AI-based Automated Lipomatous Tumor Segmentation in MR Images

[1]Quinn Kwan-Tai Ng, [2,9]Chih-Chieh Liu, [1]Yasser G Abdelhafez, [1]S. Paran Yap, [3]Francesco Acquafredda, [2]Silvia Schirò, [1]Andrew L. Wong, [1]Dani Sarohia, [1]Cyrus Bateni, [4]Morgan A. Darrow, [5]Michele Guindani, [5]Sonia Lee, [6]Michelle Zhang, [7]Ahmed W. Moawad, [7]Layla Shere, [7]Khaled M. Elsayes, [8]Roberto Maroldi, [2]Thomas M. Link, Jinyi Qi, [1]Lorenzo Nardo

[1]Department of Radiology, UC Davis, [2]Department of Radiology and Biomedical Imaging, UC San Francisco, [3]Radiologia Pediatrica - Ospedale dei Bambini, ASST Spedali Civili, Brescia, Italy, [4]Department of Pathology and Laboratory Medicine, UC Davis, [5]Department of Statistics, UC Irvine, [6]Department of Diagnostic Radiology, McGill University Health Center, Montreal, Canada, [7]Department of Diagnostic Imaging, University of Texas MD Anderson Cancer Center, Houston, TX, USA, [8]Scienze Radiologiche, Università degli Studi di Brescia, Brescia, Italy, [9]Canon Medical Research USA, Inc.

Diagnostic aids from quantitative radiomics have shown promising results for the characterization of many diseases. However, the extraction of radiomic features for complex structures/diseases such as lipomatous tumors from MRI images remains challenging. This is mainly due to the labor intensive, and error prone identification and segmentation of these tumors needed for accurate analysis. Deep learning (DL) has been proposed for the automation of medical image segmentation to improve accuracy, consistency, and efficiency; however, a major challenge of this task is data heterogeneity from multicenter data for pre-processing algorithms. We aimed to develop and study a DL-based Super Learner (SL) ensemble framework approach in comparison to any single trained neural network.

Pathologically proven LTs on pre-operative T1-weighted/proton-density MR images of 185 patients (4 centers, 20 MR scanners; 102 females; age 58.4±11.9 yrs) were manually segmented and categorized by tumor locations as distal upper limb (DUL), distal lower limb (DLL), proximal upper limb (PUL), proximal lower limb (PLL) or Trunk (T) and grouped by 80%/9%/11% for training, validation, and testing. Six configurations of correction/normalization were applied to data for 5-fold-cross-validation trainings, resulting in 30 base learners (BLs). A SL was obtained from the BLs by optimizing SL weights and performance was evaluated by dice-similarity-coefficient (DSC), sensitivity, specificity and Hausdorff distance (HD95). For predictions of the BLs, the average DSC, sensitivity and specificity from the testing data were 0.72±0.16, 0.73±0.168 and 0.99±0.012, respectively, while for SL predictions were 0.80±0.184, 0.78±0.193 and 1.00±0.010. Average HD95 of the BLs were 11.5 (DUL), 23.2 (DLL), 25.9 (PUL), 34.7 (PLL) and 49.4 (T) mm, whereas of SL were 1.7, 8.4, 15.9, 2.1 and 37.9 mm, respectively.

The proposed method could improve the segmentation accuracy and mitigate the performance instability and data heterogeneity aiding the differential diagnosis of LTs in real clinical situations.

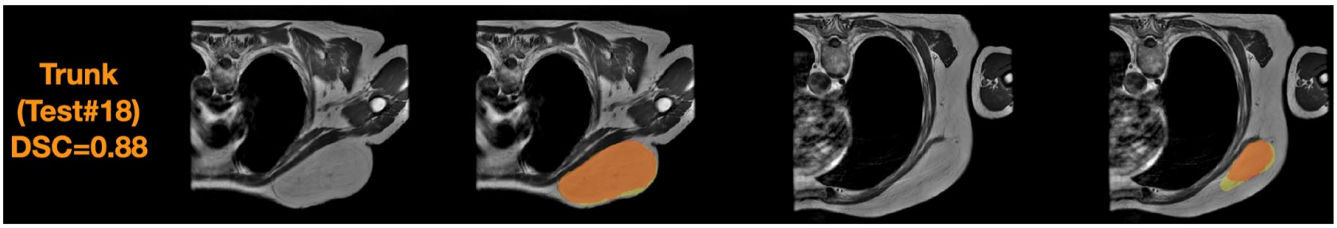

Figure 1: Example SL ensemble predictions (red) and tumor delineations (yellow) per tumor location overlaid on the T1-weighted MR images giving orange color for the correctly predicted tumor region. First two columns represent the central image slices of the tumors without and with the delineations and predictions, respectively, whereas the second two columns represent the edge image slices of the tumors. The predictions are often worse in the edge slices than in the central slices of tumor.