Artificial Intelligence for the Segmentation of Mesothelioma: Analysis of Performance Based on Computer Probability Outputs (CALGB 30901, Alliance)

Mena Shenouda[1], Eyjólfur Guðmundsson[2], Feng Li[1], Christopher M. Straus[1], Hedy L. Kindler[1], Arkadiusz Z. Dudek[3], Thomas E. Stinchcombe[4], Xiaofei Wang[5], Adam Starkey[1], Samuel G. Armato III[1]

[1]University of Chicago, Chicago, IL, USA; [2]Icelandic Radiation Safety Office, Reykjavik, Iceland; [3]Metro Minnesota Community Oncology Research Consortium, St. Louis Park, MN, USA; [4]Duke Cancer Institute, Duke University, Durham, NC, USA; [5]Alliance Statistics and Data Management Center, Duke University, Durham, NC, USA

Objectives:

- To determine whether an optimal threshold exists when applied to probability maps output by an artificial intelligence (AI) system comprised of a convolutional neural network (CNN) trained to segment pleural mesothelioma tumors on thoracic computed tomography (CT) scans.

Methods:

- 186 baseline and follow-up CT scans from 48 patients with pleural mesothelioma were collected from the Cancer and Leukemia Group B (CALGB) 30901 study, a Randomized Study of Maintenance Pemetrexed Versus Observation for Treatment of Malignant Pleural Mesothelioma.
- The CNN generated for each CT section a probability map, which displays the likelihood that each pixel contains tumor.
- Tumor boundary contours were generated from the CNN probability maps based on various probability thresholds.
- A radiologist modified the contours generated at a (default) 0.5 probability threshold, and the reference standard provided by the radiologist and the CNN outputs were compared through (1) percent difference of tumor volume and (2) overlap of the contours using the Dice Similarity Coefficient (DSC).
- These two figures of merit were calculated for thresholds ranging between 0.001 to 0.9, investigating whether an "optimal" threshold can minimize the percent difference and maximize the DSC .

Results:

- Tumor volumes calculated using the CNN contours generated at a 0.5 probability threshold were consistently smaller than those calculated from the radiologist's contours.
- Absolute percent difference decreased on average, from 42.9% to 22.1%, when decreasing the probability threshold from 0.5 to 0.1. Mean DSC values across thresholds were within the range 0.56-0.59 (Table 1).
- When inspecting the average of the two metrics across all cases, the lowest absolute percent difference occurs at a probability threshold of 0.06, while the maximum DSC occurs at a 0.2 threshold (Figure 1).
- Because the acquired CT scans incorporated patients with varying disease morphology, visual inspection of the CNN contours showed its deficiency with certain presentations, such as severe pleural effusion or disease in the pleural fissure.

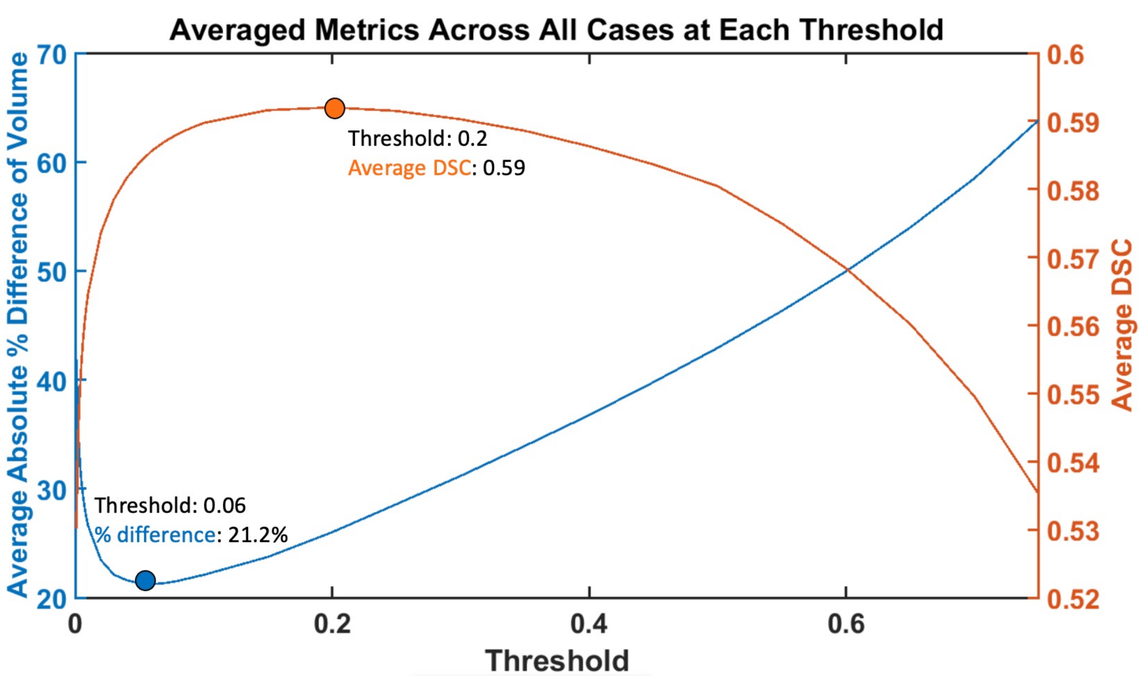

Figure 1: The average absolute percent difference of volume and the average DSC across all cases for the entire threshold range. The minimum and maximum value of each respective metric is highlighted.

Threshold	Absolute percent difference of volume (± SD)	DSC (± SD)
0.5	42.93 ± 32.99	0.58 ± 0.17
0.4	36.75 ± 30.88	0.59 ± 0.17
0.3	31.18 ± 28.17	0.59 ± 0.16
0.2	26.01 ± 24.67	0.59 ± 0.16
0.1	22.09 ± 19.11	0.59 ± 0.15
0.01	26.60 ± 17.17	0.56 ± 0.14

Table 1: The average absolute percent difference of volume and the average DSC across all cases for six thresholds. The volume and DSC values were all statistically different than the volume and DSC values acquired at the 0.5 threshold.

Conclusion:

- This investigation explored the relative values of the probability maps generated by a CNN and their impact on the clinical task of mesothelioma tumor volumetrics.
- No single CNN probability map threshold was found that optimized for both tumor volume and DSC.
- This work highlights the importance of considering the difference of tumor volume and overlap of tumor regions between the CNN and the reference standard when optimizing CNN-based tumor segmentations.