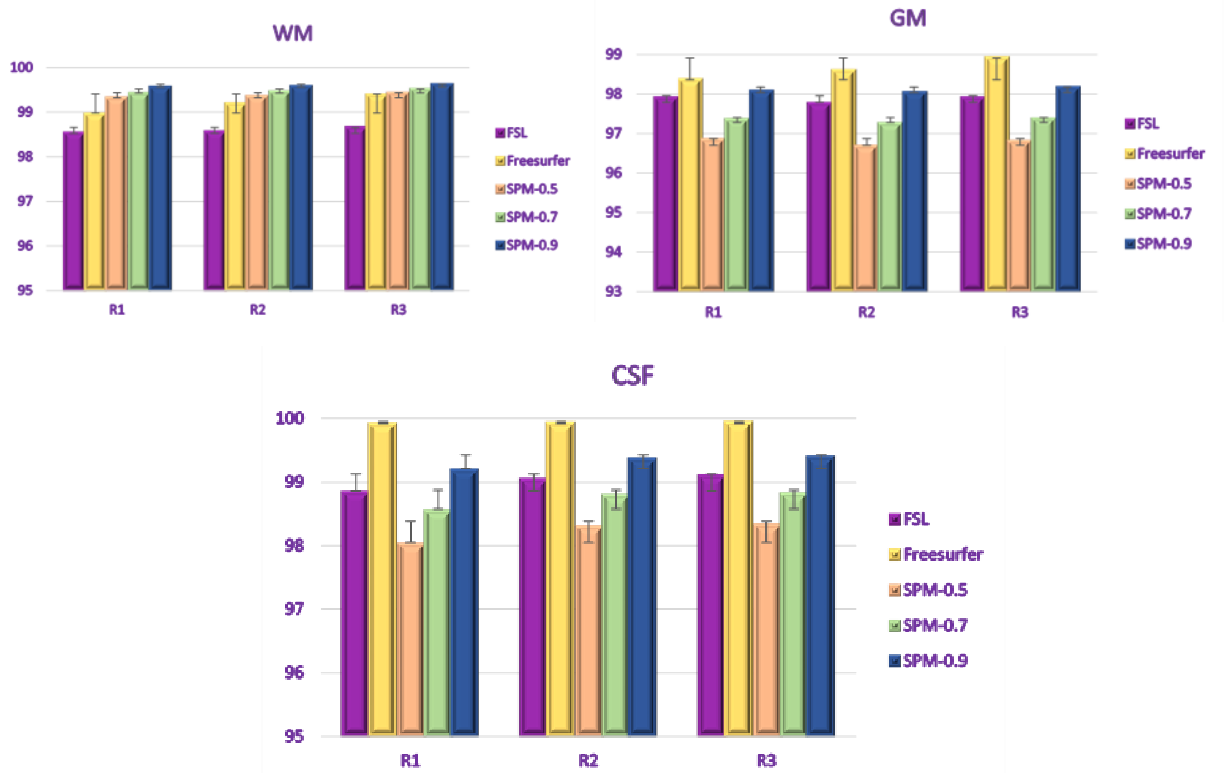

Figure 10. Specificity metric for segmented Brain from different tissues

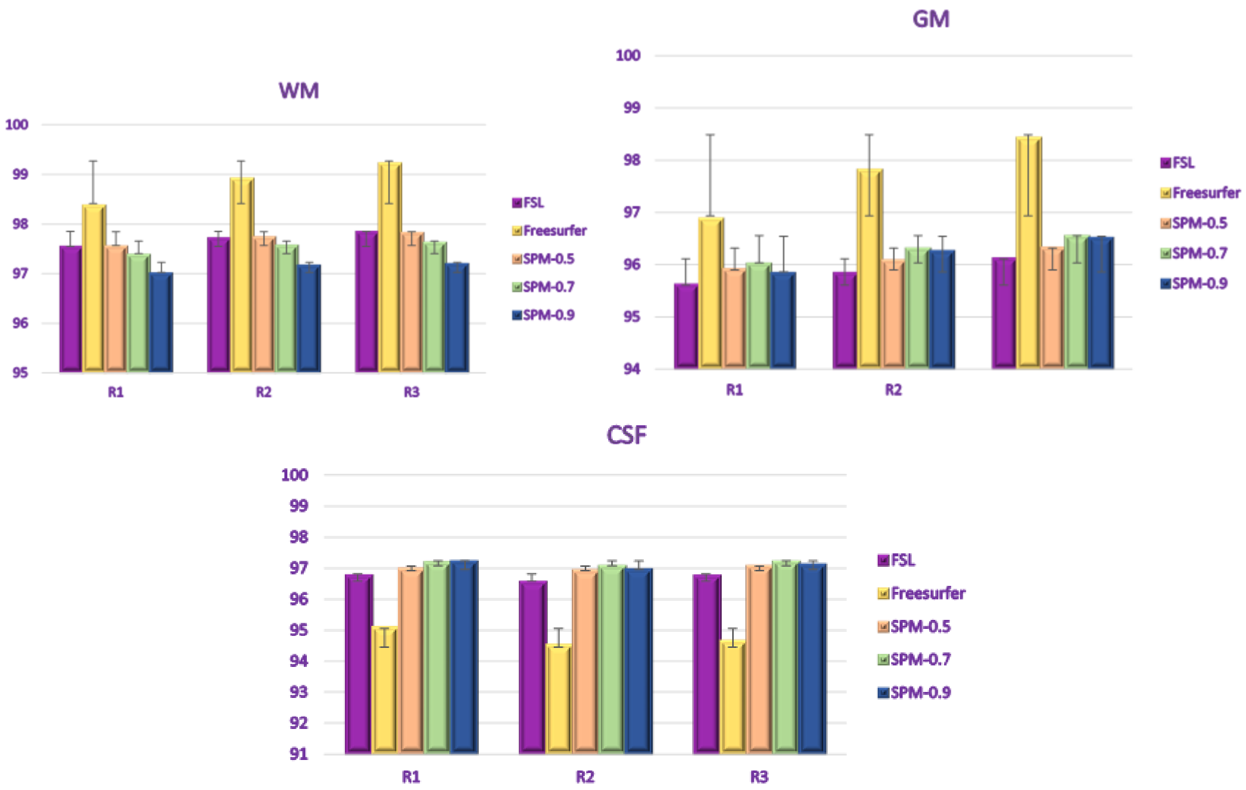

Figure 11. Accuracy metric for segmented Brain from different tissues

## Discussion

Brain tissue extraction is a time-consuming step in neuroimage processing. While there are many ways to extract the brain to perform this step automatically, their performance varies, and it may affect the results of the next steps of processing. Besides the various techniques provided for the automatic brain tissue segmentation of brain MR images, there are software packages that are mostly used in neuroimage processing. Among them, FSL, Freesurfer and SPM12 software packages are commonly used by scientists to analyze the structure and function of the brain.

Comparing the accuracy of segmentation methods based on past articles is difficult due to different evaluation measures, different manual segmentation protocols and, most importantly, different imaging data. However, we have reviewed some studies that have used different automated segmentation methods based on segmented textures in Table 5.

Kazemi et al. [40] evaluated the segmentation accuracy of GM, WM and CSF tissues using Brainweb simulated MR images and IBSR real MR images. They examined the similarities calculated between segmented tissues using different tools to introduce a suitable tool for segmentation.

Kasiri et al. [49] provided a comprehensive comparative evaluation of the three most commonly used neuroimage analysis software. This study investigated brain segmentation in GM, WM and CSFs by SPM8, FSL 4.1 and BrainSuite 9.1. The experiment results on simulated and real datasets showed that BrainSuite provides better performance in WM and GM classification compared to the other two methods, and SPM8 has significant accuracy in detecting CSF from other tissues. Palombo et al. [50] segmented gray matter (GM), white matter (WM), and subcortical structures into MRI data from healthy volunteers from the Kirby-21 and OASIS datasets. They used Pearson correlation (r), Bland-Altman diagram and dice index to evaluate reproducibility. There was a high correlation between volume measurements for both SPM and FS methods. SPM systematically provides significantly more GM volume and less WM and subcortex volume with respect to FS.

In this study, we have evaluated the segmentation performance of these tools through a series of metrics. We performed all the segmentations consisting of pre-processing and tissue segmentation by each tool, and the manual segmentation method was used as the reference standard. Our analysis was based on a comparison of automatic and manual segmentation. The results show that brain tissue segmentation tools work differently in terms of quantitative parameters. The difference in this performance is due to the fact that each of these tools is designed for a specific purpose. As can be seen, the performance of all three software in the WM and GM segmentation is acceptable, and among them, Freesurfer results has provided better. But in the case of the CSF tissue, Freesurfer performed poorly, and FSL and SPM tools performed better, especially on the probability map with threshold 0.5.

Table 5. Segmentation Studies

| Reference | Brain Regions | Methods |
|---|---|---|
| Kazemi et al. (2014) | GM, WM, CSF | SPM, FSL, Brainsuit |
| Kasiri et al. (2010) | GM, WM, CSF | SPM, FSL, Brainsuit |
| Palumbo et al. (2019) | GM, WM, Subcortical structures | SPM, Freesurfer |
| This Study | GM, WM, CSF | SPM, FSL, Freesurfer |

On the other hand, the success rate of the tools in the brain tissues segmentation were also examined. Quantitative evaluation using the metrics of sensitivity, specificity and accuracy shows that in both WM and GM tissues, Freesurfer performs better than the others, while in the CSF, it has low sensitivity and accuracy, but its specificity comparing to SPM and FSL has been better.

## Conclusion

This study provides a objective comparison of automated methods for segmentation of GM, WM, CSF brain tissues. We evaluated and compared three common software packages (FreeSurfer, FSL and SPM12) and provided an assessment of the overall performance of the methods evaluated by diagnostic criteria (Dice, Jaccard, Sensitivity, Specificity and Accuracy) that can be helped in selecting the best method for a segmentation purpose.

## References

1. Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. Computational and mathematical methods in medicine. 2015 Oct;2015.
2. Pal NR, Pal SK. A review on image segmentation techniques. Pattern recognition. 1993 Sep 1;26(9):1277-94.
3. Chan TF, Vese LA. Active contours without edges. IEEE Transactions on image processing. 2001 Feb;10(2):266-77.
4. Boykov Y, Funka-Lea G. Graph cuts and efficient ND image segmentation. International journal of computer vision. 2006 Nov;70(2):109-31.
5. Unnikrishnan R, Pantofaru C, Hebert M. Toward objective evaluation of image segmentation algorithms. IEEE transactions on pattern analysis and machine intelligence. 2007 Apr 23;29(6):929-44.
6. Cai W, Chen S, Zhang D. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern recognition. 2007 Mar 1;40(3):825-38.
7. Prastawa M. An MRI segmentation framework for brains with anatomical deviations. The University of North Carolina at Chapel Hill; 2007.
8. Shi F, Shen D, Yap PT, Fan Y, Cheng JZ, An H, et al. CENTS: cortical enhanced neonatal tissue segmentation. Human brain mapping. 2011 Mar;32(3):382-96.
9. Wang L, Shi F, Li G, Gao Y, Lin W, Gilmore JH, Shen D. Segmentation of neonatal brain MR images using patch-driven level sets. NeuroImage. 2014 Jan 1;84:141-58.
10. Weisenfeld NI, Warfield SK. Automatic segmentation of newborn brain MRI. Neuroimage. 2009 Aug 15;47(2):564-72.
11. Moreno JC, Prasath VS, Proenca H, Palaniappan K. Fast and globally convex multiphase active contours for brain MRI segmentation. Computer Vision and Image Understanding. 2014 Aug 1;125:237-50.
12. Lenchik L, Heacock L, Weaver AA, Boutin RD, Cook TS, Itri J, et al. Automated segmentation of tissues using CT and MRI: a systematic review. Academic radiology. 2019 Dec 1;26(12):1695-706.
13. Katyal R, Paneri S, Kuse M. Gaussian intensity model with neighborhood cues for fluid-tissue categorization of multisequence MR brain images. Proceedings of the MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS′13). 2013 Sep.
14. Rajchl M, Baxter JS, McLeod AJ, Yuan J, Qiu W, Peters TM, et al. ASeTs: MAP-based brain tissue segmentation using manifold learning and hierarchical max-flow regularization. Proceedings of the MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS′13). 2013 Sep;375.
15. Mahmood Q, Alipoor M, Chodorowski A, Mehnert A, Persson M. Multimodal MR brain segmentation using Bayesian-based adaptive mean-shift (BAMS). InProceedings of the MICCAI Workshops—The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS′13). 2013 Sep.
16. Sarikaya D, Zhao L, Corso JJ. Multi-atlas brain MRI segmentation with multiway cut. InProceedings of the MICCAI Workshops—The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS′13) 2013 Sep.
17. Wang C, Smedby Ö. Fully automatic brain segmentation using model-guided level sets and skeleton-based models. MIDAS J. 2013 Sep;10.
18. van Opbroek A, van der Lijn F, de Bruijne M. Automated brain-tissue segmentation by multi-feature SVM classification. InProceedings of the MICCAI Workshops—The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS′13). 2013 Sep.
19. Moeskops P, Viergever MA, Benders MJ, Išgum I. Evaluation of an automatic brain segmentation method developed for neonates on adult MR brain images. InMedical Imaging 2015: Image Processing. 2015 Mar 20; 9413:304-9.
20. Wang C, Wang Q, Smedby Ö. Automatic heart and vessel segmentation using random forests and a local phase guided level set method. InReconstruction, Segmentation, and Analysis of Medical Images. 2016 Oct 17: 159-64.

21. Chen H, Dou Q, Yu L, Qin J, Heng PA. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage. 2018 Apr 15;170:446-55.
22. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. InInternational conference on medical image computing and computer-assisted intervention. 2016 Oct 17; 424-32.
23. Andermatt S, Pezold S, Cattin P. Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data. InDeep learning and data labeling for medical applications. 2016 Oct 21;142-51.
24. Stollenga MF, Byeon W, Liwicki M, Schmidhuber J. Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. Advances in neural information processing systems. 2015;28.
25. Lenchik L, Heacock L, Weaver AA, Boutin RD, Cook TS, Itri J, et al. Automated segmentation of tissues using CT and MRI: a systematic review. Academic radiology. 2019 Dec 1;26(12):1695-706.
26. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002 Jan 31;33(3):341-55.
27. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012 Aug 15;62(2):782-90.
28. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004 Jan 1;23:S208-19.
29. Zhang Y, Brady JM, Smith SM. An hmrf-em algorithm for partial volume segmentation of brain mri fmrib technical report tr01yz1. Technical Report, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain; 2001 Jan 24.
30. Fischl B. FreeSurfer. Neuroimage. 2012 Aug 15;62(2):774-81.
31. Ashburner J, Friston K, Holmes A, Poline JB, Rouquette S, Worsley K. Statistical parametric mapping, SPM′99. The Welcome Department of Cognitive Neurology. Institute of Neurology, University College London. 1999.
32. The Image and Data Archive (IDA) is a secure online resource for archiving, exploring and sharing neuroscience data. Available from: https://ida.loni.usc.edu/home/projectPage.jsp?project=ADNI
33. Yushkevich P. ITK-SNaP integration, NLM insight. Available from: http://www.itksnap.org/pmwiki/pmwiki.php.
34. FSL - FslWiki [Internet]. fsl.fmrib.ox.ac.uk. Available from: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
35. Dale AM. Optimal experimental design for event-related fMRI. Human brain mapping. 1999;8(2-3):109-14.
36. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage. 1999 Feb 1;9(2):179-94.
37. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences. 2000 Sep 26;97(20):11050-5.
38. FreeSurfer [Internet]. Harvard.edu. 2013. Available from: https://surfer.nmr.mgh.harvard.edu/
39. SPM12 Software - Statistical Parametric Mapping [Internet]. www.fil.ion.ucl.ac.uk. Available from: https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
40. Kazemi K, Noorizadeh N. Quantitative comparison of SPM, FSL, and brainsuite for brain MR image segmentation. Journal of biomedical physics & engineering. 2014 Mar;4(1):13.
41. Ashburner J, Barnes G, Chen CC, Daunizeau J, Flandin G, Friston K, et al. SPM12 manual. Wellcome Trust Centre for Neuroimaging, London, UK. 2014 Jun 26;2464.
42. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945 Jul;26(3):297-302.
43. Jaccard P. The distribution of the flora in the alpine zone. 1. New phytologist. 1912 Feb;11(2):37-50.
44. Moore, C., Bell, D. Dice similarity coefficient. Reference article, Radiopaedia.org. 2022. https://doi.org/10.53347/rID-75056
45. Moore, C. Jaccard Index. Reference article, Radiopaedia.org. Available from: https://radiopaedia.org/articles/jaccard-index.
46. Parikh R, Mathai A, Parikh S et-al. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56 (1): 45-50.
47. Accuracy IS. of measurement methods and results—part 1: General principles and definitions. International Organization for Standardization, Geneva, Switzerland. 1994.
48. Metz, CE (October 1978). "Basic principles of ROC analysis" (PDF). Semin Nucl Med. 8 (4): 283–98. doi:10.1016/s0001-2998(78)80014-2. PMID 112681
49. Kasiri K, Javad DM, Kazemi K, Sadegh HM, Kafshgari S. Comparison evaluation of three brain MRI segmentation methods in software tools. In2010 17th Iranian Conference of Biomedical Engineering (ICBME). 2010 Nov 3; 1-4.
50. Palumbo L, Bosco P, Fantacci ME, Ferrari E, Oliva P, Spera G, et al. Evaluation of the intra-and inter-method agreement of brain MRI segmentation software packages: A comparison between SPM12 and FreeSurfer v6. 0. Physica Medica. 2019 Aug 1;64:261-72.