University Hospitals Seidman Cancer Center

CASE WESTERN RESERVE UNIVERSITY

A Deep Learning Framework for Institution-Independent PTV Delineation In Glioblastoma Using Heterogeneous MRI Datasets

Kyle O'Carroll, MS[1], Mr. Alex T. Price, PhD[1,2], Haley K Perlow, MD[1,2], Atallah Baydoun, MD[1,2], Prashant Vempati, MD[1] and Theodore Higgins Arsenault[1,2], (1)University Hospitals Seidman Cancer Center, Cleveland, OH, (2)University Hospitals Seidman Cancer Center/Case Western Reserve University, Cleveland, OH

ABSTRACT

Purpose: To develop a robust, institution-independent quality assurance model using heterogeneous datasets. By prioritizing generalizability over site-specific tuning, this model aims to predict and assist in the delineation of PTV contours for glioblastoma (GBM) using MRI T1-CE and T2-FLAIR imaging.

Methods: The Brain tumor segmentation (BRaTS) 2021 challenge made public a set of 1251 anonymized glioblastoma MRI data including T1CE and Flair MRI images which include segmentations of Edema (Ed), Necrotic/Non-enhancing tumor Core (NET) and GD- enhancing tumor (GET). Using this dataset, a MONAI based dynUnet framework was used to create a deep learning algorithm. A training, testing and validation split of (70%, 15%,15%) and a combined Dice and focal loss-based metric were utilized to train the model. The BraTS model was validated using the Burdenko GBM Progression dataset. Utilizing T1CE and FLAIR sequences, the model again predicted Ed, NET, and GET segments, which were combined into a single 'predicted Whole Tumor' (pWT) volume. Due to limitations of Burdenko's dataset only PTV information was available and thus a 15mm expansion was estimated and applied to our pWT volume based on ESTRO contouring guidelines creating our evaluation volume termed PTV_QA.

Results: Tumor delineation achieved a median/mean/IQR as follows: Overall Dice {.83, .79, .89 - .71), Dice excluding NET: {.85, .81, .90 - .76}, 95% Housdorff Distance(mm) {2.4, 3.9, 5.14 – 1.48 } and Average symmetric distance (mm) {.54, .92, 1.19 - .38}. In terms of PTV agreement with the PTV_QA the average inclusion ratio was {94.3%}, with the Median/Mean Dice {.79, .75}

Conclusion: These results demonstrate the feasibility of a novel validation metric to help mitigate institutional biases and variations. By using multi-institutional datasets, we were able to create a model which is nationally cross validated. These results support the planned implementation of local data into the model for in-house use.

CONTACT

Kyle O'Carroll
University Hospitals
Kyle.ocarroll@uhhospitals.org
347-476-3680

INTRODUCTION

Inter- and intra-institutional variability in target volume contouring poses a potential of risk of local treatment failure or healthy brain tissue damage. An automated contouring system that adapts to varied clinical practices can serve as an objective quality assurance metric to standardize contours and mitigate institutional differences. This project validates a deep-learning framework designed to automatically generate and evaluate contours, offering a robust "second check" for institutions lacking resources like dedicated chart rounds or multidisciplinary tumor boards.

METHODS AND MATERIALS

Using the BraTS 2021 challenge dataset which consisted of 1,251 glioblastoma MRIs, our team trained a MONAI-based dynUNet deep learning model to segment edema, necrotic/non-enhancing tumor core (NET), and GD-enhancing tumor (GET) from T1ce and T2 MRI sequences. The dataset was split 70/15/15% for training, testing, and validation, utilizing a combined Dice and focal loss function. The model trained for 84 epochs on 876 training patients. To address high training variability caused by tumor subregion volume imbalances averaging: 59.83 cm^3 for edema, 21.56 cm^3 for enhanced tumor, and 14.50 cm^3 for NET we implemented frequency-weighted label cropping. Two testing patients were excluded due to insufficient contours.

For external validation, we utilized the Burdenko GBM Progression dataset. In ordered to be considered evaluable we required a T1ce & FLAIR MRI image, a clinical PTV, and whole-brain contours. The model predicted edema, NET, and GET segments, which were merged into a single predicted Whole Tumor (pWT) volume for the 29 evaluable Because the original clinical PTV expansion margins were unknown in the Burdenko dataset, we generated an evaluation volume termed "PTV_QA." Following ESTRO guidelines, the pWT was expanded isotropically by 15 mm and brain-masked. The resulting PTV_QA was compared to the clinical PTV to calculate Dice scores and volume inclusion percentages.

DISCUSSION

Our results suggest that using a Monai based dynUNET machine learning algorithm has the potential to be a quality second check metric. The model demonstrated high performance on the internal test set and achieved significant overlap during external validation on the Burdenko dataset, showing strong agreement with clinical target volumes despite uncertainties in the PTV Expansions. This study was primarily limited by computational resources, which capped the training length at 84 epochs. Additionally, the lack of raw GTV contours in the Burdenko dataset introduced some ambiguity when comparing our standardized 15 mm expansion to the unknown, clinical PTV margin. Future work will focus on validating this model against datasets with well-defined GTVs to enable direct contour comparisons. Ultimately our goal is integrating the framework with local institutional data for real-time, in-house QA.

RESULTS

The initial model achieved high performance on the BraTS test set (Figures 1, Table 1) with an overall median Dice score of 0.83 (mean: 0.79, IQR: 0.71–0.89), a median 95% Hausdorff Distance of 2.4 mm (mean: 3.9 mm), and a median average symmetric distance of 0.54 mm (mean: 0.92 mm). On the external Burdenko dataset (29 evaluable patients), the initial inclusion ratio showed that 94.1% of the predicted whole tumor (pWT) fell within the clinical PTV. To compare the model's output with the clinical planning volumes, we created the PTV_QA contour (using a 15-mm isotropic expansion and brain masking per ESTRO guidelines; Figure 2). Comparing PTV_QA to the clinical PTV yielded a median Dice score of 0.79 (mean: 0.75) and a median inclusion percentage of 99.93% (mean: 94.3%).

Table 1: Metrics for the Predictions for GED, NET & GT during the testing phase with the Initial BRaTS test Dataset

Metric	Median	Average	IQR
Mean Dice	0.8345	0.7870	0.1773
Dice (excl. NCR/NET)	0.8517	0.8148	0.1407
Mean HD (95%)	2.3570	3.9049	3.6589
Mean ASD	0.5396	0.9239	0.8185

Figure 1: A Sample segmentation comparing the input MRI images (FLAIR & T1CE), the ground truth and predicted test segmentation during the testing phase of the initial BRaTs dataset.

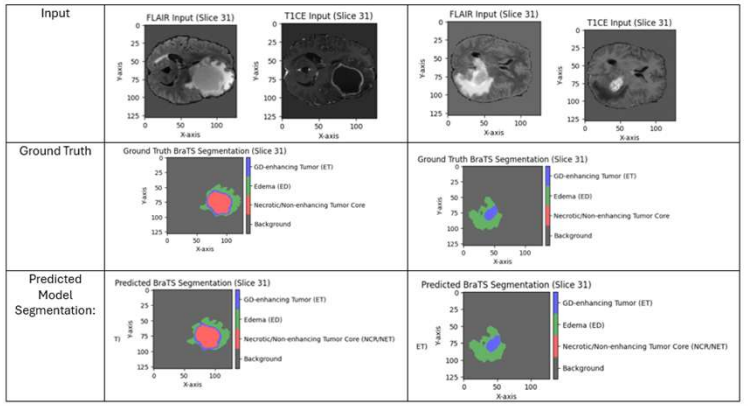
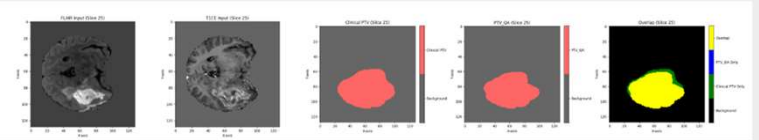


Figure 2: A Sample segmentation comparing the input MRI images (FLAIR & T1CE), the clinical PTV and predicted test segmentation, A visual representation of the overlap between the clinical PTV and PTV_QA.



CONCLUSION

This study demonstrates the feasibility of utilizing a deep learning framework to establish a robust, institution-independent quality assurance metric for glioblastoma contouring. By successfully generalizing across multi-institutional datasets without site-specific tuning, the model proves capable of standardizing target volumes and identifying geometric deviations. These findings support the planned implementation of local clinical data to refine the model for in-house use, ultimately providing an automated "second check" to mitigate institutional contouring variations and reduce the risk of geographic treatment misses.

REFERENCES

1. Zolotova, S. V., Golanov, A. V., Pronin, I. N., Dalechina, A. V., Nikolaeva, A. A., Belyashova, A. S., Usachev, D. Y., Kondrateva, E. A., Druzhinina, P. V., Shirokikh, B. N., Saparov, T. N., Belyaev, M. G., & Kurmukov, A. I. (2023). Burdenko's Glioblastoma Progression Dataset (Burdenko-GBM-Progression) (Version 1) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/E1QP-D183
2. Baid, U., Ghodasara, S., Mohan, S., et al. "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Tumor Core Detection." *arXiv preprint arXiv:2107.01047*, 2021.
3. Niyazi, M., Adeberg, S., Al-Omair, A., et al. "ESTRO-EANO guideline on target delineation and radiotherapy details for glioblastoma." *Radiotherapy and Oncology*, Vol. 184, 2023, 109663.
4. "Burdenko's Glioblastoma Progression Dataset (Burdenko-GBM-Progression)." The Cancer Imaging Archive (TCIA), 2021. DOI: https://doi.org/10.7937/E1QP-D183