Automatic Glioma Volume and Bi-directional Measurement in MRI Using Deep Learning

Ken Chang, MSE, Athinoula A. Martinos Center for Biomedical Imaging; Andrew L. Beers; Harrison X. Bai, MD; James M. Brown, PhD; K. Ina Ly, MD; Xuejun Li, MD; Joeky T. Senders; Vasileios K. Kavouridis, MD; Alessandro Boaro, MD; Chang Su; Wenya Linda Bi, MD, PhD; Otto Rapalino, MD; Weihua Liao, MD; Qin Shen, MD; Hao Zhou, MD; Bo Xiao, MD, PhD; Yinyan Wang, MD; Paul J. Zhang, MD; Marco C. Pinho, MD; Patrick Y. Wen, MD; Tracy T. Batchelor, MD; Jerrold L. Boxerman, MD, PhD; Omar Arnaout, MD; Bruce R. Rosen, MD, PhD; Elizabeth R. Gerstner, MD; Li Yang, MD, PhD; Raymond Y. Huang, MD, PhD; Jayashree Kalpathy-Cramer, PhD

Introduction

Gliomas are primary central nervous system tumors with variable natural histories and prognoses depending on their histologic and molecular characteristics. The current gold standard to determine treatment response and assess tumor progression in clinical trials are the Response Assessment in Neuro-Oncology (RANO) criteria. However, manual delineation of tumor boundaries can be difficult due to the infiltrative nature of gliomas and presence of heterogeneous contrast enhancement, which is particularly common during anti-angiogenic treatment. As a result, there can be substantial inter-rater variability in 2D measurements for both contrast-enhancing and FLAIR-hyperintense tumors. Consequently, there is great interest in developing reproducible automated methods for segmentation and calculation of the product of maximum bi-dimensional diameters.

Hypothesis

We hypothesize that deep learning can be used to automatically segments abnormal FLAIR hyperintensity and contrast-enhancing tumor, quantitating tumor volumes as well as the product of maximum bi-dimensional diameters according to RANO criteria.

Methods

Two cohorts of patients were used for this study. One consisted of 843 pre-operative MRIs from 843 patients with low- or high-grade gliomas from four institutions and the second consisted 713 longitudinal, post-operative MRI visits from 54 patients with newly diagnosed glioblastomas (each with two pre-treatment “baseline” MRIs) from one institution. Imaging from each patient was annotated by a clinical expert (radiologist, neuro-oncologist, or neurosurgery resident). We implemented multiple 3D U-Nets within our DeepNeuro framework to perform brain extraction, FLAIR hyperintensity segmentation, and contrast-enhancing tumor segmentation. We then developed an AutoRANO algorithm to automatically derive RANO measurements from our automatic deep-learning based contrast-enhancing tumor segmentations. This algorithm was designed to search for the longest diameter and corresponding perpendicular diameter within the axial slice with the largest enhancing tumor area.

Results

Our brain extraction approach had a mean Dice score of 0.935 in the testing set compared with experts, outperforming other commonly used brain extraction software. The automatically generated FLAIR hyperintensity volume, contrast-enhancing tumor volume, and AutoRANO were highly repeatable for the double-baseline visits, with an intraclass correlation coefficient (ICC) of 0.986, 0.991, and 0.977, respectively on the cohort of post-operative GBM patients. Furthermore, there was high agreement between manually and automatically measured tumor volumes, with ICC values of 0.915, 0.924, and 0.965 for pre-operative FLAIR hyperintensity, post-operative FLAIR hyperintensity, and post-operative contrast-enhancing tumor volumes, respectively. Lastly, the ICC for comparing manually and automatically derived longitudinal changes in tumor burden was 0.917, 0.966, and 0.850 for FLAIR hyperintensity volume, contrast-enhancing tumor volume, and RANO measures, respectively.

Conclusion

Our automated algorithm demonstrates potential utility for evaluating tumor burden in complex post-treatment settings.
Statement of Impact
This tool may be helpful in clinical trials and clinical practice for expediting measurement of tumor burden in the evaluation of treatment response, decreasing the time expended by clinicians for manual tumor segmentation, and decreasing inter-observer variability.

Keywords
deep learning, glioma, volumes, response assessment, RANO

Figure 1. Example of the performance different brain extraction methods on a test patient visit from the post-operative patient cohort. This patient visit had a fluid-filled resection cavity, a clear surgical point of entry, and enhancing adhesions. In this scenario, the U-Net performed better than the other brain extraction methods at removing the resection cavity while not removing brain tissue.

Figure 2. (A) Example of manual vs automatic FLAIR hypertensity segmentation (A) and enhancing-tumor segmentation (B) for the testing set in the post-operative patient cohort. (C) Examples of AutoRANO applied to automatic enhancing segmentations on the post-operative patient cohort.
Figure 3. There was high agreement between manually and automatically derived longitudinal changes in volume and RANO measures. Agreement between automatic and manual delta measures for (A) FLAIR hypertensity volume, (B) contrast-enhancing tumor volumes, and (C) RANO measure in the post-operative patient cohort. Training and Testing Sets are shown light blue/red and dark blue/red, respectively. Line of identity (x = y) is shown in all plots.