A Comparison of Convolutional Neural Network Architectures in Auto-Segmenting Primary Oropharyngeal Cancers from Contrast-Enhanced CT Scans

Onur Sahin, PhD, Medical Student, McGovern Medical School; Kareem A. Wahid, PhD; Abdallah S. Mohamed, MD, PhD; Clifton D. Fuller, MD, PhD; Mohamed A. Naser, PhD

Introduction

Contouring tumors during the development of radiotherapy plans is a labor-intensive process with a high degree of variation when manually performed. Deep learning (DL) models have been developed to automate the segmentation of tumors from anatomical imaging modalities. Previously, our group developed an auto-segmentation model for oropharyngeal cancers (OPCs) from MRI scans. While multiple studies focused on auto-segmentation models from other imaging modalities, there has been no work assessing the development of models using diagnostic-quality contrast-enhanced computed tomography (CE-CT) scans for OPCs , even though CE-CTs can be more easily acquired. In this study, we develop two models using Swin and Resunet convolutional neural networks to auto-segment OPCs from CE-CT scans.

Hypothesis

We hypothesize that DL convolution neural networks can be trained on CE-CT scans to segment OPCs.

Methods

Pre-surgical diagnostic quality CE-CT scans were collected from 474 patients with OPCs. Ground-truth segmentations of OPCs were manually performed and images were randomly assigned into a training (n=380) and test set (n=94) using a 4:1 split. 5-fold cross validation was performed on the training set to train 5 separate models using either a Resunet or Swin network from the MONAI framework. Final model predictions were generated by a consensus mask using either an averaging or simultaneous truth and performance level estimation (STAPLE) method and assessed on the test set. The Dice similarity coefficient (DSC) was used to evaluate model performance. Statistical comparisons were performed using Wilcoxon signed-ranks test.

Results

In the training/test sets, 16.8%/17% of OPCs were T1, 42.1%/42.5% were T2, 22.3%/22.3% were T3, and 18.6%/18.1% were T4. The overall average DSC performance across cross-validation folds was 0.61/0.62 for the Swin/Resunet models, 0.71/0.7 consensus masks developing via an averaging method on the test set, and 0.7/0.69 for the STAPLE method. The Wilcoxon signed-ranks test between the Swin and Resunet showed a $p < 0.05$ for only the consensus masking generated via the averaging method.

Conclusion

We developed the first DL models, to our knowledge, trained on CE-CT scans to segment OPCs, and showed that

they achieved similar performance to DL models trained on MRI images. Further work can focus on improving segmentation performance by utilizing larger patient cohort sizes.

Statement of Impact

In this study we show that more CE-CT scans can be used to develop OPC auto-segmentation models with reasonable performance. With further optimization, the proposed models could be integrated into the workflow of radiation oncologists to increase the efficiency of developing radiation treatment plans.

Keywords

Auto-segmentation; Oropharyngeal Cancer; Convolutional Neural Networks