ESTRO

Abstract

Title

Clinical evaluation of organs at risk automatic-segmentation for T2-weigthed MRI

Authors

Thais Roque[1], Neil Newman[2], Sotirios Stathakis[3], Daniela Thorwarth[4], Daniel Zips[4], Marcel Nachbar[5], Sanmady Kandiban[6], Ayoub Oumani[7], Kumar Shreshtha[7], Nikos Paragios[6,8], William E Jones III[3]

Authors Affiliations

[1]TheraPanacea, Clinical and Partnerships Affairs, Paris, France; [2]The University of Texas Health Science Center- San Antonio, Department of Radiation Oncology , San Antonio, USA; [3]The University of Texas Health Science Center- San Antonio, Department of Radiation Oncology, San Antonio, USA; [4]University Hospital and Medical Faculty, Eberhard Karls University Tübingen, Department of Radiation Oncology, Tübingen, Germany; [5]University Hospital and Medical Faculty, Eberhard Karls University Tübingen, Department of Radiation Oncology Tübingen, Germany, Tübingen, Germany; [6]TheraPanacea, Research and Development, Paris, France; [7]TheraPanacea, AI Research Department, Paris, France; [8]CentraleSupelec, University of Paris-Saclay, Computer Science and Applied Mathematics , Gif-sur-Yvette, France

Purpose or Objective

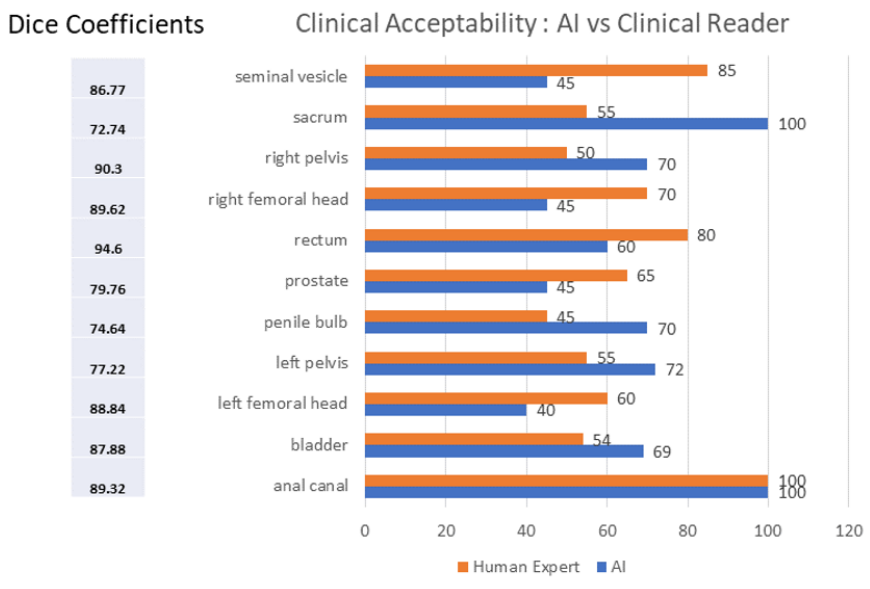
Magnetic resonance imaging (MRI) is essential for radiation therapy (RT) planning of pelvic tumors. Thanks to its excellent soft tissue contrast, MRI is systematically used to facilitate precise target and organs-at-risk (OAR) delineation. This is achieved through its superposition to the computed tomography (CT) image through deformable registration, which is sensitive to the quality of the deformable registration and fails to harness the full potential of MRI in RT. The latest advances of artificial intelligence methods offer new perspectives towards full MR-only RT (MRoRT). Automatic delineation of OARs on MR images is a prerequisite in this direction. This study aims at evaluating an AI-based auto-contouring (AC) solution and compare its clinical acceptability against contours delineated by experts (EC).

Materials and Methods

ART-Net® is a CE-marked, FDA-cleared three stage anatomically preserving deep learning ensemble architecture for AC of OARs in RT. This architecture was trained for AC of pelvic OARs for 1.5T MR Elekta Unity® T2 sequence using a range of 197 (seminal vesicle) to 295 (penile bulb) patient structures. All OARs routinely adopted for RT planning were evaluated and are listed in Tab. 1. An independent cohort of 20 additional patients was further evaluated by blending treatment experts' contours with ACs from ART-Net at 50%-50% ratio. Random blending at the patient level was performed guaranteeing that, among contours being evaluated per patient and OAR, the 50%-50% split was satisfied. Contours were scored as A/acceptable, B/ acceptable after minor corrections, and C/ not acceptable for clinical use.

Results

The mean Dice coefficient on the testing data set was 84.70% (Tab.1). Running time of ART-Net® was around 30 seconds per patient. Overall clinical acceptability after aggregating blinded evaluations coming from two independent experts for the combined categories (A+B) was 98% for ART-Net® and 95% for EC. Anal canal was the best performing structure for ART-Net® and for the EC (100% of A), whilst, left femoral head (40% of A, 60% B) and penile bulb (40% of A, 45% B) were the least performing OARs for AC and EC, respectively.



(i) *(ii)*

Table 1 (i) Dice scores for the different AC structures (ii) Percentage of structures evaluated as A (i.e. considered clinically acceptable) per organ (AI (AC) vs Clinical Reader (EC))

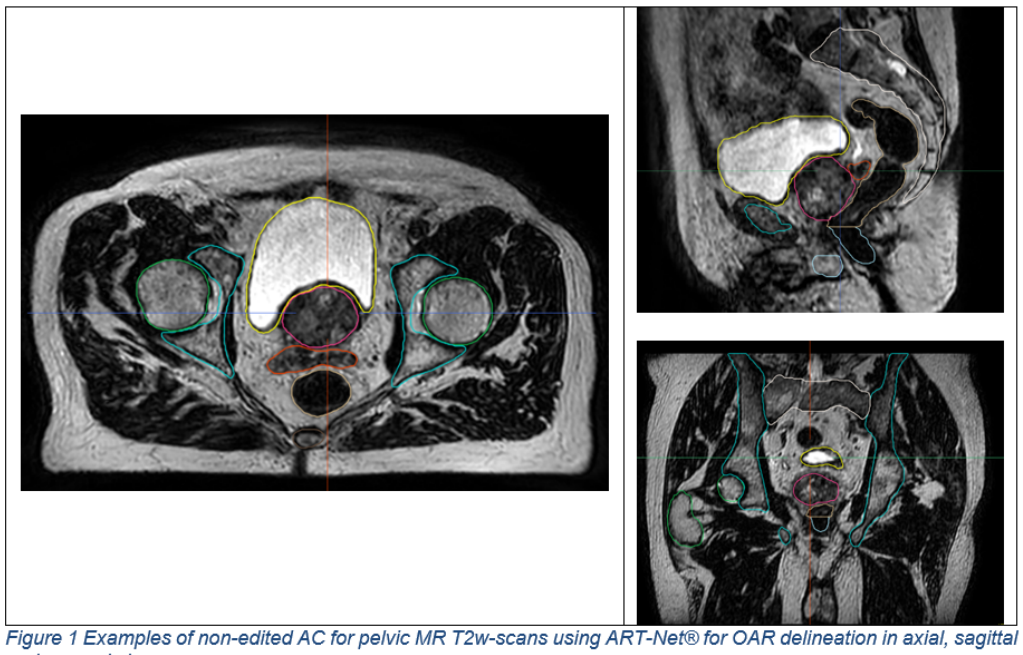


Figure 1 Examples of non-edited AC for pelvic MR T2w-scans using ART-Net® for OAR delineation in axial, sagittal and coronal views.

Conclusion

This work reports some preliminary clinical evaluation regarding the clinical relevance of AC in T2 1.5T MR Elekta Unity® T2 sequence for pelvic RT. Our results suggest that the trained ART-Net® model could be a viable clinical alternative to the human expert. Dice coefficients (84.70%) are higher than the ones usually reported in the literature and our blinded clinical evaluation yielded slightly higher acceptability for the AI contours than for EC, at a fraction of the time (around 30 seconds as compared to 15min for the expert). Introduction of additional samples at training, especially for structures with a small sample number, and of multi-vendor T2 MR cohort could be an interesting future direction to further enhance the clinical relevance of the obtained results.

Funding: The European Union’s Horizon 2020 (No. 880314).