

ID #995 Automated Segmentation Performance and Uncertainty in Pediatric Diffuse Midline Gliomas using Imaging Biomarkers

Daria Laslo, Nina Baumgartner, Laura Fontanesi, Dror Suhami, Nabaan Mir, Atlas Haddadi Avval, Deep Gandhi, Ariana Familiar, Anahita Kazerooni, Zhifan Jiang, Abhijeet Parida, Marius Linguraru, Benjamin Kann, Sabine Mueller, Arzu Coeltekin, Catherine Jutzeler, Andreas Rauschecker, Sarah Brüningk

¹⁰Department of Neurology, University of California, San Francisco, San Francisco, CA, USA, San Francisco, California, United States; ¹¹Biomedical Data Science Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland; ¹²University of Bern, Department of Digital Medicine, Bern, Bern, Switzerland; ¹³Center for AI in Radiation Oncology, Department of Radiation Oncology, Inselspital, University Hospital Bern, University of Bern, Bern, Switzerland

Background: MRI-based tumor segmentation could greatly support clinical assessment of diffuse midline glioma (DMG), yet translation of automated methods remains constrained by occasional model failures, as the performance required for clinical utility and the value of uncertainty estimates in detecting meaningful errors remain unclear. We systematically evaluate segmentation performance prediction, response label stability, and uncertainty estimation.

Methods: Whole tumor was segmented in a multicentric, international cohort of pre- and post-therapy multi-contrast MRIs (n = 403) of 107 DMG patients. Segmentations by a state-of-the-art deep learning model were dichotomized by Dice score into acceptable (Dice > 0.8) and poor (Dice < 0.8). We analyzed segmentation performance classification from image-derived features (imaging metadata, radiomic features, 3D brain MRI foundation model embeddings), and response assessments stemming from manual vs. automated segmentations (n = 51 patients with longitudinal follow-up). Using eyetracking, in a sub-study, we further quantified human segmentor (36 annotators) contour uncertainty (12 slices) contextualized with observer gaze patterns.

Results: Despite generally good performance (median Dice = 0.77–0.81), auto-segmented volumes altered 20% of trajectory-based manual response labels (n = 10), predominantly misclassifying stable/progressive disease as partial response due to undersegmentation of post-treatment scans. Segmentation performance was best classified using a combination of whole image foundation model embeddings and segmented tumor volume (ROCAUC = 0.81 ± 0.05). Segmentation error correlated (|r| = 0.9) with human contour uncertainty, supporting model-based uncertainty as a proxy for annotation difficulty. Image-derived attention features from deeper encoder layers explained substantially more uncertainty variance than eye-tracking features alone (R²: 24% vs. 2%). Human gaze attention overlapped most with U-Net bottleneck activations (Dice = 0.6). A combined model integrating model attention and human visual behavior explained 39% of uncertainty variance.

Conclusions: Jointly, these results support the integration of performance- and uncertainty-aware segmentation frameworks to enable safe clinical deployment, scalable quality assurance, and reliable endpoint extraction from automated tumor segmentations in DMG.

ABSTRACT CITATION ID: WUAG026.436

ID #995 AUTOMATED SEGMENTATION PERFORMANCE AND UNCERTAINTY IN PEDIATRIC DIFFUSE MIDLINE GLIOMAS USING IMAGING BIOMARKERS

Daria Laslo¹, Nina Baumgartner^{2,3}, Laura Fontanesi⁴, Dror Suhami⁵, Nabaan Mir⁵, Atlas Haddadi Avval⁵, Deep Gandhi⁶, Ariana Familiar⁶, Anahita Kazerooni⁶, Zhifan Jiang⁷, Abhijeet Parida⁷, Marius Linguraru⁸, Benjamin Kann⁹, Sabine Mueller¹⁰, Arzu Coeltekin⁴, Catherine Jutzeler¹¹, Andreas Rauschecker⁵, Sarah Brünink^{12,13}, ¹ETH Z, ETH Z, Zurich, ZH, Switzerland; ²University of Bern, Institute of Tissue Medicine and Pathology, Bern, Switzerland; ³Department of Digital Medicine, University of Bern, Bern, Switzerland; ⁴Institute of Interactive Technologies IIT, School of Computer Science HIS, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Windisch, Aargau, Switzerland; ⁵University of California San Francisco (UCSF), Center for Intelligent Imaging (ci2), Department of Radiology & Biomedical Imaging, California, United States; ⁶Children's Hospital of Philadelphia, Center for Data-Driven Discovery in Biomedicine (D3b), Pennsylvania, United States; ⁷Children's National Hospital, Sheikh Zayed Institute for Pediatric Surgical Innovation, Washington, District of Columbia, United States; ⁸Children's National Hospital, Washington, Sheikh Zayed Institute for Pediatric Surgical Innovation, District of Columbia, United States; ⁹Dana-Farber Cancer Institute/Brigham and Women's Hospital, Harvard Medical School, Department of Radiation Oncology, Massachusetts, United States;