Validated Automatic Brain Extraction of Head CT Images

John Muschelli[1], Natalie L. Ullman[2], Paul Vespa[3], Daniel F. Hanley[2], Ciprian M. Crainiceanu[1]

1 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
2 Department of Neurology, Division of Brain Injury Outcomes, Johns Hopkins Medical Institutions, Baltimore, MD, USA
3 Department of Neurosurgery, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Goal

Systematically analyze the performance of the brain extraction tool (BET) [3], a function of the FMRIB software library (FSL) [2], on head CT images of patients with intracranial hemorrhage by varying parameters of BET and the use of smoothing after performing CT-specific preprocessing by:

- Quantitatively comparing the results to the manual gold standard, and
- Estimating the performance using the intraclass correlation of serial CT scans.

Methods

Data were from patients with intracranial hemorrhage from MISTIE (Minimally Invasive Surgery plus recombinant-tissue plasminogen activator for Intracerebral Evacuation) stroke trial centers.

- Sample compared to gold standard: Twenty Two images from 19 patients.
- Intraclass Correlation Estimate: 1062 images from 133 patients, after excluding 115 scans for craniotomy or skull stripping failure (9.8%).

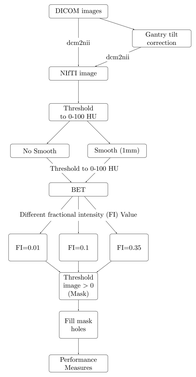


Figure: Processing Pipeline. Each image was thresholded using a $0 - 100$ Hounsfield units (HU) range. In one variant of the pipeline, data were smoothed using a Gaussian kernel ($\sigma = 1$mm) and re-thresholded to 0-100 HU; in the other, data were not smoothed. BET was applied using 1 of 3 fractional intensity (FI) thresholds: 0.01, 0.1, 0.35 and holes in the brain mask produced by BET were filled.

Measuring and Testing Brain Extraction Performance

Five common measurements of performance were calculated for each image comparing to the manually segmented images: sensitivity, specificity, accuracy, and the Dice Similarity Index (DSI) [1]. Testing paired difference of each measure using different pipelines (e.g. 0.01 vs. 0.1, smoothed data) was performed using Wilcoxon signed-rank test.

Figure: Performance Metric Distribution for Different Pipelines. Panel A displays the performance for brain extraction for the pipelines, panel B focuses on only those using smoothed images. Using an FI of 0.01 or 0.1 performed better than 0.35. Using an FI of 0.01 had a higher median sensitivity (0.9902) than an FI of 0.1 (0.9891, $p < 0.001$), lower specificity (0.998 vs. 0.998; $p < 0.001$), and no difference in accuracy (0.9971 vs. 0.9971; $p = 0.039$) or DSI (0.9892 vs. 0.9895).

We have code to do this!

- R code: `http://bit.ly/CTBET_RCODE`
- bash code: `http://bit.ly/CTBET_BASH`

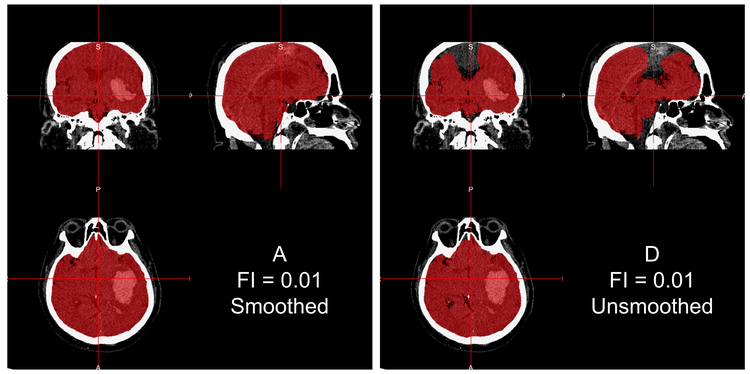
References

[1] Lee R. Dice. "Measures of the amount of ecologic association between species". In: *Ecology* 26.3 (1945), pp. 297–302.
[2] Mark Jenkinson et al. "FSL". In: *NeuroImage* 62.2 (Aug. 15, 2012), pp. 782–790.
[3] Stephen M. Smith. "Fast robust automated brain extraction". In: *Human Brain Mapping* 17.3 (2002), 143–155.

Sources of Funding

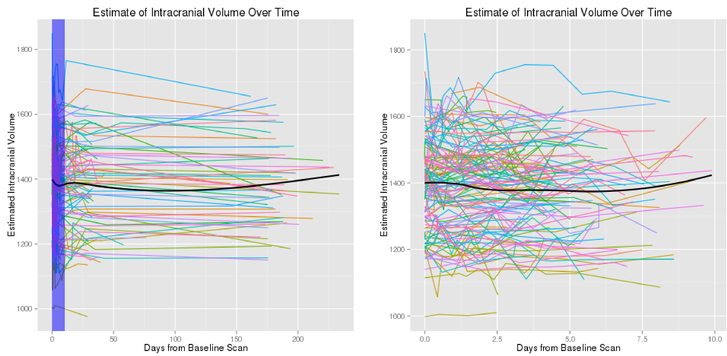
The project described and data used were supported by the NIH grants RO1EB012547, T32AG000247, R01NS046309, RO1NS060910, RO1NS085211, R01NS046309, U01NS080824, U01NS080824 and U01NS062851 and RO1MH095836.

Smoothing Images can Dramatically Increase Performance



Example Case where Smoothing before BET is Required. Panel A represents applying BET using FI of 0.01, to smoothed data. Panel D corresponds to applying BET using FI 0.01 on unsmoothed data. Smoothing images improves brain extraction with BET.

CT Skull Stripping leads to Consistent Intracranial Volume Estimates



Intracranial Volume (ICV) Estimate over Time. Each line represents an individual patient's ICV estimate over time. The data presented used an FI $= 0.01$ and smoothed data. The left panel shows all data used to estimate the intraclass correlation coefficient (ICC) of $0.93, (95\%\text{CI} : 0.91, 0.95)$.

Where does it fail?

Much more area than the brain is imaged | Patient had a craniotomy | CT ventricles are low intensity or enlarged

Conclusions

Smoothing the data using a conservative smoother (1mm Gaussian kernel) and using an FI of 0.01 provides good brain extraction.