

# Validated Automatic Brain Extraction of Head CT Images using FSL-Brain Extraction Tool (BET)

John Muschelli<sup>c,\*</sup>, Natalie Ullman<sup>d</sup>, Ciprian M. Crainiceanu<sup>c</sup>

<sup>a</sup>*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA*

<sup>b</sup>*Department of Neurology, Division of Brain Injury Outcomes, Johns Hopkins Medical Institutions, Baltimore, MD, USA*

---

## Abstract

### 1. Background

Computed Tomography (CT) brain imaging is commonly used in diagnostic settings. Although CT scans are primarily used in a clinical setting, they can be used to answer research hypotheses. Methods for brain extraction in head CT images has been informally proposed, but never formally validated.

### 2. Aim

To systematically analyze the performance of FSL's brain extraction tool (BET) on head CT images of patients with intracranial hemorrhage by varying its parameters and options after performing CT-specific preprocessing, and quantitatively comparing its results to a manual gold standard.

### 3. Methods

Nineteen images from 16 patients with intracranial hemorrhage were selected from 13 different MISTIE stroke trial centers. Before running BET, the image was: (1) not smoothed or smoothed using a 1mm Gaussian smoother, (2) thresholded to 0 – 100 Hounsfield units (HU), (3) BET was applied using 1 of 3 fractional intensity (FI) thresholds: 0.01, 0.1, 0.35, (4) holes in the mask were filled. For comparison, intracranial masks were manually created for all image volumes from expert CT readers. The resulting brain tissue masks were quantitatively compared to the manual segmentations using sensitivity, specificity, accuracy, and the Dice Similarity Index (DSI). Different results were compared using Wilcoxon signed-rank tests.

### 4. Results

Without smoothing the images, BET performed poorly regardless of FI. After smoothing, performance was high for all metrics for all scans. All reported measures are on smoothed data. Using an FI of 0.01 or 0.1 performed better than 0.35; we will focus and compare these FI. Using an FI of 0.01 had a higher median sensitivity (0.9921) than an FI of 0.1 (0.9905,  $p < 0.001$ ), lower specificity (0.9979 vs. 0.998;  $p < 0.001$ ), , and no difference in accuracy (0.9971 vs. 0.9971;  $p = 0.134$ ) or DSI (0.9894 vs. 0.9896;  $p = 0.066$ ). Overall, regardless of p-values, these measures are all high in practice and largely adequate for good brain extraction.

---

\*Principal Corresponding Author

Email addresses: jmusche1@jhu.edu (John Muschelli), nullman1@jhmi.edu (Natalie Ullman), ccrainic@jhsph.edu (Ciprian M. Crainiceanu)

## 5. Conclusion

BET performs well at brain extraction on thresholded, 1mm smoothed CT images with an FI of 0.01. Smoothing before applying BET is an important step not previously discussed. Analysis code is provided.

---

# Validated Automatic Brain Extraction of Head CT Images using FSL-Brain Extraction Tool (BET)

John Muschelli<sup>c,\*</sup>, Natalie Ullman<sup>d</sup>, Ciprian M. Crainiceanu<sup>c</sup>

<sup>c</sup>*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA*

<sup>d</sup>*Department of Neurology, Division of Brain Injury Outcomes, Johns Hopkins Medical Institutions, Baltimore, MD, USA*

---

**Keywords:** CT, skull stripping, brain extraction, validation

---

## 6. Introduction

X-ray computed tomography (CT) scanning is widely available and is a commonly used diagnostic tool in clinical settings [1]. Though much analysis of CT images are done by qualitative visual inspection, detailed quantification of information is of interest. CT scans do not discriminate in the tissues or structures captured; scans have non-brain structures such as the skull, eyes, facial and nasal features, extracranial skin, and more importantly non-human elements captured by the scanner, such as pillows or medical devices. These objects present a problem for extracting metrics on brain tissue. We propose a validated automated solution to brain extraction in head CT scans using established neuroimaging software.

In magnetic resonance imaging (MRI), brain extraction has been extensively studied and investigated (see Wang et al. [2] for an extensive list). As such, many pieces of software and algorithms exist to achieve this goal with respect to MRI scans. We have adapted one algorithm, the Brain Extraction Tool (BET) [3], a function of the FSL [4] neuroimaging software (v5.0.4), to automatically extract the brain from a CT scan. Variations of this pipeline have been presented before in Solomon et al. [5], and has been replicated in slightly more detail in Rorden et al. [6]. Neither presented a formal validation against a set of manually segmented brain images, which is the goal of our study.

## 7. Methods

### 7.1. Participants and CT data

We used CT images patients enrolled in the MISTIE (Minimally Invasive Surgery plus recombinant-tissue plasminogen activator for Intracerebral Evacuation) and ICES (Intraoperative CT-Guided Endoscopic Surgery) trials. These patients had an intracranial hemorrhage at time of scanning; for inclusion criteria, see Mould et al. [7]. CT data were collected as part of the Johns Hopkins Medicine IRB-approved MISTIE research studies with written consent from participants.

### 7.2. Imaging Data

The study protocol was executed with minor, but important, differences across the 13 sites. Slice thickness of the image varied within the scan for 2 scans, referred to as variable slice thickness. For example, a scan may have 10 millimeter (mm) slices at the top and bottom of the brain but with 5mm slices in the middle of the brain. Therefore, the scans analyzed had different voxel (volume element) dimensions and image resolution prior to registration to the template. These conditions represent how scans are presented for evaluation in many diagnostic cases.

---

\*Principal Corresponding Author

Email addresses: [jmusche1@jhu.edu](mailto:jmusche1@jhu.edu) (John Muschelli), [nullman1@jhmi.edu](mailto:nullman1@jhmi.edu) (Natalie Ullman), [ccrainic@jhsph.edu](mailto:ccrainic@jhsph.edu) (Ciprian M. Crainiceanu)

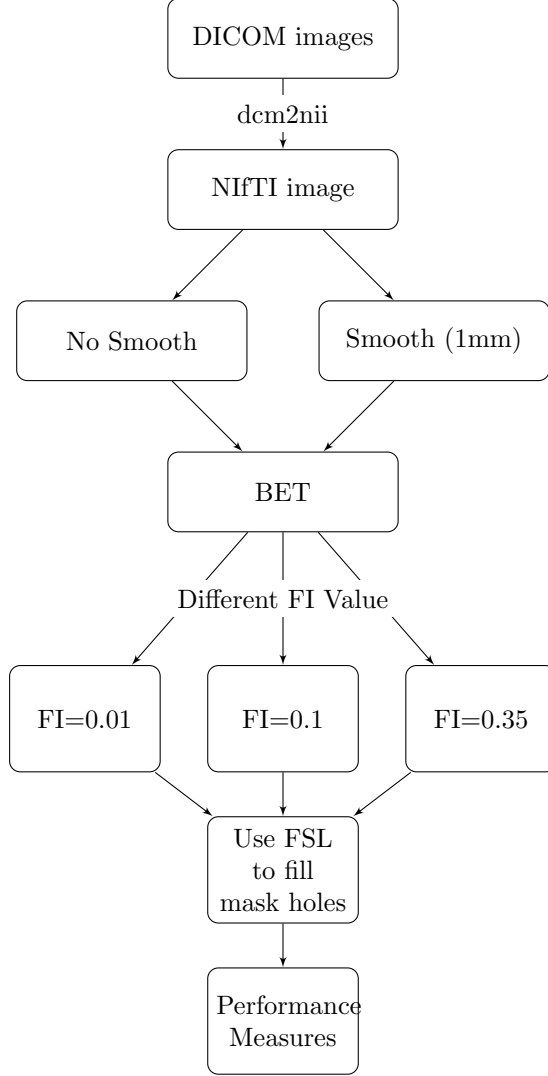


Figure 1: Overall workflow of processing of data. After the raw data has been processed and areas have been extracted, surfaces can be rendered. We are concerned with those steps in orange: creating surfaces, and export to the web. The last branch shows 2 options for export: publishing the figure to the web or enclosing it in a folder with all tools to render it. The second option would allow users to include these zipped directories as supplementary figures until more widely used.

### 7.3. Manual and Automated Brain Extraction

We analyzed 19 scans, corresponding to 16 unique patients. Brain tissue was manually segmented as a binary mask from DICOM (Digital Imaging and Communications in Medicine) images in the OsiriX imaging software (OsiriX v.4.1, Pixmeo; Geneva, Switzerland) by one expert reader (NU). CT brain images and the binary mask were exported from OsiriX to DICOM format. Images with gantry tilt were corrected using a customized MATLAB (The Mathworks, Natick, Massachusetts, USA) user-written script (<http://bit.ly/1ltIM8c>). Images were converted to the Neuroimaging Informatics Technology Initiative (NIfTI) data format using `dcm2nii` (2009 version, provided with MRIcro [8]). Images were constrained to values  $-1024$  and  $3071$  HU to remove potential image rescaling errors and artifacts. No interpolation was done for images with a variable slice thickness. Thickness was determined from the first slice converted and was assumed homogeneous throughout the image. The image processing pipeline can be seen in Figure 1.

Images were thresholded to a brain tissue range (0-100 HU). We either applied BET to the this image or smoothed this image with a Gaussian kernel ( $\sigma = 1\text{mm}$ , using FSL), re-thresholded to 0-100 HU and then applied BET. When BET was applied, we varied fractional intensity (FI) parameter to determine its

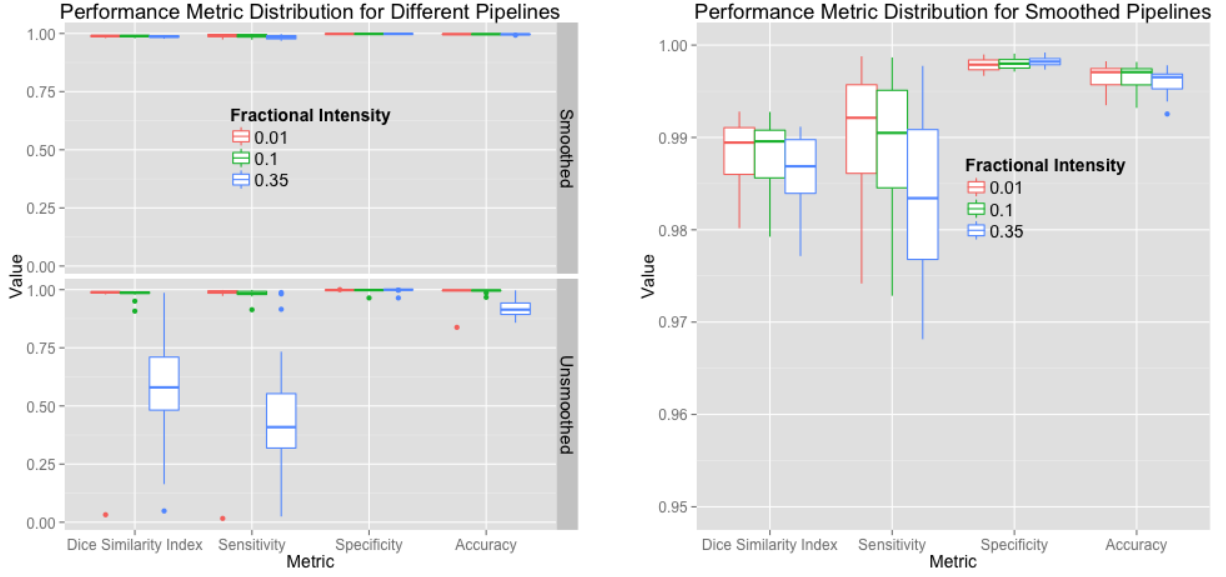


Figure 2: **Performance Metric Distribution for Different Pipelines.** Panel A displays the boxplots for performance measures when running the pipeline with a different fractional intensity (FI), using smoothed data (top) or unsmoothed data (bottom). Panel B presents the smoothed data only, rescaled to show discrimination between the different FI. Overall, FI of 0.01 and 0.1 perform better than 0.35 in all categories other than specificity. Using smoothed data improves performance in all performance metrics, markedly when an FI of 0.35 is used. Panel B demonstrates that using an FI of 0.01 on smoothed data is the best pipeline.

influence on performance: we used values of 0.35 (as recommended in Rorden et al. [6]), 0.1, 0.01.

We also present one example case which demonstrates that brain extraction performance was acceptable only after smoothing before BET was applied.

#### 7.4. Measuring and Testing Brain Extraction Performance

Five common measurements of performance were calculated for each image: sensitivity, specificity, accuracy, and the Dice Similarity Index (DSI). See Inline Supplementary Methods 1 for the calculation of each measure.

[Insert Supplementary Methods 1 here]

Performance was compared across FI using Wilcoxon signed-rank tests.

## 8. Results

Figure 2B illustrates the performance of each variation of the BET pipeline in Figure 1. The smoothed pipelines perform better than the unsmoothed pipelines and BET performs poorly on some scans without smoothing.

Figure 2A displays the performance for brain extraction in the smoothed pipelines (note the change in the y-axis). After smoothing, performance was high for all metrics for all scans. All reported measures are on smoothed data. Using an FI of 0.01 or 0.1 performed better than 0.35; we will focus and compare these FI. Using an FI of 0.01 had a higher median sensitivity (0.9921) than an FI of 0.1 (0.9905,  $p < 0.001$ ), lower specificity (0.9979 vs. 0.998;  $p < 0.001$ ), and no difference in accuracy (0.9971 vs. 0.9971;  $p = 0.134$ ) or DSI (0.9894 vs. 0.9896;  $p = 0.066$ ). Overall, regardless of p-values, these measures are all high in practice and largely adequate for brain extraction.

Although Figure 2 displays that using FI of 0.01 or 0.1 provides adequate results of brain extraction for most cases, they perform relatively well regardless of smoothing the data. Figure 3 displays an example where using unsmoothed data performs poorly for these FI, demonstrating why smoothing is essential for a general brain extraction procedure for CT.

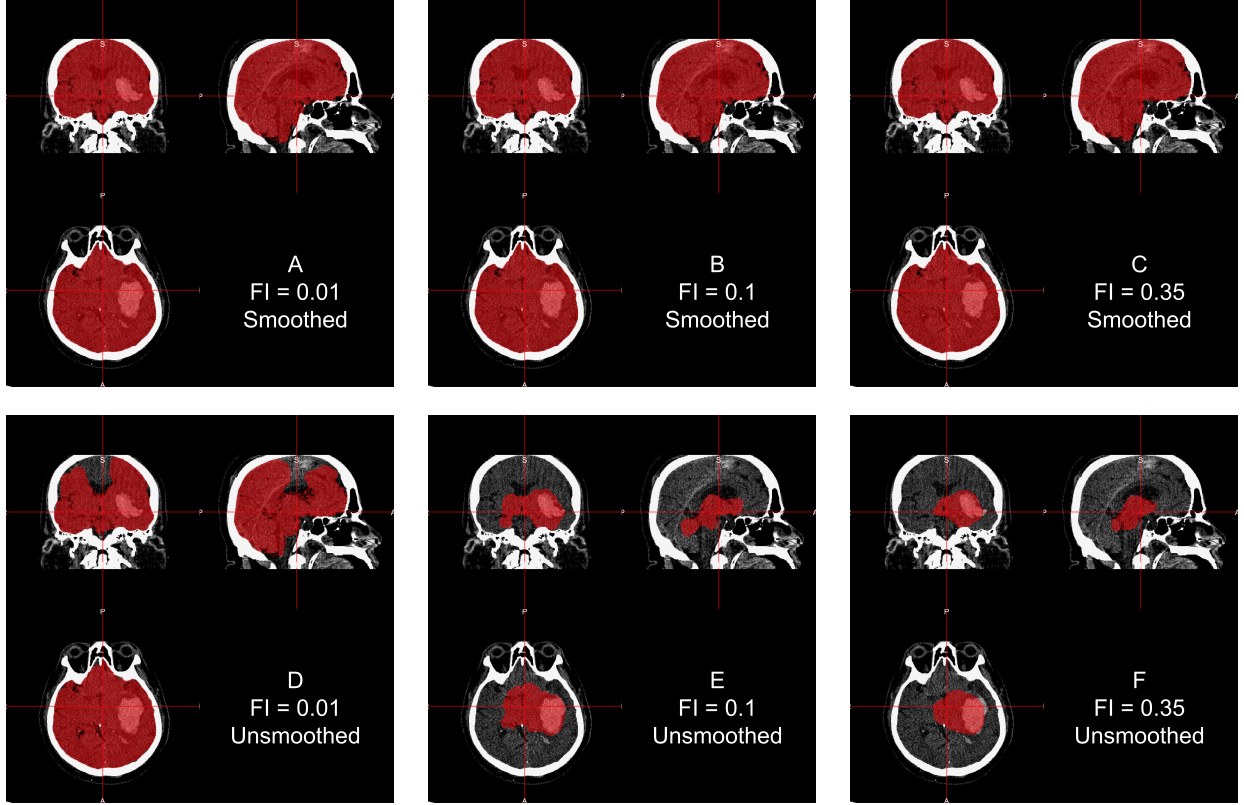


Figure 3: **Example Demonstrating How Smoothing Prior to BET is Essential** For one subject, the CT image is displayed with the brain extracted mask in red after running all pipelines. Panels A, B, and C represent applying BET using FI of 0.01, 0.01, and 0.35, respectively, to smoothed data. Panels D, E, and F correspond to applying BET using FI 0.01, 0.01, and 0.35 on unsmoothed data. Using smoothed data is required for adequate brain extraction.

## 9. Conclusions

We present an automated brain extraction pipeline for CT images that was validated using gold-standard manual segmentations of brain tissues. Overall, we found that smoothing the data with a conservative smoother (1mm Gaussian kernel) and using an FI of 0.01 provides good brain extraction for the sample studied. We believe this pipeline is generalized for most brain CT scans. These software (FSL and R) are freely available to download under public licenses are open source. Thus, this pipeline should be widely available.

With the resulting automated brain masks output from this pipeline, researchers can now do within-tissue operations, such as intensity normalization, brain volume estimation, tissue/tumor/hemorrhage segmentation, etc.

The R function to perform brain extraction is located at [https://github.com/muschellij2/CT\\_BET/blob/master/Skull\\_Strip\\_Paper/CT\\_Skull\\_Strip\\_Example.R](https://github.com/muschellij2/CT_BET/blob/master/Skull_Strip_Paper/CT_Skull_Strip_Example.R).

## Inline Supplementary Methods 1

Let  $I_{ia}, I_{im}$  be the indicators that voxel  $i$  for the automatic and manual masks, respectively, such that  $I_i = 1$  if voxel  $i$  is labeled in the brain mask and 0 otherwise.

A true positive (TP) is defined when:  $I_{ia} = 1$  and  $I_{im} = 1$ , a false positive (FP):  $I_{ia} = 1$  and  $I_{im} = 0$ , a false negative (FN):  $I_{ia} = 0$  and  $I_{im} = 1$ , and a true negative (TN):  $I_{ia} = 0$  and  $I_{im} = 0$ . Sensitivity is defined as

$$\frac{\#TP}{\#TP + \#FN} = \frac{\sum_{i=1}^V (I_{ia} \times I_{im})}{\sum_{i=1}^V I_{im}}$$

and specificity is defined as

$$\frac{\#TN}{\#TN + \#FP} = \frac{\sum_{i=1}^V ((1 - I_{ia}) \times (1 - I_{im}))}{\sum_{i=1}^V (1 - I_{im})}$$

and overall accuracy is:

$$\frac{\#TN + \#TP}{\#TN + \#FN + \#TP + \#FP} = \frac{\sum_{i=1}^V (I_{ia} \times I_{im}) + ((1 - I_{ia}) \times (1 - I_{im}))}{\sum_{i=1}^V I_{ia} + \sum_{i=1}^V I_{im}}$$

The Dice Similarity Index is defined as:

$$\frac{2 \times \#TP}{\#TN + \#FN + \#TP + \#FP} = \frac{2 \times \sum_{i=1}^V (I_{ia} \times I_{im})}{\sum_{i=1}^V I_{ia} + \sum_{i=1}^V I_{im}}$$

## References

- [1] R. Sahni, J. Weinberger, Management of intracerebral hemorrhage, *Vascular Health and Risk Management* 3 (5) (2007) 701–709, ISSN 1176-6344, URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2291314/>.
- [2] Y. Wang, J. Nie, P.-T. Yap, G. Li, F. Shi, X. Geng, L. Guo, D. Shen, A. D. N. Initiative, et al., Knowledge-guided robust mri brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates, *PloS one* 9 (1) (2014) e77810.
- [3] S. M. Smith, Fast robust automated brain extraction, *Human Brain Mapping* 17 (3) (2002) 143155, ISSN 1097-0193, doi:\bibinfo{doi}{10.1002/hbm.10062}, URL <http://onlinelibrary.wiley.com/doi/10.1002/hbm.10062/abstract>.
- [4] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, S. M. Smith, FSL, *NeuroImage* 62 (2) (2012) 782–790, ISSN 1053-8119, doi:\bibinfo{doi}{10.1016/j.neuroimage.2011.09.015}, URL <http://www.sciencedirect.com/science/article/pii/S1053811911010603>.
- [5] J. Solomon, V. Raymont, A. Braun, J. A. Butman, J. Grafman, User-friendly software for the analysis of brain lesions (ABLE), *Computer methods and programs in biomedicine* 86 (3) (2007) 245–254.
- [6] C. Rorden, L. Bonilha, J. Fridriksson, B. Bender, H.-O. Karnath, Age-specific CT and MRI templates for spatial normalization, *NeuroImage* 61 (4) (2012) 957–965, ISSN 1053-8119, doi:\bibinfo{doi}{10.1016/j.neuroimage.2012.03.020}, URL <http://www.sciencedirect.com/science/article/pii/S1053811912002935>.
- [7] W. A. Mould, J. R. Carhuapoma, J. Muschelli, K. Lane, T. C. Morgan, N. A. McBee, A. J. Bistran-Hall, N. L. Ullman, P. Vespa, N. A. Martin, I. Awad, M. Zuccarello, D. F. Hanley, Minimally Invasive Surgery Plus Recombinant Tissue-type Plasminogen Activator for Intracerebral Hemorrhage Evacuation Decreases Perihematoma Edema, *Stroke* 44 (3) (2013) 627–634, ISSN 0039-2499, 1524-4628, doi:\bibinfo{doi}{10.1161/STROKEAHA.111.000411}, URL <http://stroke.ahajournals.org/content/44/3/627>.
- [8] C. Rorden, M. Brett, Stereotaxic Display of Brain Lesions, *Behavioural Neurology* 12 (4) (2000) 191–200, ISSN 0953-4180, doi:\bibinfo{doi}{10.1155/2000/421719}, URL <http://www.hindawi.com/journals/bn/2000/421719/abs/>.