Optimization of Brain Segmentation: Local or Global Partial Volume Estimation?

Mona Omidyeganeh[1], Alison Ross[1], Arnaud Voyer[1], Robert D. Vincent[1], Claude Y. Lepage[1], Lindsay B. Lewis[1], Seun Jeon[1], Patrick Bermudez[1], Alex Zijdenbos[1,2], Reza Adalat[1], Najmeh Khalili-Mahani[1], Alan C. Evans[1]

[1]McGill Center for Integrative Neuroscience, Montreal Neurological Institute, McGill University, Montreal, Canada, [2]Biospective Inc., Montreal, Canada

Introduction

Limited spatial resolution of brain MRI leads to the mixing of signals from different tissue types at the anatomical boundaries. Therefore, the accuracy of brain tissue segmentation relies on accurate partial volume effect (PVE) estimation[1,2]. However, the accuracy of PVE estimates is difficult to ascertain, and regional variations in brain anatomy and MR image contrast may affect these estimates. In this work, we compare the manual segmentation of cortical gray matter (GM) obtained by 2 human raters, against GM obtained by thresholding the GM PVE from the method described in[3].

Methods

We defined 16 random cubic regions of interest (ROI) (covering about 15% of total brain volume; the dimensions of the cubes vary between 15 to 50 voxels in each direction) in 8 T1-weighted brain MRIs of a single adult male subject (living phantom from the multisite Infant Brain Imaging Study), that were obtained using the same acquisition protocols. We focused on GM classification, which depends on optimization of 2 sets of PVE thresholds for GM/white matter (WM) and GM/cerebrospinal fluid (CSF). The "gold standard" was obtained from manual segmentation of ROIs (using Display of MINC tools[4]) by 2 independent raters.

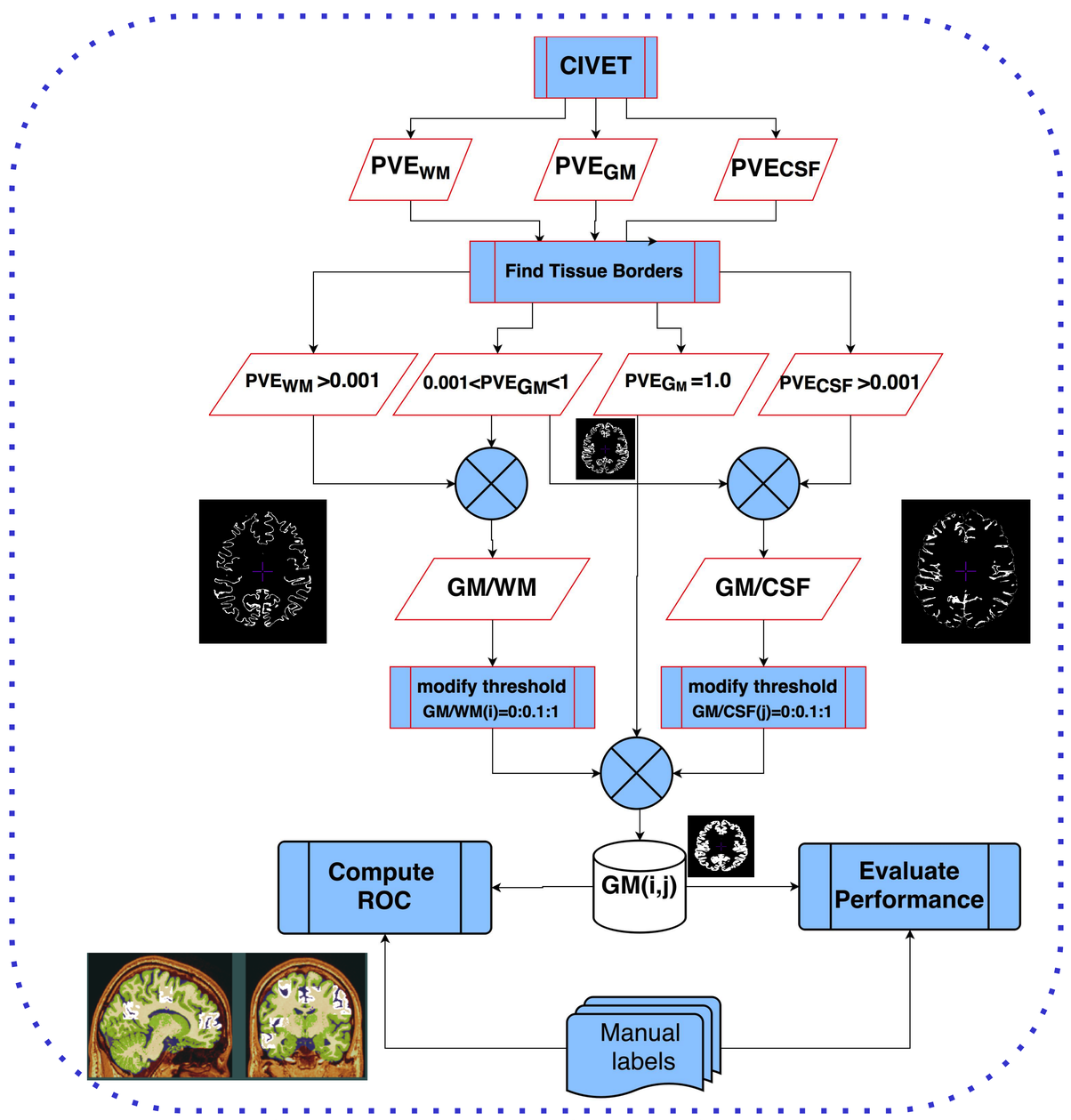

Figure 1. The flowchart of the experiments

The 8 MRIs were then processed through an automated spatial normalization[5], non-uniformity correction[6] and tissue classification pipeline[2, 3 and 7], which produced PVE-weighted (0-100%) tissue classes (Figure 1). The partial volume estimates for mixed tissue classes (GM/WM and GM/CSF) are thresholded at the 50% tissue content for both classes, globally, in the whole brain.

The receiver operating characteristic (ROC) curves were plotted to illustrate variations of sensitivity versus (! " #$%&'('&')*) for various thresholds of a classifier[8, 9] averaged over 8 scans. Sensitivity and specificity are calculated as:

$$+,-./0/1/02 \quad 3 \frac{45}{45\ 6\ 78} \quad 9 +:;;/</;/02 \quad 3 \frac{48}{48\ 6\ 75}$$

where, =>, =? , @> and @? represent the numbers of true positives, true negatives, false positives and false negatives, respectively. The sensitivity and the specificity measures of GM segmentation are computed for each ROI, for various PVE thresholds of GM/WM and various thresholds of GM/CSF boundaries. Thresholds that maximize the Youden's index (#$%&'('&')* A #%BC')'D')* " !) represent the optimum parameters of the classifier[10], and are used for regional optimization of the PVEs.

Results

Rater-Rater and Rater-CIVET similarity:

The Dice similarity coefficients of the comparison between 2 raters and the automated classification versus each rater show high similarity (>88%) between raters (Figure 2A).

Sensitivity and Specificity:

The ROC curves are sketched for 3 ROIs in Figure 2B. Results indicate that optimal classification depends on the choice of different PVE cut-off thresholds for different regions of the brain. These thresholds are selected as the values that maximize the Youden's index for the ROC curves of each region.

Regional calibration of PVE thresholds:

Regional optimization of PVE improves the performance of tissue segmentation over 16 ROIs (Table 1). Compared to automatic classification (PVE 0.5, applied globally), lowering the GM/WM thresholds (0.19-0.34) and increasing the GM/CSF threshold (0.61-0.64) achieved higher efficiency for the automatic classifier.

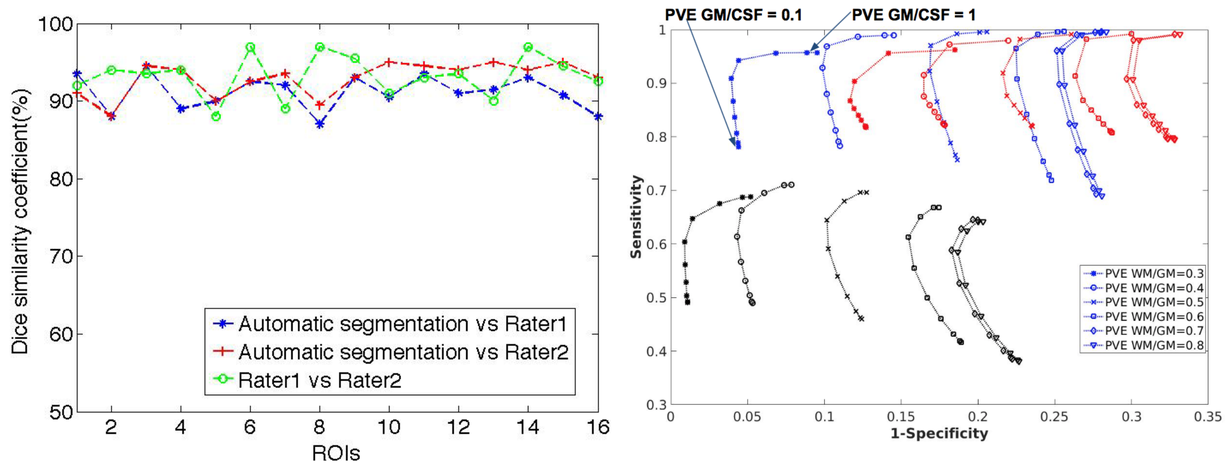

Figure 2. (A) The comparison between manual segmentations and the pipeline shows high similarity between raters and the automated method. (B) ROC curves for different ROIs (three different colors) depict regional variations in PVE-related optimization of the classifier.

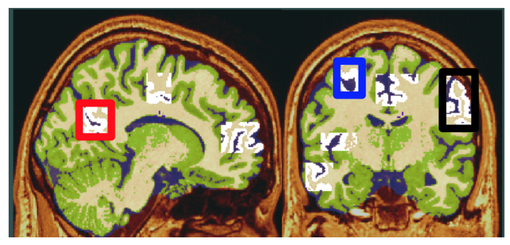

Table 1. GM segmentation performance for global and local PVE thresholding for 16 ROIs (meanE standard deviation).

	Sensitivity (%)	Specificity (%)	Accuracy (%)
Global PVE threshold (0.5)	87.98E 10.01	80.74E 10.08	83.39E 7.31
Local PVE threshold (Rater 1) (GM/CSF boundary: 0.64E 0.08, GM/WM boundary: 0.34E 0.07)	90.67E 8.34	93.47E 2.62	91.58E 4.52
Local PVE threshold (Rater 2) (GM/CSF boundary: 0.61E 0.09 GM/WM boundary: 0.19E 0.09)	92.84E 9.01	92.68E 4.72	93.14E 5.41

Conclusions

This experiment illuminates a number of issues that are critical in the assessment of segmentation accuracy, both between-rater as well as for use in the validation of an automated method. First, and well-known, there is intra-rater variability in manual segmentation. Also, there are regional variations in these findings. Differences in average estimates of local GM/WM PVE based on different raters' gold standard illustrates the challenge of subjective manual segmentation. Variations in regional PVE may be related to the imaging resolution, processing errors or even related to cytoarchitectural features of the GM boundaries. In order to better understand what drives these differences, future work will expand on these experiments, including the use of the BigBrain[11] atlas (https://bigbrain.loris.ca, http://mcin.ca) to examine the anatomical features that give rise to regional PVE variations.

References

1. Tohka, J. et al, (2014), World Journal of Radiology, 6(11):855-864. **2.** Zijdenbos, A. et al, (1998), Medical Image Computing and Computer-Assisted Interventation, W.M. Wells, A. Colchester, and S. Delp, eds. (Cambridge, MA, Springer-Verlag Berlin Heidelberg), pp. 439–448. **3.** Tohka, J. et al, (2004), NeuroImage, 23(1), pp. 84–97. **4.** Vincent, R.D. et al (2016), Frontiers Neuroinform, 10:35. **5.** Collins, D.L. et al, (1994), Journal of Computer Assisted Tomography, 18(2) pp. 192-205. **6.** Sled, J.G. et al, (1998), IEEE Transactions on Medical Imaging, vol. 17, n. 1, pp. 87-97. **7.** Ad-Dab'bagh, Y. et al, (2006), Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping, M. Corbetta, ed. (Florence, Italy, NeuroImage). **8.** Powers, D.M.S. et al, (2011), Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37-63. **9.** Prabha, D.S. et al, (2016), Indian Journal of Science and Technology, vol. 9(8):1-8. **10.** Youden, W. J. (1950), Cancer, 3: 32–35. **11.** Amunts, K., et al, (2013), Science, 340(6139):1472-1475.