Author Disclosure: D. Niraula: None. W. Sun: None. J. Jin: Research Grant; NIH-NCI. I.D. Dinov: Research Grant; NIH. K.C. Cuneo: Ann Arbor Veterans Hospital. J. Jamaluddin: None. M.M. Matuszak: Research Grant; Varian Medical Systems. Consultant; Varian Medical Systems.; Michigan Radiation Oncology Quality Consortium, AAPM. R.K. Ten Haken: Research Grant; NIH-NCI. I. El Naqa: Research Grant; NIH.; Endectra, LLC.

2223

Evaluation of an Externally Trained Deep Learning-Based Auto-Segmentation Software in the Process of Artificial Intelligence-Assisted Radiation Treatment Planning for Thoracic Cancers

M. Pan,[1,2] L. Serre,[2,3] J. Yousuf,[2] K.J. Hirmiz,[2] C. Michie,[2] L. Brown,[2] and J. Agapito[2,3]; [1]*Schulich School of Medicine & Dentistry, University of Western Ontario, London, ON, Canada,,* [2]*Windsor Regional Hospital Cancer Program, Windsor, ON, Canada,* [3]*University of Windsor, Windsor, ON, Canada*

Purpose/Objective(s): Contouring Organs-at-risk (OAR) is a laborious process that often delays radiation treatment plan design. A few FDA approved auto-segmentation software (AS) have become available. Our goal is to validate such a commercial AS in thoracic cancer OAR contouring.

Materials/Methods: We installed an externally trained AS into our AI computer. Validation is judged by our current gold standard contouring (GSC) by two experienced planners and one radiation oncologist (RO). We used 30 lung or esophageal cancer planning datasets to generate GSC and AI contours (AIC). Objective analysis included Dice Similarity Coefficient (DSC) and 95% Hausdorff distance (95% HD). Subjective analysis was done by two ROs to score 1 to 3 on all OARs by GSC and AIC that were randomly blended and anonymized with consistent nomenclature (1: no modification required; 2: minor modification required but adequate for clinical use; 3: major modification required and not suitable for clinical use).

Results: Most retrospective peer-reviewed OAR contours neglected some less important structures on CT slices typically far away from the target of the 30 patients, median age 75 years (54-90), including 22 males and 8 females, with 28 average pixel density data-sets from 4D-CT for lung cancer and 2 fast helical scans for esophageal cancer. We had to re-contour most of the OARs to generate GSC. The median GSC and AIC contouring times were 60 vs 2.5 minutes for up to 12 OARs, some of which were only partially available in the datasets (e.g., stomach and liver). Due to the inconsistency of contouring organs far away from the planning target volume, we only chose six main OARs for initial validation and analysis. Comparing AICs to GSCs, the mean DSC and 95% HD were: esophagus 0.61 and 16 mm, heart 0.85 and 13.1 mm, left lung 0.97 and 5.9 mm, right lung 0.96 and 5.7 mm, spinal cord 0.82 and 10.7 mm, trachea and proximal bronchial tree (TPB) 0.67 and 19.1 mm, respectively. The two ROs agreed with 100% of four OARs on GSC, i.e., both RO scoring 1 or 2 meaning adequate for planning purpose, with the exception of esophagus having 96.7% vs 100% and right lung having 100% vs 96.7% agreement, respectively. They had less agreement on AIC, with esophagus 90% vs 60%, heart 83.3% vs 86.7%, left lung 100% vs 96.7%, right lung 100% vs 96.7%, spinal cord 100% vs 100%, TPB 96.7% vs 86.7% agreement, respectively. The interobserver variabilities are significantly larger when ROs evaluated esophagus and heart AIC (p=0.046 and 0.05, respectively, Student's t-test).

Conclusion: The accuracy of an externally trained deep learning-based AS might not be acceptable without in-house training from local protocols. Retrospective peer-reviewed OAR contours might not be good enough in the training and evaluation of AS. Our future work involves training AS using our GSCs and re-evaluating its performance.

Author Disclosure: M. Pan: Research Grant; Windsor Cancer Centre Foundation. L. Serre: None. J. Yousuf: None. K.J. Hirmiz: None. C. Michie: None. L. Brown: None. J. Agapito: None.

2224

Evaluation of AI vs. Clinical Experts SBRT-Thorax Computed Tomography OARs Delineation

S. Stathakis,[1] G. Pissakas,[2] A. Alexiou,[2] B. Bertrand,[3] P.Y. Bondiau,[4] L. Claude,[5] T. Cuthbert,[6] A. Damatopoulou,[2] C. Dejean,[4] C. Doukakis,[2] G. Güngör,[7] L. Hardy,[8] E. Maani,[1,9] I. Martel-Lafay,[5] P. Mavroidis,[10] N. Paragios,[8,11] V. Peppa,[2] D. Remonde,[1] J.W. Shumway,[10] G. Ugurluer,[12] and E. Ozyar[12]; [1]*The University of Texas Health Science Center - San Antonio, Department of Radiation Oncology, San Antonio, TX,* [2]*Department of Radiation Oncology, Alexandra Hospital, Athens, Greece,* [3]*Therapanacea, Paris, France,* [4]*IMPT, Centre Antoine Lacassagne, Radiation Oncology, Nice, France,* [5]*Centre Léon Bérard, Radiation Oncology Department, Lyon, France,* [6]*The University of Texas Health Science Center-San Antonio, Department of Radiation Oncology, San Antonio, TX,* [7]*Acibadem MAA University, Maslak Hospital, Department of Radiation Oncology, Istanbul, Turkey,* [8]*TheraPanacea, Paris, France,* [9]*South Texas Veteran Health Affairs Care System, San Antonio, TX,* [10]*Department of Radiation Oncology, University of North Carolina, Chapel Hill, NC,* [11]*CentraleSupelec, University of Paris-Saclay, Computer Science and Applied Mathematics, Gif-sur-Yvette, France,* [12]*Acibadem MAA University, Maslak Hospital, Radiation Oncology, Istanbul, Turkey*

Purpose/Objective(s): Organ-at-risk (OAR) delineation is an essential step for radiotherapy treatment (RT) planning. Standard practice during the last decade referred to manual delineation of OARs from medical experts. Manual delineation is the current standard of practice and can be a tedious, time-consuming process, prone to errors due to intra and inter-observer variations. This is the case for lung stereotactic body radiation therapy (SBRT) due to the variety of structures to be contoured. The latest advances of artificial intelligence methods offer new perspectives towards a fully automatic delineation. This study aims at evaluating an AI-based auto-contouring solution (AC) and compare its clinical acceptability against contours delineated by experts (EC).

Materials/Methods: In this study, a CE/FDA cleared anatomically preserving ensemble deep-learning architecture contouring solution was used. The AI-solution was trained using more than 300+ fully annotated multi-centric SBRT cases according to the ESTRO guidelines (male and female) cases. A fully external cohort of 30 additional patients was considered to assess performance quantitatively and qualitatively. In terms of quantitative metrics, the Dice coefficient was used while qualitative assessment was done through a scoring mechanism: A/acceptable, B/ acceptable after minor corrections, and C/ not acceptable for clinical use. In terms of reference/comparisons with clinical experts, two independent annotations were used while treatment experts' contours were blended blindly with the AI-solution 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.

Results: The mean Dice coefficient between expert and AI was 88 % which dropped to 87% among experts. Overall clinical acceptability after aggregating blinded evaluations for the combined categories (A+B) was 95% for the AC which dropped to 81% for EC. When looking at the overall acceptability of contours that do not require any modifications (A), the AC (67%) outperformed the EC (48%) by significant margin. AC outperformed the EC on 14 structures.

Conclusion: This work reports clinical evaluation of AC solution on CT scans for Thorax-SBRT. Our results suggest that the deep learning model could be a viable clinical alternative to the human expert offering treatment standardization and potentially better outcomes.

Abstract 2224 — Table 1: Clinical results comparing AC and EC

ORGANS	A% for EC	B% for EC	A% for AC	B% for AC	Mean AC Dice (%)	Mean EC Variability Dice (%)
Aorta	49	34	52	48	72	/
Bronchial_tree	54	46	40	50	78	70
Esophagus	50	44	81	19	84	88
Heart	30	41	48	48	95	95
Brachialplexus_L	55	32	47	42	66	/
Chestwall_L	5	10	44	38	90	90
Kidney_L	67	33	86	14	98	/
Lung_L	78	19	85	15	99	/
Liver	46	50	81	19	98	/
Pulmonary_arteries	31	54	64	33	86	83
Brachialplexus_R	44	44	41	47	69	/
Chestwall_R	9	6	57	32	90	91
Kidney_R	67	33	89	11	96	/
Lung_R	76	24	77	19	99	/
Spinal_cord	50	38	89	7	89	/
Spleen	67	33	84	15	94	93
Stomach	33	67	62	31	90	88
Trachea	64	26	45	40	91	88

Author Disclosure: S. Stathakis: None. G. Pissakas: None. A. Alexiou: None. B. Bertrand: None. P. Bondiau: None. L. Claude: None. T. Cuthbert: None. A. Damatopoulou: None. C. Dejean: None. C. Doukakis: None. G. Güngör: None. L. Hardy: None. E. Maani: None. I. Martel-Lafay: None. P. Mavroidis: None. N. Paragios: None. V. Peppa: None. D. Remonde: None. J.W. Shumway: None. G. Ugurluer: None. E. Ozyar: None.

2225

A Multi-Centric Evaluation of AI-Driven OARs Low Field MRgRT Pelvic /Abdomen Contouring

D. Azria,[1] N. Andratschke,[2] P. Balermpas,[2] L. Boldrini,[3] R. Bourdais,[4] A. Bruynzeel,[5] M.D. Chuong,[6] M. De Ridder,[7] P. Fenoglietto,[1] T. Gevaert,[7] G. Gungor,[8] L. Hardy,[9] S. Kandiban,[9] F. Lagerwaard,[5] P. Maingon,[4] A.E. Marciscano,[10] K.E. Mittauer,[11] H. Nagar,[10] N. Paragios,[9] R. Pennell,[10] L. Placidi,[3] O. Riou,[1] J.M. Simon,[4] S. Tanadini-Lang,[2] G. Ugurluer,[8] S. Valdes,[6] V. Valentini,[3] B. Vanspeybroeck,[7] and E. Ozyar[8]; [1]*L'Institut du Cancer de Montpellier, Department of Radiation Oncology, Montpellier, France,* [2]*Department of Radiation Oncology, University Hospital Zurich, Zurich, Switzerland,* [3]*Fondazione Policlinico Universitario A. Gemelli IRCCS, Radiation Oncology, Rome, Italy,* [4]*AP-HP Sorbonne Université, Charles-Foix Pitié-Salpêtrière, Paris, France,* [5]*Amsterdam UMC, Dept. of Radiation Oncology, Amsterdam, Netherlands,* [6]*Miami Cancer Institute, Department of Radiation Oncology, Miami, FL,* [7]*UZ Brussel, Vrije Universiteit Brussel, Radiotherapy Department, Brussels, Belgium,* [8]*Acibadem MAA University, Maslak Hospital, Radiation Oncology, Istanbul, Turkey,* [9]*Thera-Panacea, Paris, France,* [10]*NewYork-Presbyterian/Weill Cornell Hospital, Radiation Oncology, New York, NY,* [11]*Department of Radiation Oncology, Miami Cancer Institute, Baptist Health South Florida, Miami, FL*

Purpose/Objective(s): Organs at risk annotation is a strong bottleneck of Magnetic resonance imaging guided radiotherapy (MRgRT) in the context of adaptive treatment. It is a time-consuming task that reduces patient throughput (20% of the fraction duration dedicated to contouring) while suffering standardization and reproducibility across physicians, hampering the accuracy of high precision MRgRT and diminishing its adoption potential. AI-contouring becomes a game changer in radiation oncology since it is able within seconds to provide a full OAR delineation that could be close to clinical acceptance with little modifications. The aim of this study is to evaluate the performance of AI-contouring within a multi-centric cohort for patients with pelvic / abdominal tumors treated with low field (0.35T) MRgRT.

Materials/Methods: In the context of this study, a CE/FDA cleared anatomically preserving ensemble deep-learning architecture contouring solution was considered. The adopted solution was trained using more than 350 0.35T MR fully annotated pelvic cases according to the ESTRO guidelines and 270 annotated abdomen fractions samples. A retrospective cohort involving 42 test cases coming from seven different institutions (US: 1, EU: 5, AS: 1) was considered. The clinical delineations used for treatment planning from expert physicians/ medical physicists were associated with these cases.

Results: It appears that treatment practices can be very different between institutions since the use of OAR constraints were far from being uniformly adopted. Bladder & liver dosimetry constraints were the most frequently used (100% & 90%) while abdominal aorta and seminal vesicle were the least adopted (24% & 15%). The average DSC between Clinical experts and AI annotations was 78% across all structures. Bladder and left/right kidney were the structures for which the highest DSC were observed (93%, 91% & 90%), while penile bulb and duodenum were the ones with the lowest agreement (54% & 59%). AI solutions seem to have important discrepancies with clinical contours in organs on which either the volume is small or there are practice-related uncertainties with respect to the definition of beginning and the end of the structure. For quantitative evaluation, dice similarity coefficients (DSC) and 95% Hausdorff distances (HD95) were calculated.

Conclusion: This retrospective multi-centric study demonstrates that AI-driven contours could be a reliable alternative to clinical contours offering performance that appears to be close to the human expert for many of the structures while increasing throughput and offering automatization & standardization.

Abstract 2225 — Table 1

Organ	Average dice (%)	Average HD95 (mm)
Anal canal	65	6
Bladder	93	3
Left femoral head	84	10
Penile bulb	54	7
Prostate	83	6
Rectum	81	15
Right Femoral Head	85	8
Seminal Vesicle	81	3
Sigmoid	77	13
Abdominal aorta	74	42
Duodenum	59	45
Large bowel	64	57
Left kidney	91	5
Liver	89	29
Pancreas	48	27
Right Kidney	90	8
Stomach	84	22
Vena cava inferior	72	22

Author Disclosure: D. Azria: None. N. Andratschke: None. P. Balermpas: None. L. Boldrini: None. R. Bourdais: None. A. Bruynzeel: None. M.D. Chuong: Research Grant; ViewRay. Honoraria; ViewRay, Sirtex. Speaker's Bureau; ViewRay, Sirtex. Advisory Board; ViewRay, Advanced Accelerator Applications. Travel Expenses; ViewRay.; Miami Cancer Institute, Particle Therapy Cooperative Group. Oversees the individual disease site committees; Proton Collaborative Group. M. De Ridder: None. P. Fenoglietto: None. T. Gevaert: None. G. Gungor: None. L. Hardy: None. S. Kandiban: None. F. Lagerwaard: None. P. Maingon: None. A.E. Marciscano: None. K.E. Mittauer: None. H. Nagar: Research Grant; ViewRay, Lantheus, Teleflex. Advisory Board; ViewRay, Boston Scientific. Advisory Board; Boston Scientific; ViewRay. N. Paragios: CEO; Thera-Panacea. R. Pennell: None. L. Placidi: None. O. Riou: None. J. Simon: None. S. Tanadini-Lang: None. G. Ugurluer: None. S. Valdes: None. V. Valentini: None. B. Vanspeybroeck: None. E. Ozyar: None.