BHPA-symposium 2022

Scientific Committee

Milan Tomsej (chair), CHU Charleroi

Alain Seret, Université de Liège

Caro Franck, Universitair Ziekenhuis Antwerpen

Damien Dumont, UCLouvain

Kristof Baete, KU Leuven

Lara Struelens, SCK-CEN

Federica Zanca, Palindromo

Koen Tournel, Jessa Ziekenhuis

Nico Buls, UZ Brussel

Gert Van Gompel, UZ Brussel

Thierry Gevaert, UZ Brussel

Plenary Keynote

Chair: Nico Buls (UZ Brussel)

Friday 29/04/2022 09h15-10h35

Auditorium 2000

Comprehensive evaluation of ProtegeAI Prostate 2.0 auto-segmentation: time-gain and accuracy

N. Jullian[1], **Z. Paquier[2,3]**, M. Burghelea[2,3], D. van Gestel[1], N. Reynaert[2,3], **A. Gulyban[2,3]**
[1]Radiation Oncology department, Jules Bordet Institute, Anderlecht, Belgium
[2]Medical Physics, Jules Bordet Institute, Anderlecht, Belgium
[3]Radiophysics and MRI physics laboratory, Faculty of Medicine, Université Libre de Bruxelles, Brussels, Belgium.

ABSTRACT

KEY WORDS – Autosegmentation, Optimization, AI-implementation, Contouring

Introduction

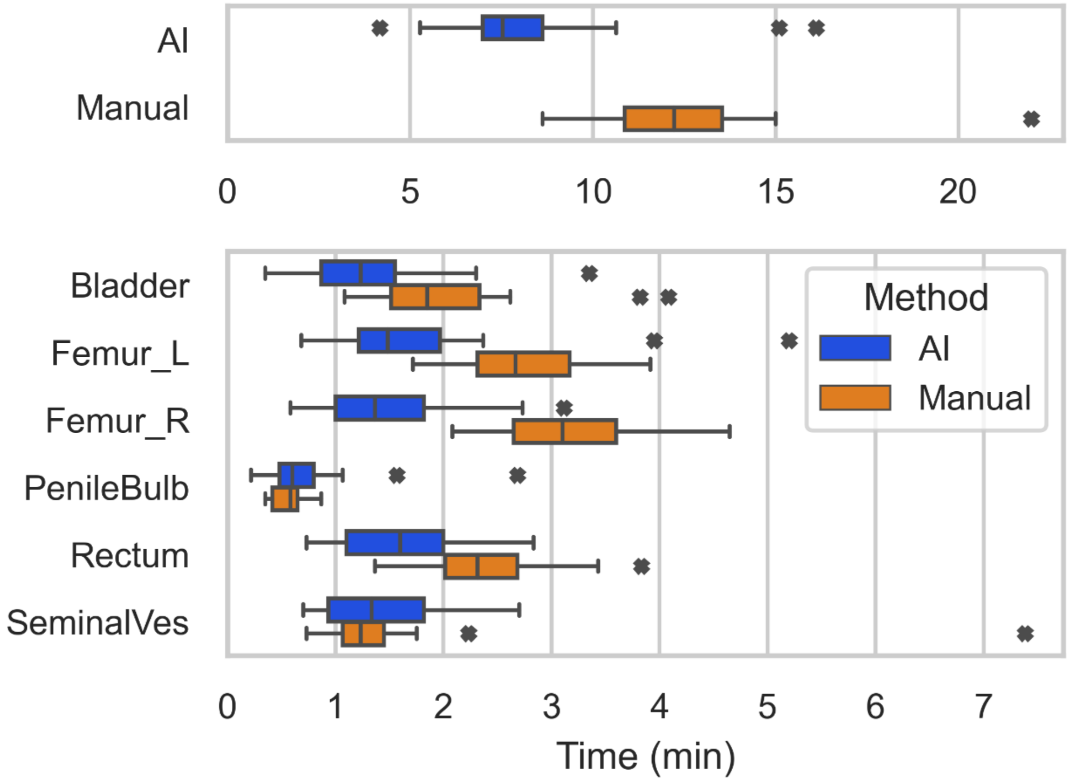
The aim of this study was to evaluate the time gain and accuracy of the MIM ProtegeAI 2.0 auto-segmentation solution (version 7.1.5, MIM software Inc, Cleveland OH, USA). A second objective was to assess intra-observer variability and familiarization bias when using auto-segmentation.

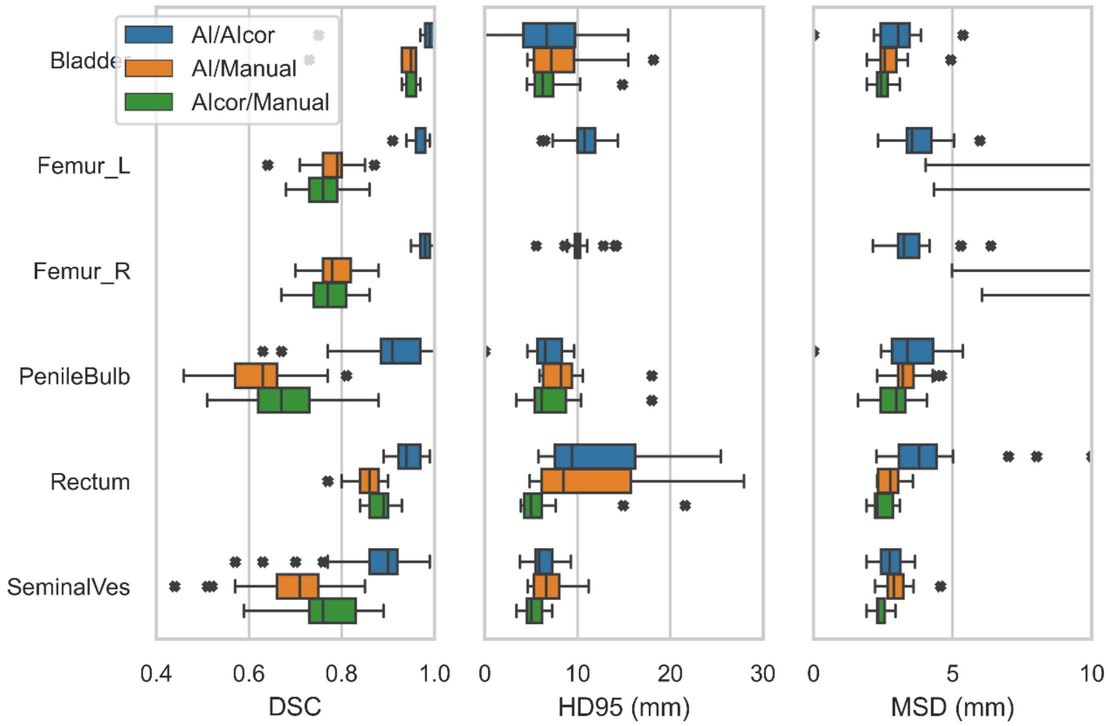
Materials and methods

Twenty-five patients with prostate cancer were included. For each case a planning CT scan (from vertebrae L1/2 to 3cm below the ischial tuberosity, 3 mm slice thickness) was performed, followed by auto-segmentation using the ProtegeAI Prostate 2.0 model (***AI***) and manual delineation by a single observer (***Manual***). Femur_L/_R, PenileBulb, Rectum, SeminalVes, Bladder were evaluated; while another five AI-generated OARs did not match our institutional template, hence were not evaluated. Time of AI delineation scoring (***AIscor***: major/minor/no correction needed), AI correction (***AIcor***), total ***AI*** (=AIscor+AIcor) and manual delineation was measured. Time gain was also calculated per individual OAR. Half of the cohort started with ***AIscor*** and ***AIcor*** followed by ***Manual***, while the other half started with ***Manual***, followed by ***AIscor*** and ***AIcor***. For both groups ***Manual*** and ***AIcor*** were compared separately to evaluate familiarization bias. For time-gain and bias evaluationt-test at $p<0.05$ significance level were used. Dice Similarity Coefficient (DSC), 95% Hausdorff and median surface distance (HD95, MSD) were also determined for ***AI/AIcor***, ***AI/Manual*** and ***AIcor/Manual*** comparisons. ***AIcor***/***Manual*** was used to define intra-observer variability as both contours were considered clinically acceptable.

Results

A total of 235 contours were generated by AI (5 min per patient). For 20 patients, AI failed to generate Kidney_L/_R. Major, minor or no correction was considered in 14%, 72% and 14% of delineations, respectively. ***Manual*** took on average 12:25 (min:sec; range:8:21-21:59), ***AIscor*** and ***AIcor*** 1:55 (r: 1:21-3:32) and 6:18 (r:2:49-14:14), respectively (figure 1). AI gave up to 13:06 time gain, with an average of 4:12 ($p<0.001$), although for two patients ***AI*** took more time than ***Manual*** (3:05 and 2:08). Per OAR, the average time gain was 0:42 (r:-0.11-1:45). The familiarization bias, observed for ***Manual*** ($p=0.029$), was on average 2:25 faster when AI workflow started first, while for ***AIcor*** no significant bias was observed ($p=0.168$). Good DSC (>0.8) was observed for ***AI/AIcor***, while HD95 and MSD (figure 2) showed larger discrepancy. For Femur (***AI*** and ***AIcor***) vs. Femural Head (***Manual***) agreement was moderate due to difference in intended delineation. Intraobserver (***AIcor***/***Manual***) variability was worse for DSC and better for HD95 and MSD compared to ***AI*** vs. ***AIcor***.





Conclusion

ProtegeAI Prostate 2.0 auto-segmentation provides on average >4 minutes gain per patient while requiring only minor corrections. Realistic time gain is likely higher, as AIscor+AIcor prior manual delineation significantly reduced manual delineation time. Intraobserver variability remains a substantial source of differences, especially based on DSC.

Fully-Integrated Auto-Contouring with Artificial intelligence (AI) in Eclipse TPS

H. Cavus[1], P. Bullens[1]

[1]Jessa Hospital, Belgium

ABSTRACT

KEY WORDS – Contouring, AI, ESAPI, SQL and Prostate

Introduction

Auto-contouring with AI is a recent solution instead of manual contouring. To use it easily in clinic, a fully integration in the TPS is necessary.

Materials and methods

In this study, we used Contour ProtégéAI™ to compare it with manual contouring. Two parameters were used: completion time and Dice Coefficient Index [2].

Contour ProtégéAI™ [1] is fully-integrated in Eclipse TPS with ESAPI and SQL.

Results

Figure 1: Box whisker plot showing the Dice coefficient index

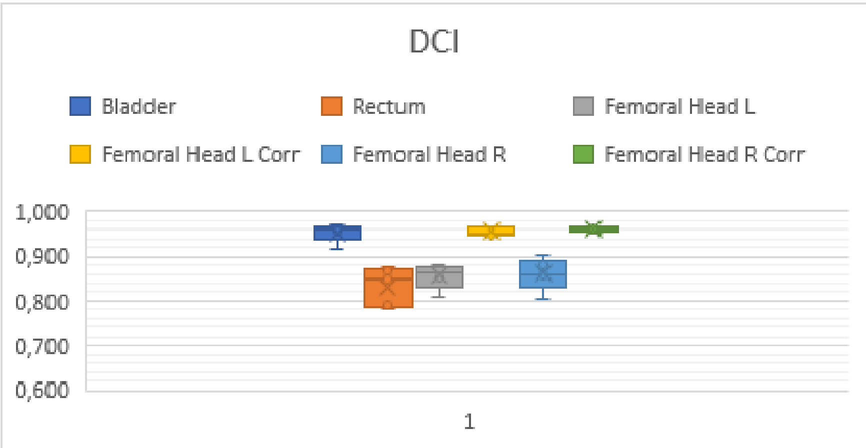


Table 1: Manual contouring time and correction time of auto-segmented contours

	Manuel Time (s)	**Correction AI (s)**	**Ecart (%)**
Case 1	*593*	*350*	*-40,98*
Case 2	*636*	*244*	*-61,6*
Case 3	*767*	*533*	*-30,5*
Case 4	*531*	*288*	*-45,7*
Case 5	*651*	*322*	*-50,5*

Conclusion

The study show positive results. To have adequate conclusion, it is necessary to perform a study with more patients.

References

[1] H. Wan, "Automated Contouring Using Neural Network", MIM Software Inc, 2020

[2] V. Thada, V. Jaglan, "Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm", International Journal of Innovations in Engineering and Technology, Pp 202-205, 2013.

Lung Nodule Volumetry and Morphology in Chest CT: Effect of Deep Learning versus Iterative Reconstruction at different dose levels

C. Franck[1], F. Zanca[2], K. Carpentier[1], H. El Addouli[1], M. C. Niekel[1], M. Spinhoven[1], A. Van Hoyweghen[1], A. Snoeckx[1]
[1]Antwerp University Hospital, Belgium
[2]Palindromo Consulting, Belgium

KEY WORDS

Computed Tomography, Artificial Intelligence, Chest, Lung nodules

Introduction

Deep learning image reconstruction (DLIR) methods have been shown to preserve the FBP-like image texture at reduced dose levels attained by Iterative Reconstruction (IR) [1-4]. This is relevant for repeated CT scanning in the context of lung cancer screening. It is unknown however whether volume measurements of detected lung nodules on low dose DLIR images are comparable with measurements on IR images. In addition, the lung nodule's morphology may change depending on the level of dose reduction and reconstruction algorithm used. The study's objective was to assess the value of DLIR compared to IR at different dose levels, in terms of lung nodule volumetry and morphology perception.

Materials and methods

An anthropomorphic chest phantom (Lungman, Kyoto Kagaku) containing 6 spherical, 6 lobulated and 6 spiculated 3D printed solid nodules (volume range 28-392 mm^3), was scanned at six dose levels (0.2, 0.4, 0.8, 1.5, 3, 6 mGy). Images were 1.25 mm reconstructed with ASIR-V 60% and three levels of DLIR (TrueFidelity Low, Medium, High).

The volumes of 432 nodules (18 nodules x 6 doses x 4 reconstructions) were measured by five experienced chest radiologists in a semi-automatic fashion. In addition, the nodule's image quality (IQ) was scored on a five-point scale (1=poor, 5=excellent). Readers were blinded for dose and reconstruction algorithm.

Mean percentage error in nodule volume measurements was assessed for all reconstructions and dose levels, with respect to the ground truth (high dose scan, 11 mGy). A smaller absolute percentage error indicates a higher accuracy. Percent IQ score frequency was calculated per reconstruction algorithm and dose level. An IQ of 3 was considered diagnostic. Subsequently, volume measurements and IQ score were stratified per nodule type.

Results

In general, mean % errors decreased with increasing dose. On average, errors were significantly lower with TrueFidelity (3.6/3.4/3.0% for Low/Medium/High) than with ASIR-V (4.1%), for all dose levels (p=0.001). With increasing DLIR level, errors decreased for the lower dose range (0.2-0.8 mGy), while for higher doses (1.5-6 mGy) values were comparable. When stratifying per morphology, the largest error was found for lobulated nodules (4.8%), followed by spiculated (3.3%) and spherical (2.8%) nodules.

Overall, IQ was higher for TrueFidelity compared to ASIR-V, with 94% of the DLIR cases having an IQ ≥ 3, versus 84% for ASIR-V. Spherical nodules had a significant better IQ score compared to lobulated nodules, for both reconstructions (p=0.003). When stratifying per dose level, the percent frequency of IQ ≥ 3 for ASIR-V/DLIR algorithms was: 0.2 mGy 26%/70%, 0.4 mGy 80%/94%, 0.8 mGy 96%/100%, 1.5 mGy 100%/100%, 3.6 mGy 100%/99%, 6.4 mGy 100%/100%.

Conclusion

In chest CT, volume measurements with TrueFidelity showed a significantly higher accuracy compared to ASIR-V, for all dose levels and all nodule types. Lobulated nodules showed the highest absolute error in volume measurements. Lung nodule morphology perception performs equally or better with TrueFidelity compared to ASIR-V, for all nodule types. At very low dose levels (0.2-0.8 mGy) DLIR outperforms ASIR-V, while at higher dose levels (≥1.5mGy), DLIR and ASIR-V are comparable in terms of nodule perception.

References

[1] Solomon, J., Lyu, P., Marin, D. and Samei, E. Noise and spatial resolution properties of a commercially available deep-learning based CT reconstruction algorithm. Solomon Med. Phys. 47(9), 3961–3971 (2020). doi: 10.1002/mp.14319.

[2] Greffier, J., Hamard, A., Pereira, F., Barrau, C., Pasquier, H., Beregi, J. P. and Frandon, J. Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. Eur Radiol. 30, 3951–3959 (2020). doi: 10.1007/s00330-020-06724-w.

[3] Racine, D., Becce, F., Viry, A., Monnin, P., Thomsen, B., Verdun, F. R. and Rotzinger, D. C. Task-based characterization of a deep learning image reconstruction and comparison with filtered back-projection and a partial model-based iterative reconstruction in abdominal CT: a phantom study. Physica Medica: Eur. J. Med. Phys. 76, 28–37 (2020). doi: 10.1016/j.ejmp.2020.06.004.

[4] Franck C, Zhang G, Deak P, Zanca F. Preserving image texture while reducing radiation dose with a deep learning image reconstruction algorithm in chest CT: A phantom study. Phys Med. 2021 Jan;81:86-93. doi: 10.1016/j.ejmp.2020.12.005. Epub 2021 Jan 11. PMID: 33445125.