Advances in Radiation Oncology (2023) 8, 101177

advances in radiation oncology

www.advancesradonc.org

Scientific Article

Clinical Validation of Siemens' Syngo.via Automatic Contouring System

Óscar Pera, MSc,[a,b,*] Álvaro Martínez, MD,[a] Christian Möhler, MSc,[c] Bob Hamans, MSc,[c] Fernando Vega, MSc,[c] Fernando Barral, MSc,[c] Nuria Becerra, AS,[a] Rafael Jimenez, AS,[a] Enric Fernandez-Velilla, PhD,[a,b] Jaume Quera, MSc,[a,b] and Manuel Algara, PhD[a,b,d]

[a]*Radiation Oncology Department, Hospital del Mar, Barcelona, Spain;* [b]*Institut Hospital del Mar d'Investigacions Mèdiques, Barcelona, Spain;* [c]*Siemens Healthineers, Erlangen, Germany; and* [d]*Autonomous University of Barcelona, Barcelona, Spain*

Received 28 July 2022; accepted 5 January 2023

Abstract

Purpose: The manual delineation of organs at risk is a process that requires a great deal of time both for the technician and for the physician. Availability of validated software tools assisted by artificial intelligence would be of great benefit, as it would significantly improve the radiation therapy workflow, reducing the time required for segmentation. The purpose of this article is to validate the deep learning–based autocontouring solution integrated in syngo.via RT Image Suite VB40 (Siemens Healthineers, Forchheim, Germany).

Methods and Materials: For this purpose, we have used our own specific qualitative classification system, RANK, to evaluate more than 600 contours corresponding to 18 different automatically delineated organs at risk. Computed tomography data sets of 95 different patients were included: 30 patients with lung, 30 patients with breast, and 35 male patients with pelvic cancer. The automatically generated structures were reviewed in the Eclipse Contouring module independently by 3 observers: an expert physician, an expert technician, and a junior physician.

Results: There is a statistically significant difference between the Dice coefficient associated with RANK 4 compared with the coefficient associated with RANKs 2 and 3 ($P < .001$). In total, 64% of the evaluated structures received the maximum score, 4. Only 1% of the structures were classified with the lowest score, 1. The time savings for breast, thorax, and pelvis were 87.6%, 93.5%, and 82.2%, respectively.

Conclusions: Siemens' syngo.via RT Image Suite offers good autocontouring results and significant time savings.

Introduction

The development of a radiation therapy (RT) plan requires images of the tumor and the normal structures surrounding it, and it must be possible to delineate both the tumor (gross tumor volume) and the organs at risk (OAR). These delineations are made manually by the RT oncologist or technician, with the precise segmentation of the gross tumor volumes and OARs forming the basis for good RT. To be able to increase reproducibility and

Sources of support: Siemens provided the software and hardware necessary for the operation of syngo.via, as well as proofreading and translation services.

Disclosures: This research was conducted in collaboration with Siemens, but it were not involved in any way during the clinical phase of the study or in the data collection. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Research data are available upon request from the corresponding author. The raw data were generated in PSMar. Because of commercial restrictions, model data supporting the conclusions will not be available.

*Corresponding author: Óscar Pera, MSc; E-mail: opera@psmar.cat

https://doi.org/10.1016/j.adro.2023.101177

reduce uncertainty and errors, international or institutional guidelines, contouring atlases, case libraries, and numerous recommendations exist.[1-3] The manual delineation of critical organs is a process that requires a great deal of time both for the technician and for the physician. Having validated software tools assisted by artificial intelligence would be of great benefit, as these tools would significantly improve the RT workflow, reducing the time required for segmentation.[4-6] These tools also may contribute toward reducing interobserver variability, standardizing the critical organs delineated in different clinical pathologies, and reducing uncertainty when comparing results between them.[7]

Syngo.via RT Image Suite is the dedicated multimodality imaging software solution for RT offered by Siemens Healthineers (Forchheim, Germany) including automatic contouring of organs required for RT planning. The automatic contouring algorithm is based on deep learning, and it has been trained and validated with patient images from multiple institutions across the world.

To use the contours generated automatically by a software program for assistance, they must be validated. The Dice index is often used to perform this type of validation. Also known as similarity coefficient, the Dice index is a statistic widely used in different specialties[8] to gauge the similarity of 2 samples. However, it has limitations, such as the effect of the size of the structure and clinical interpretability, because it does not provide local information. A qualitative classification system with a clear and straightforward clinical interpretation may be more suitable.[8−10]

In this work, we present the validation of 600 contours corresponding to 18 different OARs delineated automatically by syngo.via RT Image Suite VB40. For this purpose, we have used a specific qualitative classification system, RANK.

Methods and Materials

Algorithm description

The autocontouring algorithm, developed by Siemens Healthineers for syngo.via RT Image Suite, computes segmentation masks of an organ or anatomic structure for a given computed tomography (CT) data set.[11] The segmentation is performed on the region of interest of individual organs instead of on the entire image volume. To locate the regions of interest, anatomic landmarks (including vessel bifurcations, bony structures, organ center, and boundary points) are detected in the input image based on Ghesu et al.[12]

A Deep Image-to-Image Network (DI2IN) is employed for organ segmentation according to the technique described by Yang et al.[13] It consists of a convolutional encoder−decoder architecture combined with multilevel feature concatenation. During the training process, the network is driven by a cross-entropy loss based on a learning rate of 0.001 using the Adam optimization. An adversarial network is selectively used to regularize the training process of DI2IN by discriminating the output of DI2IN from the ground truth in a patch-by-patch manner using binary cross-entropy.

Radiation therapy images from multiple institutions across Europe, Asia, and North and South America were used for training and validation. Because the application only segments normal anatomy, no ethnicity-based dependency is expected. Depending on the segmented organ, the number of annotated patients in the training set ranges from few hundreds to many thousands. The algorithm is trained with a fixed number of cases during the process of product development; it does not learn from new cases after product release and cannot be retrained in the hospital itself.

The training, validation, and testing data were annotated based on predefined clinical annotation protocols. These protocols were defined based on guidelines from the Radiation Therapy Oncology Group.[14]

A team of experienced radiologists and a radiation oncologist mentored and supervised the entire process. For each organ, a detailed protocol was established based on the aforementioned international guidelines. A quality assessment based on peer review was performed for each data set before further use in model training and validation.

Patient selection

Computed tomography images from 105 different patients were included: 30 patients with lung, 30 patients with breast, and 35 patients with prostate cancer. All the CT scans were performed with a 64-slice, single-source, flat-table SOMATOM Confidence scanner (Siemens Healthineers). The characteristics of the CT scanning protocols are described in the Appendix. The CT scans of patients with lung cancer were acquired in the supine position with the stereotactic body RT Conchest Supra (Lorca Marin S.A., Murcia, Spain) immobilization system and respiratory control using the RGSC (Respiratory Gating for Scanners) system (Varian Medical Systems, Palo Alto, CA). For the contouring of OARs, the “Average” series was used. Patients with breast cancer were positioned with their arms abducted and were immobilized with the same system as those with lung cancer. The CT scan was performed with deep breaths and with respiratory control for all left breasts, as well as for right breasts requiring radiation of the ganglia regions, provided the patient could complete it. In total, there were 15 patients with the right breast affected (3 breath hold and 12 free breathing) and 15 patients with the left breast affected (12 breath hold and 3 free breathing). All CT scans were performed with the patient in the supine position.

Workflow

The CT images were sent from our CT scanner to syngo.via RT Image Suite VB40 software and ARIA (Varian Medical Systems). The structures were manually segmented in ARIA following the usual clinical workflow established in our service: an expert technician segments the structures and then a radiation oncologist reviews and validates them. In parallel, all studies were automatically contoured with the syngo.via RT Image Suite VB40 software, which is based on a server-client system.

The autocontoured organs that were evaluated in this study are shown in Table 1. The heart was not taken into consideration, because our own protocol has changed in relation to the craniocaudal limits and presented discrepancies with the guidelines used by Siemens Healthineers. The bilateral organs (lung and breast) were delineated separately. The body structure corresponds to the patient's external contour.

Once the CT had been performed and the study concluded, the images were sent to syngo.via, which automatically contoured the structures assigned in the model. Approximately 2 minutes later, the images together with the set of structures were filed in a network folder without having to open syngo.via. The autocontoured structures from syngo.via were then reviewed in the Eclipse Contouring module independently by 3 observers: an expert physician, an expert technician, and a junior physician (completely independent of the clinical course of the patient and their manually contoured structures). The evaluation was done based on a categorical system of 4 values (Table 2). The specifics of the respective RT treatment were taken into account. In patients with prostate cancer, for example, the upper parts of the abdominal cavity were not considered. In cases, for which the contouring of an additional organ was required for treatment, this was performed manually by the technician or physician responsible. Subsequently, all required structures were reviewed and edited. The responsible physician then delineated the planning target volume, checked the OARs, and approved all the structures. Before the dosimetric planning, all the patients underwent a clinical session in which a multidisciplinary group consisting of physicians, physicists, and technicians rechecked the correct delineation of the structures. Images from other modalities such as magnetic resonance imaging or positron emission tomography were not required in the process.

Table 1 Contoured organs

Computed tomography type	Organ
Lung	Lung right Lung left Aorta Esophagus Spinal cord Body
Breast	Breast right Breast left Lung right Lung left Esophagus Spinal cord Body
Pelvis	Bladder Rectum Proximal femur right Proximal femur left Prostate Abdominopelvic cavity Body

Evaluations and measurements

The Dice coefficient was measured by means of the Varian Eclipse 16.1 software for each of the autosegmented organs, considering manual segmentation as ground truth. Each segmented structure, together with its corresponding Dice value, was assigned to a group based on its RANK value. The RANK 1 category was excluded for this part of the analysis because it corresponds to segmentations that could not be used and had to be completely redone manually.

The time invested in delineating the OARs was self-timed, after opening the CT series, manually for 19 cases (6 breast, 6 thorax, and 7 pelvis) in the Eclipse "Contouring" module. The contouring time of OARs not included in the study has not been considered. Similarly, the time taken by the expert RT technician to review and correct the autocontoured OARs was recorded. With Excel 365 (Microsoft, Redmond, WA), a *t* test analysis was performed to investigate the correlation between the 4-level categorical system and the Dice coefficient.

Table 2 Categories of the specific qualitative evaluation system, RANK

RANK 4	Clinically usable for RT planning	May contain very minor errors that do not require any type of correction
RANK 3	Usable after minor edits	May be used but requires minor corrections on a maximum of 1-3 slices
RANK 2	Usable after major edits	May be used but requires changes to at least 4 slices
RANK 1	Must redo	Must be rejected and redone manually

Abbreviation: RT = radiation therapy.

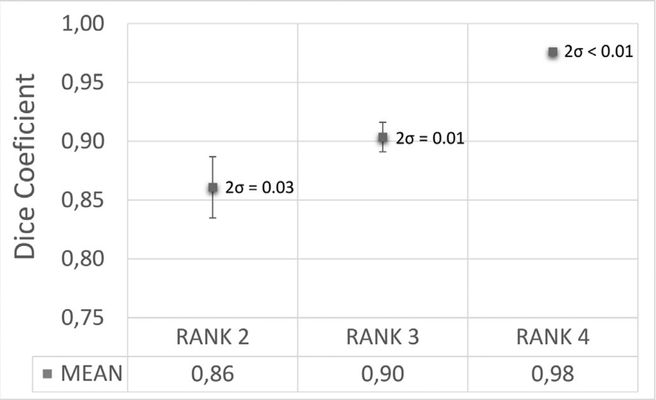

Figure 1 Dice values depending on the associated RANK.

Results

Dice coefficient versus RANK

Figure 1 shows the results of the Dice coefficient obtained for the various RANK values.

There was a statistically significant difference between the Dice coefficient values associated with RANK 4 compared with the coefficient associated with RANKs 2 and 3 ($P < .001$).

The Dice coefficient associated with RANK 4, 3, and 2 was 0.98, 0.90, and 0.86, respectively, with a 95% confidence level. The average Dice coefficient across the entire data set was between 0.95. In total, 95% of the autosegmentations obtained a Dice coefficient greater than 0.9 and the 100% was greater than the common benchmark of 0.7.

RANK by OAR

Figure 2 shows the distribution of all the 600 organs segmented automatically based on their RANK. In total, 64% received the maximum score, 4. Only 1% of the structures were classified with the lowest score, 1. The structures obtaining a generally greater score (90% of

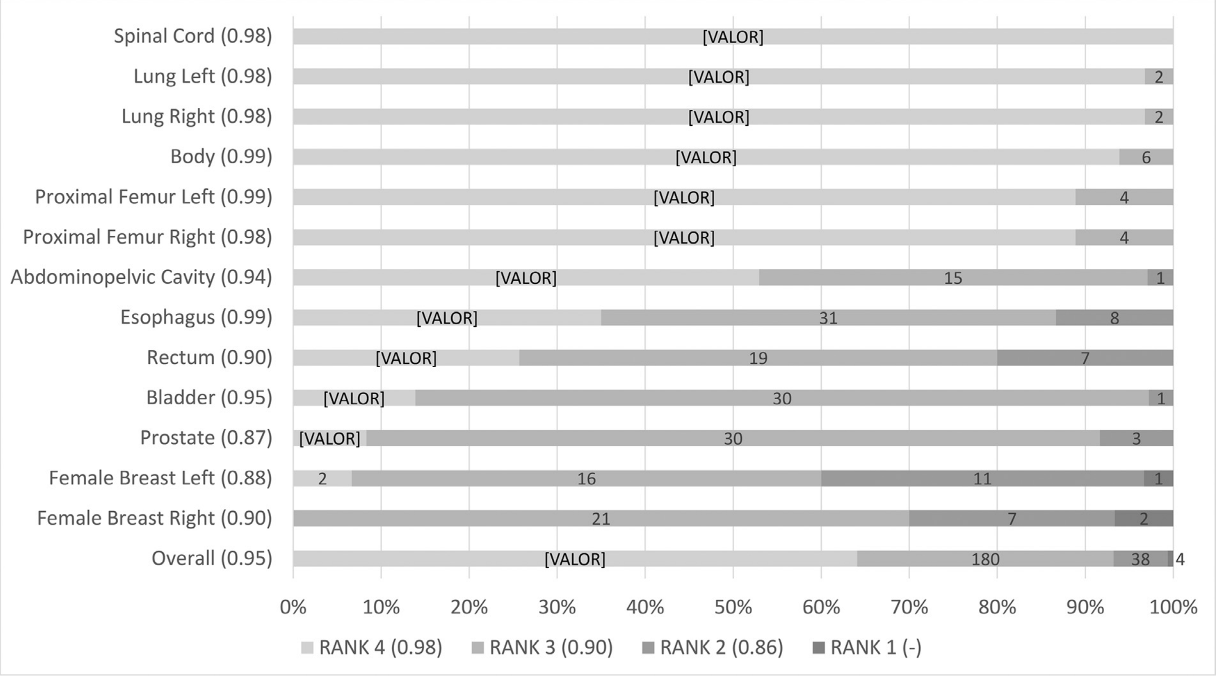

Figure 2 Distribution of the RANK scores obtained in all the structures and in each one separately. The absolute values are indicated in the bars.

Table 3 Time invested in contouring all the structures associated with breast, thorax, and pelvis treatments

	Breast	Thorax	Pelvis
Manual contouring, min:s (Eclipse)	32:44 ± 09:05	19:35 ± 10:30	34:40 ± 11:30
Corrections, min:s (syngo.via)	03:44 ± 02:36	01:16 ± 02:48	06:11 ± 04:18
Average time saved, min:s	29:00 ± 10:12	18:19 ± 10:54	28:29 ± 12:18

Average ± 2σ in minutes and seconds.

cases rated 3 or 4) were femur right, proximal femur left, body, spinal cord, prostate, bladder, abdominopelvic cavity, lung right and lung left, whereas those obtaining a lower score were breast right and breast left, scoring 1 or 2 in 30% and 40% of cases, respectively. Esophagus and rectum obtained a score of 3 or 4 in 80% and 85% of cases, respectively.

Time invested

The average time invested in checking and correcting the autocontoured structures is a fraction of the time dedicated to manual contouring in Eclipse (Table 3). The time savings for breast, thorax, and pelvis was 88.6%, 93.5%, and 82.2%, and the number of volumes requiring correction was 2.5, 0.8, and 2.9 in average, respectively. Table 3 does not show the cases in which the RANK1 event occurs as the autosegmented structure could not be used in these cases. These cases accounted for only 0.67% (4 of 600) of the total.

Discussion

Current techniques such as intensity modulated RT or volumetric modulated arc therapy require the contouring of various critical OARs for treatment planning. The definition and reliable segmentation of the normal tissues also have significant clinical and dosimetric consequences. For OARs that limit the dose in particular, the more information is provided to the system, the better the result will be. However, if performed manually, this process requires a significant amount of time and effort of the expert clinician. Autocontouring software plays an important role in the future of RT and will be an essential tool to manage the increasing patient volumes in cancer.

Artificial intelligence is demonstrating its importance with automatic contouring tools and autoplanning systems.[15-18] New technologies will appear in the short term[19] that will change the way of working, as happened with the appearance of 3-dimensional RT.

The atlas-based automatic contouring systems[20] significantly improve precision and reduce variability, which is further improved by systems based on artificial intelligence. Both can reduce the time taken and improve the workflow. Accordingly, we have verified the clinical applicability of syngo.via RT Image Suite in terms of the time invested and the clinical benefit of deep learning−based autocontouring.

Comparisons between contouring systems by means of geometric overlap metrics may not lead to meaningful results. Therefore, as proposed by Brouwer et al,[21] we decided to evaluate the manual adjustments made to the automatic contours to be able to establish the applicability of the system.

The performance of the autosegmentation software may be evaluated by the contour similarity indices as well as possible time savings. Different parameters exist, such as the Jaccard index, the Dice coefficient, or the Hausdorff distance,[6-10] which are used to compare the results of the tested delineation to a ground-truth delineation. However, aiming for a complete overlap between 2 structures contoured manually by 2 observers or by 1 observer and the software may not lead to any insights, as there will always be a certain degree of variability, which is inevitable in everyday clinical practice. Moreover, given that artificial intelligence is still at the developmental stage, the contours generated by computer tools must always be subject to revision and manual correction by the radiation oncologist.[22] For that reason, we decided to perform a clinical applicability analysis, based on the possible changes that the technician would make to the automatically generated contour, and we performed a usability analysis.

Walker et al[23] published the results of a randomized double-blind study in which the manually drawn contours were compared with the automatically created contours in 40 patients with head and neck cancer. The automatic system allowed a time savings of 30%, with a similar Dice coefficient between manual and autosegmented series. However, in their conclusions, the authors affirm that autocontouring must be validated by a physician. Choi et al[24] compared the contours that were drawn manually by an expert to those generated automatically by 2 algorithms, one based on atlas and the other based on deep learning in patients with breast cancer. They concluded that the algorithm based on deep learning is more reliable than the one based on atlas. Zabel et al[25] performed a similar study on the bladder, prostate, and rectum contours, reaching the same conclusion; similar results have been demonstrated by Van Dijk et al.[26] The advantage of deep learning methods was also shown for delineations of greater complexity such as the parotid glands[27] and masticatory muscles,[28] as well as in the delineation of the hippocampus on magnetic resonance images.[29]

In the present study, we quantified the time saved with autosegmentation. It is important to measure the total time for all parts of the process and to include the time invested in the definition of the atlas cases and in the choice of the correct atlas case (manual or automatic).

Several benefits are expected from the use of autocontouring systems. However, they must be implemented and carefully validated in everyday clinical practice.

To conclude, autosegmentation tools provide an attractive opportunity with the potential both to reduce segmentation time and to increase adherence to existing guidelines. By generating contours that require minimal editing time, these systems can be easily incorporated into the routine workflows of a radiation oncology department.

Conclusion

Siemens Healthineers' syngo.via RT Image Suite offers good autocontouring results and may be incorporated into the clinical practice of a radiation oncology department with significant time savings in the delineation of the critical organs and, therefore, in the total preparation time for RT.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.adro.2023.101177.

References

1. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiother Oncol*. 2016;121:169-179.
2. Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. *J Med Imaging Radiat Oncol*. 2016;60:393-406.
3. Wright JL, Yom SS, Awan MJ, et al. Standardizing normal tissue contouring for radiation therapy treatment planning: An ASTRO Consensus Paper. *Pract Radiat Oncol*. 2019;9:65-72.
4. Boldrini L, Bibault J, Masciocchi C, Shen Y, Bittner M. Deep learning: A review for the radiation oncologist. *Front Oncol*. 2019;9:977.
5. Giraud P, Giraud P, Gasnier A, et al. Radiomics and machine learning for radiotherapy in head and neck cancers. *Front Oncol*. 2019;9:174.
6. Oktay O, Nanavati J, Schwaighofer A, et al. Evaluation of deep learning to augment image-guided radiotherapy for head and neck and prostate cancers. *JAMA Netw Open*. 2020;3:e2027426.
7. Men K, Geng H, Biswas T, Liao Z, Xiao Y. Automated quality assurance of OAR contouring for lung cancer based on segmentation with deep active learning. *Front Oncol*. 2020;10:986.
8. OpenReview.net. Common limitations of performance metrics in biomedical image analysis. Available at: https://openreview.net/forum?id=76X9Mthzv4X. Accessed February 10, 2023.
9. Liu C, Gardner SJ, Wen N, et al. Automatic segmentation of the prostate on CT images using deep neural networks (DNN). *Int J Radiat Oncol Biol Phys*. 2019;104:924-932.
10. Cardenas CE, McCarroll RE, Court LE, et al. Deep learning algorithm for auto-delineation of high-risk oropharyngeal clinical target volumes with built-in Dice similarity coefficient parameter optimization function. *Int J Radiat Oncol Biol Phys*. 2018;101:468-478.
11. Marschner S, Datar M, Gaasch A, et al. A deep image-to-image network organ segmentation algorithm for radiation treatment planning: Principles and evaluation. *Radiat Oncol*. 2022;17:129.
12. Ghesu F-C, Georgescu B, Zheng Y, et al. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. *IEEE Trans Pattern Anal Mach Intell*. 2019;41:176-189.
13. Yang D, Xu D, Zhou K, et al. Automatic liver segmentation using an adversarial image-to-image network. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, eds. *Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017. Lecture Notes in Computer Science, vol 10435*. Cham: Springer; 2017.
14. NRG Oncology. Radiation Therapy Oncology Group (RTOG) Atlas. Available at: https://www.nrgoncology.org/About-Us/Center-for-Innovation-in-Radiation-Oncology. Accessed November 11, 2022.
15. Brouwer CL, Steenbakkers RJ, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiother Oncol*. 2015;117:83-90.
16. Kaderka R, Hild SJ, Bry VN, et al. Wide-scale clinical implementation of knowledge-based planning: An investigation of workforce efficiency, need for post-automation refinement, and data-driven model maintenance. *Int J Radiat Oncol Biol Phys*. 2021;111:705-715.
17. Chen X, Men K, Li Y, Yi J, Dai J. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning. *Med Phys*. 2019;46:56-64.
18. Nguyen D, Jia X, Sher D, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-Net deep learning architecture. *Phys Med Biol*. 2019;64:065020.
19. Poortmans PM, Takanen S, Marta GN, Meattini I, Kaidar-Person O. Winter is over: The use of artificial intelligence to individualise radiation therapy for breast cancer. *Breast*. 2020;49:194-200.
20. Casati M, Piffer S, Calusi S, et al. Methodological approach to create an atlas using a commercial auto-contouring software. *J Appl Clin Med Phys*. 2020;21:219-230.
21. Brouwer CL, Boukerroui D, Oliveira J, et al. Assessment of manual adjustment performed in clinical practice following deep learning contouring for head and neck organs at risk in radiotherapy. *Phys Imaging Radiat Oncol*. 2020;16:54-60.
22. McCarroll RE, Beadle BM, Balter PA, et al. Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: A step toward automated radiation treatment planning for low- and middle-income countries. *J Glob Oncol*. 2018;4:1-11.
23. Walker GV, Awan M, Tao R, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. *Radiother Oncol*. 2014;112:321-325.
24. Choi MS, Choi BS, Chung SY, et al. Clinical evaluation of atlas- and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. *Radiother Oncol*. 2020;153:139-145.
25. Zabel WJ, Conway JL, Gladwish A, et al. Clinical evaluation of deep learning and atlas-based auto-contouring of bladder and rectum for prostate radiation therapy. *Pract Radiat Oncol*. 2021;11:e80-e89.

26. van Dijk LV, Van den Bosch L, Aljabar P, et al. Improving automatic delineation for head and neck organs at risk by deep learning contouring. *Radiother Oncol*. 2020;142:115-123.
27. Hänsch A, Schwier M, Gass T, et al. Evaluation of deep learning methods for parotid gland segmentation from CT images. *J Med Imaging*. 2018;6:011005.
28. Chen W, Li Y, Dyer BA, 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;15:176.
29. Pan K, Zhao L, Gu S, et al. Deep learning-based automatic delineation of the hippocampus by MRI: Geometric and dosimetric evaluation. *Radiat Oncol*. 2021;16:12.