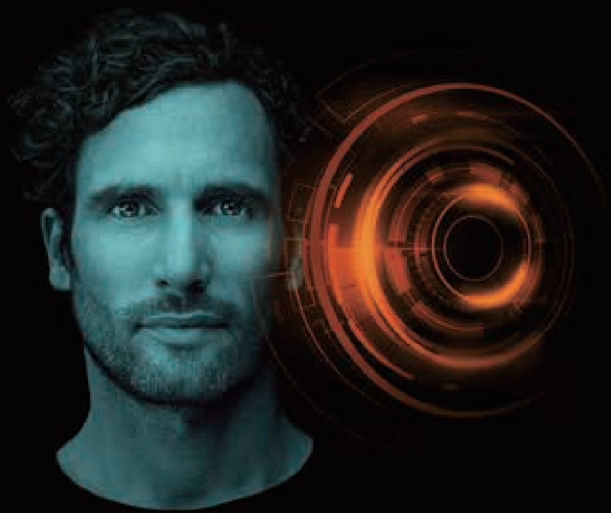
OAR Turing Test: A blinded comparison of expert-based OAR segmentation and three commercial auto contouring solutions



www.siemens-healthineers.com/jp

日本放射線腫瘍学会第34回学術大会（JASTRO2021）
スポンサードセミナー2

オンタイム配信

日　時：2021年11月12日（金）12:00～13:00

演　題：OAR Turing Test: A blinded comparison of expert-based OAR segmentation and three commercial auto contouring solutions

演　者：Dr. med. Florian Putz
Department of Radiation Oncology, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

参加登録、視聴方法は、学会ホームページをご覧ください。
学会ホームページURL：
http://www.congre.co.jp/jastro2021/outline/index.html

共催：日本放射線腫瘍学会第34回学術大会
　　　シーメンスヘルスケア株式会社

JASTRO

SIEMENS Healthineers

OAR Turing Test: A blinded comparison of expert-based OAR segmentation and three commercial auto contouring solutions

Background

While Organ at risk (OAR) autosegmentation for radiotherapy treatment planning has traditionally been atlas- or model-based, it is only recently that Deep Learning-based OAR autosegmentation has become commercially available. In this study, we compared Deep Learning-, atlas- and model-based OAR autosegmentations with expert OAR segmentations in a completely blinded fashion.

Methods

We evaluated three commercial autosegmentation solutions based on the three different autosegmentation core technologies: Deep Learning-based (Siemens Healthineers AutoContouring Solutions, Siemens Healthineers DirectORGANS and syngo.via RT Image Suite), atlas-based (Pinnacle 9.10 SPICE autosegmentation) and model-based OAR autosegmentation (Raystation 9B).
Across 5 major radiotherapy treatment sites (head and neck, thorax, breast, abdomen, and pelvis) OARs in ten clinical patient cases each were autosegmented with every solution (Total of 50 patient cases, 6-10 OARs per site). In addition, three experienced radiation oncologists manually segmented all 50 patient cases (syngo.via RT Image Suite, 3x50 manually segmented datasets) and the time required for manual delineation was taken.
OAR autosegmentations and expert OAR segmentations were then presented to the three physicians in a completely blinded manner and in randomized order using a custom-built software and the physicians were tasked with rating every OAR (auto)segmentation on a 1 - 4 scale (1 - must redo, 2 - requiring major corrections, 3 - requiring minor corrections, 4 - clinically usable). In a second step, physicians manually corrected all OAR autosegmentations to measure the manual time required to make OAR autosegmentations clinically usable. Furthermore, OAR autosegmentations were quantitatively compared to STAPLE consensus structures created from the three expert segmentations using DICE and Hausdorff metrics.

Results

In total 2140 manually and automatically created OAR segmentations were rated by the three radiation oncologists (total of 5400 ratings). There was no significant difference in the blinded ratings for Deep Learning- and physician-generated OARs in a paired comparison (mean rating, 3.6 vs. 3.5, $p = 0.054$ trend in favor of Deep Learning, OAR Turing Test), whereas atlas- and model-based OARs were rated significantly lower than physician-generated OARs (3.1 vs. 3.5 and 2.7 vs. 3.5, $p < 0.001$ each). Deep Learning-generated OARs also scored significantly higher than atlas- or model-based OARs in a direct comparison ($p < 0.001$ each).
Moreover, Deep Learning-generated OARs required the least amount of time for manual correction both at an individual OAR- and at a case-level across all 5 radiotherapy treatment sites. Finally, Deep Learning-generated OARs were highly significantly more similar to the physician consensus OARs in terms of DICE score and Hausdorff distance ($p<0.001$ each).

Conclusions

In a completely blinded evaluation, no difference in expert ratings was observed between Deep Learning- and physician-created OARs (OAR Turing Test). In multiple analyses, Deep Learning-based OAR autosegmentation showed clinically meaningful and statistically significant improvements over atlas- and model-based OAR autosegmentation.