Deep learning-based automatic segmentation of rectal tumors in endoscopy images

Alana Thibodeau-Antonacci[1,2], Luca Weishaupt[3], Aurélie Garant[4], Corey Miller[5,6], Té Vuong[7], Philippe Nicolaï[2], Shirin Enger[1,3]

1. Medical Physics Unit, McGill University, Montreal, Canada
2. Centre Lasers Intenses et Applications (CELIA), Université de Bordeaux, Bordeaux, France
3. Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Canada
4. Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, USA
5. Department of Medicine, McGill University, Montreal, Canada
6. Division of Gastroenterology, Jewish General Hospital, Montreal, Canada
7. Department of Oncology, Jewish General Hospital, Montreal, Canada

Purpose: Radiotherapy is commonly used to treat rectal cancer. Accurate tumor delineation is essential to deliver precise radiation treatments. Endoscopy plays an important role in the identification of rectal lesions, but is prone to errors as tumors are often difficult to detect. Previous deep learning methods to automatically segment malignancies in endoscopy images have used single expert annotations or majority voting to create ground-truth data, but this ignores the intrinsic inter-observer variability associated with this task. The goal of this study was to develop an unbiased deep learning-based segmentation model for rectal tumors in endoscopy images.

Methods: Three annotators identified tumors and treatment change (*i.e.,* radiation proctitis, ulcers and tumor bed scars) in 464 endoscopy images from 18 rectal cancer patients. In cases where the image quality was too low to confidently classify the tissues, the image was labeled as “poor quality”. The inter-observer variability was evaluated for whole image classification using the F1-score and on a contour level using the dice score in cases where two annotators agreed a class was present in the image. A deep learning model was trained with all annotators’ labels simultaneously to perform pixel-wise classification. Grouped stratified splitting was used to divide the data into the training (375 images; 1125 labels), validation (29 images; 87 labels) and test (60 images; 180 labels) sets. A dice loss function was used to train a DeepLabV3 model with a pre-trained ResNet50 backbone. Automatic hyperparameter tuning was performed with Optuna. Images were normalized according to the COCO standard and resized to 256x256 pixels. Random transformations (horizontal and vertical flipping, rotation) were applied once per epoch to the training set with a 50% probability to increase the robustness of the model.

Results: Significant variability was observed in the manual annotations: the average F1-score was 0.76 and 0.59 for tumor and treatment change, while the average dice score was 0.83 and 0.61 for the same classes. The deep learning model was able to segment each image in the test set in less than 0.05 seconds on CPU, while manual contours took on average 21 seconds per image. The model achieved an average dice score of 0.75 and 0.64 for tumor and treatment change, respectively.

Conclusion: This study illustrated the important inter-observer variability associated with the task of segmenting endoscopy images. The results suggest that ground-truth data should be composed of labels from multiple annotators to avoid bias. Additional data are needed to increase the developed model’s performance and generalizability.