Minimal Studies Needed to Train Novel PACS-Based Glioma Auto-Segmentation Tool to Translate to Clinical Practice

Sara Merkaj, Research Fellow, Yale School of Medicine

Introduction

Gliomas, the most common primary brain malignancy, are classified according to World Health Organization (WHO) criteria into grades 1-4. A rate limiting step in development of glioma grade predictive algorithms is generation of ground truth tumor segmentations, a prerequisite for volumetric information and radiomic features extraction. There is a critical need to develop automatic segmentation algorithms that are accurate and reproducible in clinical practice.

Hypothesis

Development of a PACS-based pipeline for deep learning-based auto-segmentation of gliomas that allows generation of annotated images during clinical workflow is feasible and requires a minimum training set to translate to new clinical practice.

Methods

We pretrained a deep learning algorithm (U-Net) on 1251 tumors from the multi-institutional BraTS 2021 dataset and evaluated on our internal test dataset from our home institution.

We consequently retrained the U-Net in 5 batches of 50 tumors from our home institution to auto-segment whole, core, and necrotic portions of glioma.

We established a segmentation workflow within PACS, in which de-identified images are transferred from clinical to research PACS, and are auto-segmented in 3D using the U-Net. PyRadiomics was natively embedded into PACS and once activated, feature extraction of the segmentations in all sequences was done automatically. The accuracy of the AI algorithm was measured based on Dice Scores (DSC) between the automatic segmentations and the manually modified segmentations.
Results

The baseline U-Net algorithm was trained on 1251 tumors from BraTS 2021 dataset and evaluated on our internal test dataset of 35 gliomas. The internal dataset used for the gradual retraining of U-Net consisted of 250 gliomas divided into 5 batches of 50 tumors.

The baseline U-Net trained on BraTS 2021 achieved Dice Scores of 0.82. Dice Scores gradually improved from 0.82 to 0.84 through the incremental training (Figure 1).

Figure 2 demonstrates an example of a segmentation of a brain tumor generated before our incremental retraining (BrATS data only) and after our incremental retraining process (BraTS and hospital data). We noticed that signal intensity contrasts were less pronounced on segmentations generated after our retraining process, resulting in ground base segmentations that are closer to what a neuroradiologist would segment in daily clinical practice (Figure 2).

Conclusion

We demonstrated that clinical implementation of segmentation algorithms in neuroradiology practice is feasible through a PACS-integrated workflow. This workflow allows real-time generation of large and annotated datasets of brain tumors and translation of algorithms into new clinical environments with minimal required annotated segmentations.

Figures

Figure 1. Dice Scores (DSC) measured between the automatic segmentations and the manually modified segmentations for each step of the incremental training process. 0 = BrATS pretrained algorithm (baseline). 50-250 = batches of data used for step 1-5 of the gradual training process.
Figure 2. Example of a high-grade glioma automatically segmented through using U-Net before retraining (left) and after retraining on YNHH data (right) on FLAIR MRI sequence.

Keywords
Applications; Artificial Intelligence; Clinical Workflow & Productivity; Imaging Research

SIIM22 Presentation
Neuroradiology & MSK Machine Learning Abstracts | 3015V | Virtual