Multi-center validation of a model for prostate tumor delineation using multi-parametric MRI

C. Dinh1, K. Haustermans1, P. Steenbergen1, G. Ghobadi1, E. Lerut1, R. Oyen2, H.V.D. Poel1, J.D. Jong3, S. Heijmink4, U.V.D. Heide1

1The Netherlands Cancer Institute, Department of Radiation Oncology, Amsterdam, The Netherlands
2University Hospitals Leuven, Department of Oncology KU Leuven, Leuven, Belgium
3University Hospitals Leuven, Department of Imaging & Pathology, Leuven, Belgium
4The Netherlands Cancer Institute, Department of Radiology, Amsterdam, The Netherlands
5The Netherlands Cancer Institute, Department of Pathology, Amsterdam, The Netherlands

Purpose/Objective: Focal boosting in radiotherapy for prostate cancer aims at improving tumor control without increasing side effects by giving a high dose to the index lesion only. Here, an accurate delineation of the tumor is essential. Manual delineations are time consuming and vary among observers. Recently a model has been published for automatically delineating tumors in the peripheral zone using multi-parametric MRI (Model 1). The model that was trained on a database of 87 radiotherapy patients. Here we perform an independent validation of this model with and without updating its parameters using data from independent centers to investigate the robustness of this model w.r.t. different equipment and scan protocols.

Materials and Methods: The study involves 18 patients (database A) from the first center and 17 patients (database B) from the second center (Table 1). There are one and three patients from A and B, respectively, of whom index lesions are not in the peripheral zone. Prior to prostatectomy, all patients received MRI, including T2-weighted (T2w), diffusion-weighted (DWI), and dynamic contrast enhanced (DCE) sequences. The MRI protocols differ among the centers and are not intended for quantitative evaluation. Hematoxylin and Eosin stained slices were obtained from the prostate specimens, on which a pathologist delineated the tumors. These slices were registered to the T2w images and then used as ground truth for evaluation. Voxels within ±1.25 mm margin of the pathological tumor contour were ignored to reduce the influence of registration errors. Model 1 is a logistic regression model which predicts the probability of tumor presence using ADC, Ktrans, and coordinate features. We compare the performance of Model 1 and two retrained models, A and B, created by fitting the coefficients of Model 1 on databases A and B, respectively. The retrained models are tested using a leave-one-patient-out cross validation.

Results: Model 1 correctly detects 16 of 17 and 13 of 14 index lesions on databases A and B, respectively. However, it introduces 7 and 8 false alarms. The T1w images indicate that the tumor missed in database A might be obscured by post-biopsy hemorrhage effects. The tumor missed in database B is sparse on histology, and therefore difficult to detect on MRI. Figure 1 shows an example of the classification results.

On database A, models 1 and A produce average AUC values of 82.2% and 83.6%, respectively. On database B, models 1 and B produce average AUC values of 73.6% and 77.6%. That means the retrained models perform slightly better than Model 1, however, the difference in AUC is not statistically significant.

Conclusions: Model 1 provides comparable results to the retrained ones, i.e., it is robust to differences in scan protocol. This suggests that the model can be implemented widely as an additional diagnostic tool for radiation oncologists to delineate focal lesions in prostate.