Clinics of Oncology

Clinical Paper

ISSN: 2640-1037 | Volume 6

Clinical Validation and Quantitative Evaluation of a Novel Artificial-Intelligence Automatic Segmentation Tool in Radiotherapy Planning

Peng JL*, Godwin WJ, Maynard MR, Rapchak AK, Roles SA and McDonald DG

Department of Radiation Oncology, Medical University of South Carolina, Charleston, South Carolina, USA

***Corresponding author:**

Jean L Peng,
Department of Radiation Oncology, Medical University of South Carolina
169 Ashely Ave, Charleston, SC, USA 29425,
E-mail: pengl@musc.edu

Received: 14 Oct 2022
Accepted: 26 Oct 2022
Published: 31 Oct 2022
J Short Name: COO

Citation:

Peng JL. Clinical Validation and Quantitative Evaluation of a Novel Artificial-Intelligence Automatic Segmentation Tool in Radiotherapy Planning. Clin Onco. 2022; 6(13): 1-7

Keywords:

Artificial-intelligence auto-segmentation; Head and Neck (H&N); CNS; Breast; Radiotherapy

1. Abstract

1.1. Purpose: Organ-at-risk (OAR) contouring for radiation treatment planning requires significant physician effort. Manual segmentation (MS) of patient organs remains a large time sink for physicians in radiation oncology. Auto-contouring systems aim to reduce this effort, but quality remains inconsistent. Artificial-intelligence auto-segmentation (AI-AS) has emerged as an alternative to atlas-based methods, promising improved results without the effort necessary to create a functional atlas. This study evaluates a novel, commercial AI-AS system for head and neck (H&N), central nervous system (CNS), and breast OAR contouring via comparison to manually delineated contours by an experienced radiation oncologist.

1.2. Methods: A total of thirty patients, ten per treatment site, with H&N, CNS or breast cancer treated at our institution were randomly selected retrospectively for this study. OARs, chosen by treatment site, were delineated manually on each dataset by the physician. The delineated OARs included the brainstem, optic nerves, optic chiasm, and normal brain for CNS patients, larynx, left and right parotids, oral cavity, brainstem, spinal canal, and mandible for H&N patients, and contralateral (normal) breast, heart, lungs, esophagus, and spinal cord for breast patients. Contours for all structures were then generated using the AI-AS system -AI-Rad Companion OrgansRT™ (Siemens healthineers, Forchheim, Germany). All MS structures were compared to their respective AI-AS structures via Dice Similarity coefficient (DSC), mean distance to agreement (DTAmean) and max distance to agreement, also known as the Hausdorff distance (HD). Similarity metrics for each structure were then averaged across all evaluated datasets.

1.3. Results: In CNS cases, DSCs ranged from 0.36 ± 0.22 to 0.98 ± 0.00 for the structures analyzed. DTAmean values ranged from 0.76 ± 0.13 mm to 3.25 ± 0.99 mm for the structures analyzed. HD values ranged from 8.53 ± 5.29 mm to 17.80 ± 15.09 mm. For H&N cases, average structure DSCs ranged from 0.37±0.05 to 0.85±0.03. Average DTAmean values ranged from 1.04±0.26 mm to 10.15±0.94 mm. Average HD values ranged from 5.18±2.38 mm to 36.64±5.00 mm. Average DSCs for the left and right parotids, and mandible were greater than 0.75, indicating good agreement. The mandible showed particularly strong agreement, with a minimum DSC of 0.80 across all datasets analyzed. The larynx and oral cavity did not agree well, with average DSCs below 0.40. For breast cases, correlation between the AI-AS generated and MS volumes was strong for the evaluated OARs, with a mean DSC higher than 0.90 for all OARs except the esophagus (mean DSC 0.83 ± 0.09)). Mean DSC values ranged from 0.98 ± 0.01 to 0.83 ± 0.09, while the DTAmean ranged from 0.68 ± 0.3 to 0.84 +0.54 mm to mm. HD values ranged from 9.88±15.99 mm to 25.61±8.78 mm. For all body sites, the AI-AS system was able to generate a full set of contours in approximately 2 minutes, while manual contouring of these structures required approximately 30 minutes.

1.4. Conclusions: The AI-AS system was able to reasonably match manual contours for multiple structures for CNS, H&N and breast radiation treatment planning. The brainstem, left and right parotids, spinal canal, mandible, lung, heart, esophagus and normal breast consistently showed the best agreement. In CNS, optic structure definitions are often aided by MR-imaging that can significantly improve a user's ability to delineate structures when compared to an AI-AS working solely on CT imaging. All contours generated with AI-AS require review by an experienced physician. Although edits will be needed for most structures due to system limitations and physician-preference, these edits should be minor for many structures, and significant time-savings can be achieved compared to a solely manual contouring approach. Furthermore, initial volumes created by AI-AS are physician-independent, and may help to improve contour consistency both within a clinic, and across clinics, when compared to strictly manual approaches.

2. Introduction

Modern radiation therapy (RT) planning is a complex process that primarily relies on computed tomography (CT)-based three-dimensional (3D) imaging for target and organ at risk (OAR) delineation. In standard practice, before planning can begin, these structures are drawn manually by the radiation oncologist or dosimetrist. Accurate contour delineation is essential for a good treatment outcome, but requires a significant time commitment from dosimetrists and physicians. In addition, definition of targets and OARs adjacent to targets require expert knowledge of the complex anatomy and disease pathways in the affected region, particularly the lymph nodes, which are the most frequent sites of cancer recurrence and metastasis [1]. The average physician and or dosimetrist time commitment for contouring of a standard head and neck (H&N) IMRT plan may, in some cases, exceed three hours [2]. In addition, depending on differences in the clinical experience of physicians and dosimetrists, targets and OARs definition may also have high intra- and inter-observer variability [3]. Because OARs are more predictable than treatment target volumes, they may lend themselves more readily to automatic segmentation (AS) methods. AS has drawn enormous attention in radiation therapy because it has the potential to reduce the time required for OAR contouring drastically and to reduce intra- and inter-observer variability [4, 5]. Reduced segmentation time may allow radiotherapy to be initiated sooner. This is especially beneficial for rapidly growing cancers, such as cancers of the H&N. AS may also allow for more widespread use of adaptive radiotherapy [6].

An earlier framework of AS systems, atlas-based auto-segmentation (AB-AS) has been shown to effectively reduce contouring time and improve contouring consistency of OARs in radiation therapy [5,7]. AB-AS takes advantage of deformable image registration to map well-defined contours on the atlas, which typically come from previous patients' treatment data, to the new patient's CT image for auto contouring. However, many OARs do not have the constant image intensity or clear anatomic boundaries necessary for computer vision-based algorithms to work effectively in a uniform manner. Low spatial accuracy of auto-contours relative to conventional manual contours have been a barrier to widespread adoption of auto-contouring tools in the clinical setting.

The high variability in different patients' anatomies and the low contrast of lymph nodes or blood vessels pose a significant challenge for targets and OAR auto segmentation. Several AB-AS methods [8, 9] have been proposed for H&N lymph node region delineation, and discussions have been conducted regarding the uncertainty and limitation of AB-AS methods from these low contrast organs. Therefore, researchers have aimed to improve auto-contouring recently using artificial intelligence (AI) technology based on deep-learning neural networks. Machine learning approaches, especially deep-learning with multi-layered neural networks, can process large datasets in order to provide automated solutions without the need for manual feature extraction. AI-based auto-segmentation (AI-AS) utilizes a database of previously contoured CT-datasets as training data for the neural network. New patients are then automatically contoured based on the data within the learned network.

Previous investigations have evaluated the accuracy of AI-AS algorithms developed in-house for various body sites [10-12]. AI-AS of OARs and CTVs has been shown to provide greater accuracy and time savings compared to AB-AS13. However, development of in-house AI-AS models is not feasible in many clinical settings, due to the required technical coding skill necessary, and the large effort needed to collect sufficient training datasets. As a result, the use of AI-AS in clinical practice remains uncommon.

Recently, commercial options for AI-AS have been introduced, which aim to make this technology more accessible. These systems utilize centralized networks, relieving the customer of the work needed for algorithm training, and increasing the speed of implementation. However, specific features developed during training and the effect on final output may not be readily available or apparent to the customer. For this reason, robust studies evaluating performance of deep learning algorithms are required prior to clinical implementation. In this study, AI-Rad Companion Organs RT™ (Siemens healthineers, Forchheim, Germany), a completely trained, fully automated commercially available AI-AS model, was evaluated for accuracy and efficiency gain versus contours generated manually by a qualified physician. CT-based AI-AS generated organs typical for H&N, central nervous system (CNS), and breast cancers were evaluated. Individual AI-AS structure models may behave differently, and, as a result, each structure must be independently validated. This study is the first clinical validation report on this novel commercial AI-AS software program with available structures. However, the new structures would be

added in future version release and each new structure model in AI-AS should be validated individually.

3. Materials and Methods

3.1. Artificial-Intelligence-based auto-segmentation (AI-AS)

The commercial deep learning-based auto-segmentation software, AI-Rad Companion Organs RT™ Contour build 1.0 (Siemens healthineers, Forchheim, Germany), uses deep convolutional neural network models. AS models were developed prior to the inception of this study; complete details of this software, including neural networks and architecture, have not been made public by the manufacturer. These models were trained using publicly available data; no local institutional data was used. The auto-segmentation models of OARs for H&N, CNS, and breast RT planning were selected for validation based on our institution's demands. The training data for each model consisted of publicly available contoured 2D axial computed tomography (CT) images. Specifically, data used for CNS model training included contours of brainstem, optic nerve, optic chiasm, normal brain and optic globe. H&N model training included contours of brainstem, spinal canal, larynx, left and right parotids, oral cavity, and mandible. Breast RT model training included contours of contralateral breast, heart, lungs, esophagus, and spinal cord. Data augmentation (flipping, brightness adjustments, and elastic deformations) and regularization techniques (dropout, batch normalization) were used during training to improve model performance and prevent overfitting. AI-Rad Companion Organs RT™ is able to utilize the graphics processing unit (GPU) for segmentation generation; GPU is the manufacturer's default option, since it allows for implementation into a wider range of clinical systems, and was chosen for this study.

3.2. Patient Selection and OAR Segmentations

A total of thirty patients with H&N, CNS or breast cancer treated at our institution were randomly selected retrospectively for this study. Each site had ten patients. A SOMATOM Confidence® RT Pro CT scanner (Siemens healthineers, Forchheim, Germany) was used to acquire 120 kV CTs,600 mm field of view and 512 × 512 matrix. For each patient, a 3-dimensional planning CT with 3-mm slice thickness was used for contouring. Contrast was not used for any patients' planning CT in this study. Five breast patients received breast-conserving surgery with breast implant. The CT scans for these patients were reconstructed with the iterative metal artifact reduction (iMAR) algorithm (Siemens healthineers, Forchheim, Germany) per institutional protocol. OARs were manually contoured on all image studies by a qualified radiation oncologist according to institutional guidelines and published recommendations. Specific OARs were contoured based on treatment site. For CNS datasets, brainstem, optic nerves, optic chiasm, and normal brain were contoured. For H&N datasets, larynx, parotids, oral cavity, brainstem, spinal canal, and mandible were contoured. For breast datasets, contralateral (normal breast), heart, lungs, esophagus, and spinal cord were contoured. with specific OARs selected based on treatment site. The Organs RT™ system was then used to automatically generate contours of the applicable OARs for these same datasets. Generally, automatically generated contours were left unchanged. For H&N datasets, the automatically generated brainstem was cropped in the inf-sup direction to match the extent of brainstem contoured by the physician, as not all H&N cases required a full brainstem contour. In addition, for H&N cases, the spinal canal was contoured by the physicians. Organs RT™ contours the spinal cord only. This automatically generated spinal cord was expanded by 4mm to create a spinal canal for evaluation. Of note within the CNS patient cohort, physician-created "Whole Brain" structure was compared to the combination of the "Brain," "Brainstem," and "Optic Chiasm" substructures within AI-Rad Companion Organs RT™.

Thus, two groups of contours were collected for analysis: The manually contoured group used for clinical treatment and meeting clinical acceptance, and the AI-AS group made up of the non-edited automatically created contours. Manually contoured structures on a given dataset were directly compared with their AI-AS counterparts using a commercial analysis software Velocity™ (Version 4.1, Varian Medical Systems, Inc, Palo Alto, CA for both numeric and visual omparison.

3.3. Comparison Metrics

Geometric indices comparing the manually contoured and AI-AS structures were calculated for each structure, including dice similarity coefficient (DSC), mean distance to agreement (DTA_{mean}) and Hausdoff distance (HD). 1. Dice similarity coefficient (DSC) is a geometric volumetric similarity measure used to determine the degree of overlap of two set of contours, which provides a value that simultaneously quantifies differences in volume and orientation for nonsymmetric shape of contours. DSC normalizes the intersection volume to a value between 0 (no overlap) and 1 (perfect overlap) and is defined as:

$$DSC = \frac{2|V_m \cap V_a|}{|V_m| + |V_a|}$$

where V_m and V_a are the volumes of the manual drawn and auto-segmented contours, respectively.

2. The mean distance to agreement (DTAmean) is another measure of relative contour overlap. Distance to agreement is defined as the distance from any point on a contour surface to the nearest point on the second structure's surface. The mean distance to agreement is the mean of this set of minimum distances. Mathematically this is given by:

$$DTA_{mean} = mean \begin{Bmatrix} \min d(a) & \min d(b) \\ a \in A & b \in B \end{Bmatrix}$$

"a" and "b" represent the points on contour A and contour B, where $\min_{a \in A} d(a)$ is the minimum distance of all points on the con-

tour A to points on the contour B, so as the same definition used for $\min_{b \in B} d(b)$. The DTA_{mean} is the mean distance over all distances from points in A to their closest point in B.

3. The Hausdorff distance is the maximum distance to agreement of the structures being compared. Mathematically, Hausdorff distance (HD) is given by:

$$HD = DTA_{max} = max\left\{\min_{a \in A} d(a) \quad \min_{b \in B} d(b)\right\}$$

4. Results

4.1. CNS

Within the CNS patient cohort, each structure's comparison metrics were averaged across the cohort. Data for each organ within the cohort is provided in Table 1a. The resulting average structure DSCs ranged from 0.36 ± 0.22 to 0.98 ± 0.00, showing a wide range of reliability highly dependent on the structure of interest. DTAmean ranged from 0.76 ± 0.13 mm to 3.25 ± 0.99 mm. HD for this set ranged from 6.64 ± 4.42 to 17.80 ± 15.09 mm. The highest HD values among all evaluated structures belonged to the whole-brain structure, and prompted further visual inspection of notable differences in whole brain contouring. AI-AS for a CNS patient required approximately 2 minutes. MS was estimated to require approximately 30 minutes of professional time by the dosimetrist or physician.

Table 1: The Dice similarity coefficient (DSC), mean distance to agreement (DTAmean), and Hausdorff distance (HD) over all distances from points in contours A to their closest point in contours B between AI-AS (artificial intelligence-based auto segmentation) and manually segmentation (MS) are listed as mean ± standard deviation format. (a) CNS (b) H&N (c) breast

(a)

Structures	DSC	DTA_{mean} (mm)	HD (mm)
Brainstem	0.74 ± 0.05	3.25 ± 0.99	14.93 ± 3.81
Brain	0.98 ± 0.00	0.76 ± 0.13	17.80 ± 15.09
Right optical nerve	0.67 ± 0.06	1.06 ± 0.57	6.64 ± 4.42
Left optical nerve	0.67 ± 0.06	1.14 ± 0.86	8.80 ± 5.96
Chiasm	0.36 ± 0.22	2.13 ± 1.54	8.53 ± 5.29

4.2. H&N

Within the H&N cohort, average structure DSCs ranged from 0.37 ± 0.05 to 0.85 ± 0.03. Average DTAmean values ranged from 1.04 ± 0.26 mm to 10.15 ± 0.94 mm. Average HD values ranged from 5.18 ± 2.38 mm to 36.64 ± 5.00 mm. Average DSCs for the brainstem, left and right parotids, spinal canal, and mandible were greater than 0.75, indicating good agreement. The mandible and spinal canal showed particularly strong agreement, with minimum DSCs of 0.80 and 0.77, respectively, across all datasets analyzed. The larynx and oral cavity did not agree well, with average DSCs below 0.40. Data for each organ within the cohort is provided in Table 1b. The AI-AS system was able to generate a full set of contours in approximately 2 minutes, while MS of these structures required approximately 30 minutes

(b)

Structures	DSC	DTA_{mean} (mm)	HD (mm)
Brainstem	0.79±0.07	2.43±1.19	10.48±4.30
Spinal Canal	0.85±0.03	1.04±0.26	5.18±2.38
Larynx	0.37±0.05	6.21±0.35	23.49±1.53
Parotid Lt	0.81±0.05	2.31±0.88	18.50±8.21
Parotid Rt	0.78±0.11	2.64±1.50	17.25±8.05
Oral Cavity	0.39±0.04	10.15±0.94	36.64±5.00
Mandible	0.84±0.04	1.32±0.50	13.78±5.50

4.3. Breast

Correlation between the AI-AS and MS generated volumes was strong for the evaluated OARs within the breast cohort, with a mean DSC higher than 0.90 for all OARs except the esophagus (mean DSC 0.83 ± 0.09). In addition, ipsilateral breast tissue (clinical treatment target) volumes also showed excellent results, with a mean DSC higher than 0.93. For the OARs examined, mean DSC values ranged from 0.83 ± 0.09 to 0.98 ± 0.01, while the DTAmean ranged from 0.68 ± 0.30 to 0.84 +0.54 mm. HD values ranged from 9.88 ± 15.99 mm to 25.61 ± 8.78 mm. Data for each organ within the cohort is provided in Table 1c. AI-AS generated volumes were produced in approximately 2 min, while MS of the same volumes takes the physician an estimated 30 min.

(c)

Structures	DSC	DTA_{mean} (mm)	HD (mm)
Ipsilateral breast	0.93±0.03	1.76±0.68	25.61±8.78
Contralateral breast	0.94±0.04	1.71±0.91	24.48±14.01
Left Lung	0.97±0.02	0.82±0.48	17.37±9.31
Right Lung	0.98±0.01	0.84±0.54	23.73±14.62
Heart	0.94±0.02	1.66±0.65	10.15±2.44
Spinal Cord	0.89±0.02	0.99±0.78	24.94±66.17
Esophagus	0.83±0.09	0.68±0.30	9.88±15.99

5. Discussions

5.1. CNS

AI-AS struggled to accurately identify most structures analyzed in the CNS patient cohort. Only the brain had an average DSC of above 0.75, indicating good agreement. The brainstem had an average DSC of 0.74±0.05, indicating near agreement. Visual review identified additional clinically relevant issues. Brain and brainstem contour results varied depending on the inferior extent of the brainstem contour. This contour was truncated by the physician and limited to the clinically useful extent. OrgansRT™ attempted to contour the entire extent. For the H&N patient cohort, the AI-AS brainstem was truncated to match the MS extent, and agreement increased significantly. Similarly, the brain contour, which includes the brainstem, was affected by the extent of the brainstem

contoured by the physician. AI-AS brain structures also consistently missed normal brain tissue in the cribriform plate region. Review of the optic structures showed significant discrepancies. AI-AS optic nerves and chiasms often did not connect (Figure 1). This could lead to meaningful changes in inversely obtained dose distributions, as optimization algorithms will not reduce dose in a region unless explicitly instructed to do so. Leaving false gaps between optic structures could result in unintended dose in these areas. Finally, as seen in Figure 2, optic structures are often difficult to identify on CT images. In the clinical setting, magnetic resonance (MR) imaging can significantly improve a user's ability to delineate these structures. While OrgansRT™ currently lacks the ability to utilize MR imaging, the addition of this feature would likely improve segmentation of these structures. If quantifiable metrics are to be used heavily in analysis of AI-AS across a cohort of patients, as opposed to strict visual inspection, special care must be taken to observance of the standard deviation and outliers within the set. The current generation of AI-AS create a useful head start for physicians' contours of the brain, but require significant editing based on anatomical interpretation inaccuracies and physician preference.

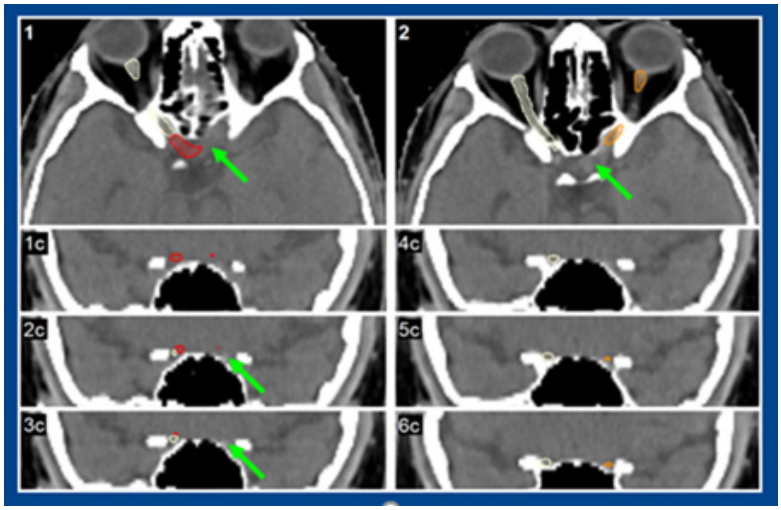

Figure 1: Clinically significant disconnect between left optic nerve and optic chiasm contours generated by AI-AC shown in axial and coronal views. (Numeric labeling to identify immediately adjacent image slices. Green arrows to emphasize locations of potential nerve-chiasm connection that would impact optimization decisions

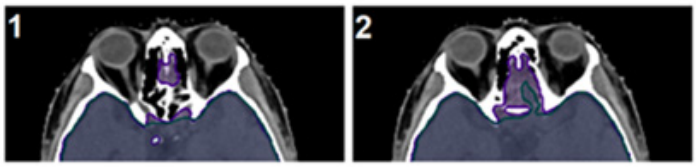

Figure 2: Discrepancy in definition of brain tissue in proximity to cribriform plate. (Clinically approved contours PURPLE, AI-generate contours GREEN)

5.2. H&N

The AI-AS system was able to reasonably match manual contours for multiple structures typically used for H&N treatment planning. A structure was considered to have good agreement if the average DSC was greater than 0.75. In addition, mean distance to agreement was less than 3mm for structures showing good agreement. Structures such as the parotids and mandible showed very good agreement with no post-processing of the automatically generated contours necessary. The brainstem and spinal canal also showed good agreement, but required some minor manipulation of the automatically generated contours in order to make them more clinically relevant. For the brainstem, OrgansRT™ contours the entire brainstem, which may not be necessary depending on the target volume location. As a result, meaningful comparison required cropping of the automatically generated structure. However, in general use, it should not affect the planning process to retain the entire automatically generated volume. For the spinal canal, a 4mm expansion was added to the automatically generated spinal cord to create a spinal canal volume. Clinically, the spinal canal is used as a conservative surrogate for the spinal cord. Although the addition of a margin to the automatically generated contour requires an additional step, margin expansions are standard practice in commercial treatment planning systems, and can themselves be automatically generated through scripting or templated protocols. Both the larynx and oral cavity did not show good agreement. This was primarily due to a fundamental difference in the definition of these structures by the physician and OrgansRT™. The physician defined the larynx to include the airspace inside the larynx, thyroid cartilage, cricoid cartilage, arytenoid cartilage, and a portion of the subglottic larynx. This definition is consistent with our treatment

planning process, and allows for more effective shielding of the larynx during treatment plan optimization. OrgansRT™ primarily contoured the true and false vocal cords, stopping superior to the subglottic larynx. Although OrgansRT™ accurately outlined the true and false vocal cords, the resulting contour was too limited to be considered useful for treatment planning. Similarly, the oral cavity defined by the physician was more generous to allow for shielding of this area during optimization, while the contour generated by Organs RT™ was more limited.

5.3. Breast

No clinically significant differences between contours in the breast cohort structures including targets (ipsilateral breasts) were identified. Each structure achieved high DSC (0.83-0.98, very close to 1) and low DTAmean values (0.68-1.76mm, less than 2mm).

Despite generally high-quality AI-As structures, high HD values (>20mm) were observed in this study. The inconsistencies were generally located near low contrast organs such as contralateral (normal) and ipsilateral (treated) breasts (targets). During the manual contouring process, physicians must refer to data not contained in CT imaging such as diagnosis and surgical history to determine whether lymph nodes, chest walls, or ribs should be included in target regions. Auto-segmentation cannot utilize such information. Performance in this regard may be improved by: (1) increasing the diversity of the training data (with and without potential positive lymph nodes) and (2) improving the consistency of the training data (e.g., all lymph nodes are included, or none are included).

5.4. Contouring Time

Although no formal analysis was performed regarding the efficiency, defined as the time required for the manual contouring of each structure in this study, it is clear that the AI-AS has a significant advantage over manual contouring. The AI-AS method required less than 2 minutes to contour an entire image set for a given patient across all cohorts. Review and modification of the AI-AS contour set regularly required less than 5 minutes of professional time. A remaining issue exists in terms of target contouring. The complexity of a given target and the experience of the physician determines the duration of the manual contouring process. If the AI-AS could be used as an assistance tool, the time required for the contouring of targets in addition to OARs, the time savings will likely be very significant to the treatment workflow of the clinic. It is noteworthy that the number of the cases evaluated in this study is relatively small. Additionally, comparison of the time required for manual contouring by physicians with different qualifications and for AI-AS-assisted contouring in larger cohorts is warranted.

6. Conclusions

In conclusion, we have demonstrated that the accuracy of deep learning-based auto-segmentation is comparable to segmentation by an expert radiation oncologist for many organs at risk used for radiation treatment planning. The brain, brainstem, left and right parotids, spinal canal, mandible, lung, heart, esophagus and normal breast consistently showed the best agreement. Structures modified by the physician to accomplish treatment planning objectives, such as the larynx and oral cavity did not show good agreement, primarily due to differing intents between the AI-AS system and the physician. In addition, optic structures such as the optic nerves and chiasm showed poor agreement. AI-AS of these structures could be improved by incorporating anatomical rules, such as ensuring connection between the chiasm and nerves. In addition, MR-based auto-contouring would likely be more effective for these structures.

All contours generated with AI-AS require review by an experienced physician. Although edits will be needed for most structures due to system limitations and physician-preference, these edits should be minor for many structures, and significant time-savings can be achieved compared to a solely manual contouring approach. Of note, in the current clinical workflow, manual contours of OARs are often generated by resident physicians or dosimetrists first, followed by review and modification by the attending radiation oncologist. AI-AS can be used by residents and dosimetrists to save time and improve contour consistency and accuracy, so that time spent by senior physicians modifying the contours can also be reduced. This works to improve training, a key goal, especially for radiation oncology residency programs, and save clinical time for all parties.

As the methods for generating auto-segmented contours continue to evolve, new approaches and algorithms will emerge, and it is critical for these to be evaluated by metrics that reflect clinically meaningful outcomes. Use of AI-AS in clinical practice will likely lead to significant benefits to RT planning workflow and resources. Prospective evaluation of the OrganRT™ AI-AS software in a multi-center workflow study will be included in future studies, which will help define the impact of incorporating machine learning into clinical practice.

7. Acknowledgement:

Thanks Siemens for providing the software installation and setting up the server and configuration.

References

1. Jeon W, Wu HG, Song SH, Kim JI. Radial displacement of clinical target volume in node negative head and neck cancer. Radiat Oncol J. 2012; 30: 36-42.
2. Harari PM, Song S, Tomé WA. Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2010; 77: 950-958
3. Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol. 2012; 7: 32
4. Yang J, Wei C, Zhang L, Zhang Y, Blum RS, Dong L. A statistical

modeling approach for evaluating auto-segmentation methods for image-guided radiotherapy. Comput Med Imaging Graph. 2012; 36: 492-500

5. Reed VK, Woodward WA, Zhang L, et al. Automatic segmentation of whole breast using atlas approach and deformable image registration. Int J Radiat Oncol Biol Phys. 2009; 73: 1493-1500.
6. Nelson TCF, Wai MH, Chun KS, Michael CHL, Wai TN. Automatic segmentation for adaptive planning in nasopharyngeal carcinoma IMRT: time,geometrical, and dosimetric analysis. Med Dosim. 2019.
7. Chao KSC, Bhide S, Chen H, et al. Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach. Int J Radiat Oncol Biol Phys. 2007; 68: 1512-1521
8. Chen A, Deeley MA, Niermann KJ, Moretti L, Dawant BM. Combining registration and active shape models for the automatic segmentation of the lymph node regions in head and neck CT images. Med Phys. 2010; 37: 6338-6346.
9. Stapleford LJ, Lawson JD, Perkins C, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2010; 77: 959-966.
10. Tim L, Johan VS, Mark G, Devis P, Paul A, Judith VDS, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol. 2018; 126: 312–7.
11. Sang HA, Adam UY, Kwang HK, Chankyu K, Youngmoon G, Shinhaeng C, et al. Comparative clinical evaluation of atlas and deep-learning-based auto-segmentation of organ structures in liver cancer. Radiat Oncol. 2019.
12. Wen C, Yimin L, Brandon AD, Xue F, Shyam R, Stanley HB, et al. Deep learning vs. atlas-based models for fast auto-segmentation of the masticatory muscles on head and neck CT images. Radiat Oncol. 2020.
13. Lustberg T, van Soest J, Gooding M, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol. 2018; 126(2): 312–7.