Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

The Dosimetric Impact of Deep Learning-Based Auto-Segmentation of Organs at Risk on Nasopharyngeal and Rectal Cancer

Hongbo Guo
Fudan University Shanghai Cancer Center
Jiazhou Wang
Fudan University Shanghai Cancer Center
Xiang Xia
Fudan University Shanghai Cancer Center
Yang Zhong
Fudan University Shanghai Cancer Center
Jiayuan Peng
Fudan University Shanghai Cancer Center
Zhen Zhang
Fudan University Shanghai Cancer Center
Weigang Hu (jackhuwg@gmail.com)
Fudan University Shanghai Cancer Center

Research Article

Keywords: Treatment planning, Dosimetric, Deep learning, Auto-segmentation

Posted Date: March 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-328649/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

The dosimetric impact of deep learning-based auto-segmentation of organs at risk on nasopharyngeal and rectal cancer

Abstract

Purpose: To investigate the dosimetric impact of deep learning-based auto-segmentation of organs at risk (OARs) on nasopharyngeal and rectal cancer.

Methods and Materials: Twenty patients, including ten nasopharyngeal carcinoma (NPC) patients and ten rectal cancer patients, who received radiotherapy in our department were enrolled in this study. Two deep learning-based auto-segmentation systems, including an in-house developed system (FD) and a commercial product (UIH), were used to generate two auto-segmented OARs sets (OAR_FD and OAR_UIH). Treatment plans based on auto-segmented OARs and following our clinical requirements were generated for each patient on each OARs set (Plan_FD and Plan_UIH). Geometric metrics (Hausdorff distance (HD), mean distance to agreement (MDA), the Dice similarity coefficient (DICE) and the Jaccard index) were calculated for geometric evaluation. The dosimetric impact was evaluated by comparing Plan_FD and Plan_UIH to original clinically approved plans (Plan_Manual) with dose-volume indices and 3D gamma pass rates. Spearman's correlation analysis was performed to investigate the correlation between dosimetric deviation and geometric metrics.

Results: FD and UIH could provide similar geometric performance in parotids, temporal lobes, lens, and eyes (DICE, $p > 0.05$). OAR_FD had better geometric performance in the optic nerves, oral cavity, larynx, and femoral heads (DICE, $p < 0.05$). OAR_UIH had better geometric performance in the bladder (DICE, $p < 0.05$). In dosimetric analysis, both Plan_FD and

Plan_UIH had nonsignificant dosimetric differences compared to Plan_Manual for most PTV and OARs dose-volume indices. The only significant dosimetric difference was the D_{max} of the left temporal lobe for Plan_FD vs. Plan_Manual (p = 0.05). Only one significant correlation was found between the mean dose of the femoral head and its HD index (R = 0.4, p = 0.01).

Conclusions: Deep learning-based OARs auto-segmentation for NPC and rectal cancer has a nonsignificant impact on most PTV and OARs dose-volume indices. Correlations between the auto-segmentation geometric metric and dosimetric difference were not observed for most OARs.

1. Introduction

Organs at risk (OARs) delineation is a critical task in radiotherapy. It affects many aspects of treatment planning, which can further affect the probability of local tumor control and normal tissue complications [1−4]. However, manual OARs delineation is time-consuming and tedious work. This fact is especially true for cancers with complex anatomy, such as nasopharyngeal carcinoma (NPC).

Auto-segmentation can reduce the work intensity of oncologists and improve work efficiency [5−10]. Recently, deep learning-based auto-segmentation has become a mainstream assistance segmentation technique provided by many software vendors [7,11−13]. The latest relevant studies have shown promising results for these systems, improving consistency among oncologists and shortening the delineation time [14−16].

As an emerging technique, sufficient clinical application assessment is required. Although many studies have evaluated the performance of auto-segmentation in terms of geometric

metrics [7,14–15,17–19], few studies have focused on dosimetric impact [11,20]. Because OARs delineation will have an impact on treatment planning optimization and evaluation, geometric metric evaluation is not sufficient for clinical application.

For dosimetric impact, Van Dijk et al. [11] compared the dosimetric difference between auto-segmented and manually delineated OARs with a clinically approved treatment plan. The results proved that more accurate auto-segmentation translated into smaller dosimetric differences compared to the manual contours. Kaderka et al. [20] used an atlas-based method for cardiac substructure segmentation and proved that the quality of auto-segmented contours cannot be determined by geometric metrics only, and geometrical measures did not predict the accuracy of dosimetric parameters. These two studies used clinically approved treatment plans based on manual delineation and assessed on auto-segmented contours.

Because OARs delineation has an impact on treatment plan optimization. Different delineations will generate different plans. A straighter method to assess dosimetric impact should reoptimize the treatment plan based on auto-segmented contours and then use manual contours to evaluate the dosimetric differences between the reoptimized plans and the original clinical treatment plans.

To further assess the dosimetric impact of deep learning-based auto-segmentation, we have designed a dosimetric comparison study. Two sites, including the nasopharynx and rectum, and two deep learning-based auto-segmentation systems, including a commercial tool from United Imaging Healthcare (UIH, Shanghai, China) and an in-house auto-segmentation tool developed by our institution, were investigated. To evaluate the application of deep learning-based auto-segmentation in clinical situations, the whole planning process was

following our clinical routine requirement. Meanwhile, the correlation between geometric metric and dosimetric performance was investigated.

2. Methods

A schematic workflow of this study is presented in **Fig. 1**. After auto-segmentation, the assessment was divided into three parts. First, the accuracy of auto-segmentation was evaluated based on geometric metrics. Second, we reoptimized the plan based on the auto-segmented OARs and compared it with the original treatment plan to evaluate the dosimetric differences. Third, we explored the correlation between the geometric metrics and dosimetric differences.

2.1 Patients and treatment protocol

Two sites, including the nasopharynx and rectum, were investigated. Ten patients for each site who received radiotherapy at XXXX between 2017 and 2019 were randomly selected from our database and enrolled in this study. The details of the patient characteristics are shown in **Additional file 1: Supplement A, Table S1**. For NPC patients, the prescription was 70.4 Gy in 32 fractions for T3-T4 stage patients and 66 Gy in 30 fractions for T1-T2 stage patients. For rectal cancer, all of the patients received 50 Gy in 25 fractions.

2.2 OARs manual delineation

Manual delineation was performed on the Pinnacle (Pinnacle, v9.10, Philips Corp, Fitchburg, WI, USA) treatment planning system. The targets and OARs are presented in **Table**

1. These contours were delineated by radiation oncologists with more than 5 years of experience in radiation oncology and revised and approved by senior radiation oncologists. All of the manually delineated OARs were used for patient treatment.

2.3 Deep learning-based auto-segmentation

Two deep learning-based auto-segmentation systems were used in this study. FD is an in-house deep learning-based auto-segmentation system, and the details of the network and model training have been presented in our recent studies [XXXX]. Briefly, we used approximately 200 NPC and 200 rectal cancer cases from our institution as the training dataset. The delineation of the training dataset came from clinical routine without modification for this task. The network was a modified 2D U-Net. It was used in 2018 for clinical testing on NPC and rectal OARs auto-segmentation. The OARs segmented by this system were marked as OAR_FD.

UIH is a commercial treatment planning system developed by UIH Corporation [21-22]. It uses a two-phase 3D U-Net for OARs location and segmentation. The training data did not come from our institution. We used UIH from 2019 for clinical testing. This system provided NPC and rectal cancer OARs auto-segmentation, which was used in this study. The OARs segmented by this system were marked as OAR_UIH.

2.4 Treatment planning

Pinnacle (Pinnacle, v9.10, Philips Corp, Fitchburg, WI, USA) and Varian Trilogy Linac (Varian, Polo Alto, CA, USA) with 120 multileaf collimator were used for treatment planning for

all plans. All of the treatment planning processes were the same as our clinical routine for consistency.

The NPC clinical treatment plans used the 9-field static intensity modulated radiotherapy (sIMRT) technique, and the gantry angles were 0°, 45°, 85°, 120°, 160°, 200°, 245°, 275°, and 315°. The field could be split based on field width. The maximum number of segmented subfields was set to 55. The rectal cancer clinical treatment plans adopted 7 fields of the sIMRT technique. The beam angles were chosen based on clinical experience. Here, we mainly considered having the bladder and femoral heads receive less radiation exposure. The maximum number of segmented subfields was set to 35. For all of the plans, the minimum subfield area was set to 10 cm^2, and the minimum subfield monitoring unit was set to 10 MU.

The prescription was normalized to the mean dose of PTV as in our clinical routine. For NPC, we prescribe 220 cGy per fraction to 97% of the PTV70.4 mean dose for 32 fractions or 220 cGy per fraction to 97% of the PTV66 mean dose for 30 fractions. For rectal cancer, we prescribe 200 cGy per fraction to 96% of PTV mean dose for 25 fractions. In this setting, the D_{95} of PTV was close to the prescription dose. All of the treatment plans were completed by medical physicists with more than 3 years of experience.

Each patient had three plans: Plan_Manual, Plan_FD and Plan_UIH. Plan_Manual was a clinically approved plan that was used for patient treatment. Plan_FD and Plan_UIH were reoptimized based on manually delineated PTVs and auto-segmented OARs. For OARs that were not generated by the auto-segmentation system (temporomandibular joints and chiasm), we used manually delineated OARs to replace them. The beam angles and initial optimization parameters for the reoptimization plan (Plan_FD and Plan_UIH) were consistent with

Plan_Manual. The physicist could adjust the optimization objective based on his or her experience and judgment, the same as the routine clinical treatment planning process.

2.5 Geometric evaluation

Manual delineated contours were used as references. The performance of auto-segmentation was evaluated by the following four geometric metrics: Hausdorff distance (HD), mean distance to agreement (MDA), Dice similarity coefficient (DICE), and Jaccard index [23-25]. HD and MDA were used to quantify the maximum and mean 3D distances between contours A and B, respectively. DICE and the Jaccard index were measures of the overlap between contours A and B. The definitions are as follows:

$$\mathrm{HD}\ (\mathrm{A},\mathrm{B}) = max\{H\ (A,B),\quad H(B,A)\}$$

$$H\ (A,B) = \underset{a\epsilon A}{max}\{\underset{b\epsilon B}{min}\{d\ (a,b)\}\}$$

where *d* (a, b) represents the 3D Hausdorff distance between point a from contour A and point b from contour B.

$$\mathrm{MDA}\ (\mathrm{A},\mathrm{B}) = \frac{h\ (A,B) + h\ (B,A)}{2}$$

$$h\ (A,B) = \underset{a\epsilon A}{mean}\{\underset{b\epsilon B}{min}\{d\ (a,b)\}\}$$

$$\mathrm{DICE} = 2 * \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A}| + |\mathrm{B}|}$$

$$\mathrm{Jaccard} = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|}$$

For a perfect overlap, the values of HD and MDA are 0, and the values of DICE and Jaccard are 1. For an imperfect overlap, the values of HD and MDA are large, and the values of DICE and Jaccard are close to 0.

2.6 Dosimetric evaluation

The dose-volume indices of PTV and OARs were extracted from Plan_FD, Plan_UIH and Plan_Manual. All of the dose-volume indices were calculated on manually delineated contours. For serial organs, we mainly focused on D_{max}. For parallel organs, we mainly focused on D_{mean}, V_{30} or V_{40} (**Table 1**).

A 3D gamma analysis was performed with 3% and 3 mm for a dose distribution comparison. The homogeneity index (HI) and conformity index (CI) for PTV were further calculated using the following formulas:

$$\mathrm{HI} = \frac{\mathrm{D_2} - \mathrm{D_{98}}}{\mathrm{D_p}}$$

$$\mathrm{CI} = \frac{\mathrm{V_R} * \mathrm{V_R}}{\mathrm{V_P} * \mathrm{V_{dose}}}$$

where D_P is the prescription dose, V_P and V_{dose} are the volume of PTV and the prescription dose region, respectively, and V_R is the intersection volume of V_P and V_{dose}.

2.7 Correlation between the geometric metric and dosimetric metric

The correlation between the geometric metric and the ΔDose was analyzed by Spearman's correlation test. The ΔDose is the dose deviation between reoptimized treatment plans (including Plan_FD and Plan_UIH) and Plan_Manual.

2.8 Statistical analysis

R software (v4.0) was used for statistical analysis. For a value comparison, the Shapiro-Wilk normality test was performed first. If a normal distribution was found, the paired-sample t test between groups was performed; otherwise, the Wilcoxon's paired-sample nonparametric

signed-rank test was performed. $p < 0.05$ indicates that the difference is statistically significant. The correlation between geometric metrics and dose difference indices was evaluated with Spearman's correlation coefficient R.

3. Results

3.1. Geometric evaluation

Fig. 2 shows the geometric evaluation results of auto-segmentation. Both deep learning systems can provide similar results in some OARs, including the parotids, temporal lobes, lens, and eyes (DICE, $p > 0.05$). For the brainstem and spinal cord, although there was a significant difference ($p < 0.05$), the deviation was small (less than 0.05 in DICE), while OAR_FD had better performance in the optic nerves, oral cavity, larynx, and femoral heads. OAR_UIH had better performance in the bladder. Representative rectal cancer and NPC examples of auto-segmentation are illustrated in **Fig. 3** and **Additional file 1: Supplement B, Fig. S1**. More examples are presented in **Additional file 1: Supplement D, Figs. S6-S12**.

3.2. PTV dosimetry evaluation

Table 2 lists the PTV dosimetric parameters of Plan_Manual, Plan_FD and Plan_UIH. No significant differences were found by comparison with Plan_Manual.

3.3. OARs dosimetry evaluation

Table 3 lists the OARs dosimetry parameters. No significant dosimetric differences were

found except for left temporal lobe D_{max} for Plan_FD vs. Plan_Manual (6376 ± 2126 cGy vs. 6444 ± 2156 cGy, p = 0.05). **Fig. 3** and **Additional file 1: Supplement B, Fig. S1** present an example of the dose distributions of Plan_Manual, Plan_FD and Plan_UIH for representative rectal cancer and NPC cases.

3.4. The correlation between the geometric metrics and dosimetric differences

No strong relationship was found between the ∆Dose metrics and the geometric metric DICE of all OARs (**Additional file 1: Supplement D, Figs. S2-S5**). The only significant correlation was found between the femoral head mean dose and geometric metric HD (R = 0.4, p = 0.01 for femoral head ∆mean dose vs. HD). However, the DICE vs. brainstem ∆max dose was significantly correlated. This outcome might have been a statistical random error since the trend is contrary to our expectations.

4. Discussion

In this study, we assessed the dosimetric impacts of deep learning-based OARs auto-segmentation on nasopharyngeal and rectal cancers. Our results showed that deep learning-based OARs auto-segmentation had no significant impact on the PTV dose distribution or most OARs dose-volume indices, while the correlation between the geometric metrics and OARs dosimetric differences was weak.

Two deep learning auto-segmentation systems were investigated. Both systems are under clinical testing in our institution. The clinical test for FD started in November 2018. Radiation oncologists can use this system for NPC and rectal OARs auto-segmentation in our institution.

These auto-segmented contours were usually reviewed and modified by radiation oncologists before clinical approval. This process has been applied on more than 500 patients. For UIH, we started testing it in March 2019. Similar to the FD system, radiation oncologists are required to review auto-segmented contours before clinical approval. The preliminary feedback of these two systems can reduce radiation oncologists' workload. More detailed data are being collected.

For quantitative geometric evaluation, both systems can provide similar performance for five OARs (eyes, parotids, lens, oral cavity and temporal lobes, $p > 0.05$ DICE). These OARs performance also similarly to those in other researches have reported [7,11]. Although the differences for the spinal cord and brainstem were significant, the deviation value was small (approximately 0.04 in DICE and <0.5 mm in MDA). Six OARs, including the bladder, femoral heads, spinal cord, brainstem, optic nerves and larynx, were significantly different between the two systems ($p < 0.05$, DICE). The reasons might be as follows.

FD can provide a better performance than UIH ($p < 0.05$ DICE) for the femoral heads, optic nerves, spinal cord, and larynx, which might be caused by the different OARs definitions between our clinical routine and UIH training data. For example, we did not include the femoral necks in femoral head segmentation. UIH included the femoral necks (**Fig. 3C**, red arrow). Additionally, OARs that do not have clear visible boundaries on CT images like temporal lobes can have large delineation variations (**Additional file 1: Supplement D, Figs. S10**). By retraining the auto-segmentation model on our institution data, these deviations could be eliminated.

The performance of the bladder for FD was worse than that for UIH ($p < 0.05$ DICE). This finding might have been caused by the algorithm difference between the two systems. Our

system used a 2D U-Net network, which could have some outliers, as our previous study demonstrated [XXXX]. UIH used a two-phase algorithm, which was more robust according to region location.

In dosimetric analysis, no difference was found for the PTV target ($p > 0.05$). The most significant dose difference was rectal PTV D_2 (Manual: 5349 ± 177 cGy, FD: 5384 ± 167 cGy, UIH : 5383 ± 160 cGy, $p = 0.08$). This outcome might have been caused by the interoperative difference between dosimetrists. For OARs dose-volume indices, the most significant dose difference was in the left temporal lobe D_{max} for Plan_FD vs. Plan_Manual (6376 ± 2126 cGy vs. 6444 ± 2156 cGy, $p = 0.05$). This finding might have been caused by the large variation in the delineation of the temporal lobes (**Additional file 1: Supplement D, Figs. S10**).

However, no significant dose-volume index difference was found for PTV and OARs. A plan dose distribution review remains necessary to fully investigate the dosimetric impact of an auto-segmentation system. The delineation could have a different impact on the final dose distribution. As we demonstrated in **Fig. 3C** and **Fig. 3G**, the femoral neck delineated by the UIH system was spared from 10% dose coverage. The low-dose isodose lines (10 Gy and 25 Gy) of Plan_UIH have different shapes compared to Plan_Manual and Plan_FD. In contrast, the difference between oral cavity delineation for UIH and manual delineation did not cause a significant dose distribution difference (**Additional file 1: Supplement B, Fig. S1. C** and **G**, red arrow).

This study showed that there was no clear monotonic relationship between the geometric metrics and dosimetric differences for most OARs. The only significant correlation was shown for the femoral head mean dose. There could be several reasons for this result. First, the

difference between manual and automatic delineation might be too small to cause a dosimetric difference beyond the random noise dose levels. In other words, the performance of our two auto-segmentation systems was “good enough”. When the delineation difference is sufficiently large, such as with the femoral head definition, the correlation between geometric metrics and dosimetric difference can still be observed. Second, the interoperator difference or intraoperator difference during treatment planning could cause a larger difference than auto-segmentation. These interoperator differences were difficult to avoid by the manual planning process. By automatic planning, these subjective deviations can be decreased. To analyze the impact on routine clinical practice, we did not implement it.

In this study, we used manually delineated contours as references. This fact does not mean that manual delineation is “better” or more “accurate” than deep learning-based delineation. In our ongoing evaluation study, radiation oncologists preferred auto-segmented contours over manual delineation for the parotids, optic nerves, lens and eyes. This phenomenon was also observed in [11]. Manual delineation represents a clinically acceptable and approved contour quality, which also implies some clinical experience or the habits of local institutions. Therefore, using a commercial auto-segmentation system that is not trained on local data requires more investigation.

For segmentation evaluation, geometric evaluation is a straightforward method for auto-segmentation performance. Many studies using these indices have been published in recent years [17,23-25]. Geometric metrics, such as DICE and MDA, are the critical indices for segmentation algorithm development. Using high-quality and consistent training or validation data, the algorithm performance can be quantified and compared. However, the clinical

assessment of auto-segmentation can be much more complicated and should be based on clinical purposes. A small improvement in geometric metrics, for example, DICE increase of 0.05, could represent substantial progress in the algorithm. However, its clinical value is likely to improve only marginally. A more practical assessment procedure should mimic clinical practice as much as possible. This principle is also consistent with some task-based evaluation procedures proposed by other studies [26-27].

The main limitation of this study was that it did not investigate interoperator variations. Using the auto-planning technique might reduce these variations, in turn increasing objectivity when plans are compared. These tasks were left for the future to complete.

5. Conclusion

Deep learning-based OARs auto-segmentation for NPC and rectal cancer might not have a significant impact on PTV and OARs doses. Correlations between the auto-segmentation geometric metric and dosimetric difference were not observed for most OARs. A dosimetric evaluation is recommended for applying auto-segmentation systems in the clinic.

Additional file

Additional file 1:

Supplement A. The details of the patient characteristics.

Supplement B. Representative nasopharyngeal carcinoma examples of auto-segmentation and dose distribution.

Supplement C. The correlation analysis between the geometric metrics and dosimetric differences.

Supplement D. More rectal cancer and nasopharyngeal carcinoma examples of auto-segmentation and dose distribution.

Abbreviations

OARs: Organs at risk; NPC: Nasopharyngeal carcinoma; sIMRT: Static intensity modulated radiotherapy; HD: Hausdorff distance; MDA: Mean distance to agreement; DICE: Dice similarity coefficient; HI: Homogeneity index; CI: Conformity index (CI).

Acknowledgements

Not applicable.

Authors' contributions

Study concept and design (WH, ZZ, JW, HG); acquisition of data (HG, XX, YZ, JP); analysis and interpretation of data (HG, JW); statistical analysis (HG, XX, YZ, JP); drafting of the manuscript (HG, JW, WH). All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Fudan University Shanghai Cancer Center Institutional Review Board and all methods were performed in accordance with the guidelines and regulations of this ethics board. Informed consent was obtained from all individual participants included in the study.

Consent for publication

Written informed consent for publication of their clinical details and/or clinical images was obtained from the patients.

Competing interests

The authors declare that they have no competing interests.

References

1. Brodin NP, Kabarriti R, Garg MK, et al. Systematic review of normal tissue complication models relevant to standard fractionation radiation therapy of the head and neck region published after the quantec reports. Int J Radiat Oncol Biol Phys 2018;100:391-407.

2. Liu Y, Lei Y, Wang Y, et al. Mri-based treatment planning for proton radiotherapy: Dosimetric validation of a deep learning-based liver synthetic ct generation

method. Phys Med Biol 2019;64:145015.

3. Wang J, Qing G, Ou X, et al. The impact of target dosimetry on patients' locoregional recurrence in nasopharyngeal carcinoma: A propensity score-matched analysis. Radiother Oncol 2019;141:67-71.

4. Ma JL, Hennessey DB, Newell BP, et al. Radiotherapy-related complications presenting to a urology department: A more common problem than previously thought? BJU Int 2018;121:28-32.

5. Kim N, Chang JS, Kim YB, et al. Atlas-based auto-segmentation for postoperative radiotherapy planning in endometrial and cervical cancers. Radiat Oncol 2020;15:106.

6. Cardenas CE, Yang J, Anderson BM, et al. Advances in auto-segmentation. Semin Radiat Oncol 2019;29:185-197.

7. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. Radiother Oncol 2020;144:152-158.

8. Ayyalusamy A, Vellaiyan S, Subramanian S, et al. Auto-segmentation of head and neck organs at risk in radiotherapy and its dependence on anatomic similarity. Radiat Oncol J 2019;37:134-142.

9. Elguindi S, Zelefsky MJ, Jiang J, et al. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. Phys Imaging Radiat Oncol 2019;12:80-86.

10. Savenije MHF, Maspero M, Sikkes GG, et al. Clinical implementation of mri-based

organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy. Radiat Oncol 2020;15:104.

11. 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.

12. Vergalasova I, Cai J. A modern review of the uncertainties in volumetric imaging of respiratory-induced target motion in lung radiotherapy. Med Phys 2020;47:e988-e1008.

13. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. Med Phys 2014;41:050902.

14. 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.

15. 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 2020.

16. 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.

17. van der Veen J, Willems S, Deschuymer S, et al. Benefits of deep learning for delineation of organs at risk in head and neck cancer. Radiother Oncol 2019;138:68-74.

18. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks. Med Phys 2017;44:547-557.

19. Zhu W, Huang Y, Zeng L, et al. Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. Med Phys 2019;46:576-589.

20. Kaderka R, Gillespie EF, Mundt RC, et al. Geometric and dosimetric evaluation of atlas based auto-segmentation of cardiac structures in breast cancer patients. Radiother Oncol 2019;131:215-220.

21. Liu J, Zhang J, Shao Y, et al. Radiation therapy system. US Patent 9,974,980, 2018-5-22.

22. Zhou J, Wang L, Ni C, et al. Systems and methods for generating radiation treatment plan. US Patent Application 16/109,707, 2019-11-7.

23. Taha AA, Hanbury A. Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool. BMC Med Imaging 2015;15:29.

24. Heimann T, van Ginneken B, Styner MA, et al. Comparison and evaluation of methods for liver segmentation from ct datasets. IEEE Trans Med Imaging 2009;28:1251-65.

25. Kuperman VY, Figueiredo G. Technical note: New similarity index for radiotherapy and medical imaging. Med Phys 2020;47:4325-4331.

26. Barrett HH, Myers KJ, Hoeschen C, et al. Task-based measures of image quality and their relation to radiation dose and patient risk. Phys Med Biol 2015;60:R1-75.

27. Conzelmann J, Schwarz FB, Hamm B, et al. Development of a method to create uniform phantoms for task‐based assessment of ct image quality. J Appl Clin Med Phys 2020;21:201-208.

Figure Captions:

Fig. 1. The workflow of this study.

Fig. 2. Geometric evaluation results of auto-segmentation. A. The DICE; B. The mean distance to agreement (MDA); C. The Jaccard; D. The Hausdorff distance (HD)

Fig. 3. An example of rectal cancer patient. A. Manual OARs; B. FD OARs; C. UIH OARs; D. Contour comparison; E. Plan_Manual dose distribution; F. Plan_FD dose distribution; G. Plan_UIH dose distribution; H. PTV contour; I. Plan_FD with manual OARs; J. Plan_UIH with manual OARs; K. 3D Gamma analysis of Plan_FD; red color represents gamma index > 1; L. 3D Gamma analysis of Plan_UIH, red color represents gamma index > 1; M. Plan_Manual DVH; N. Plan_FD DVH; O. Plan_UIH DVH.

Figures

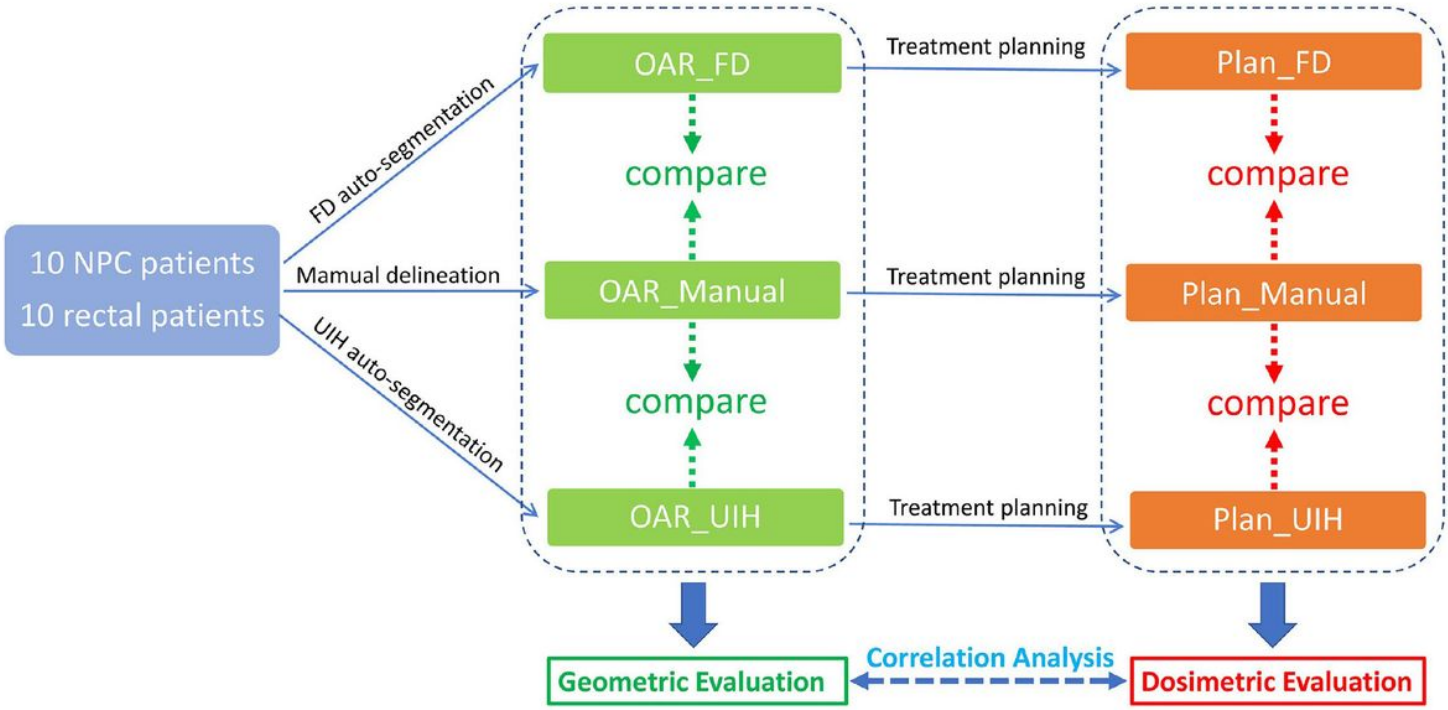

Figure 1

The workflow of this study.

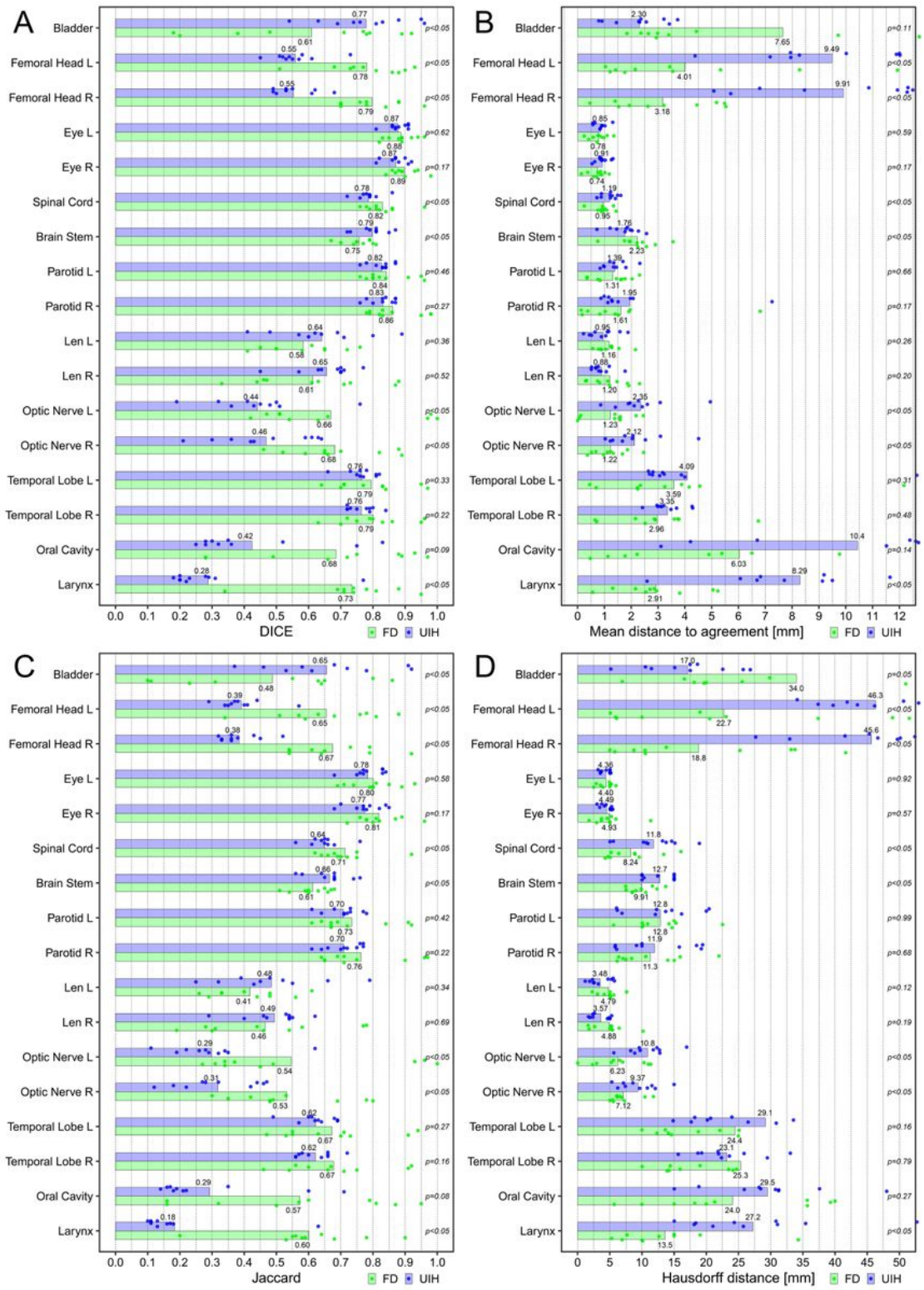

Figure 2

Geometric evaluation results of auto-segmentation. A. The DICE; B. The mean distance to agreement (MDA); C. The Jaccard; D. The Hausdorff distance (HD)

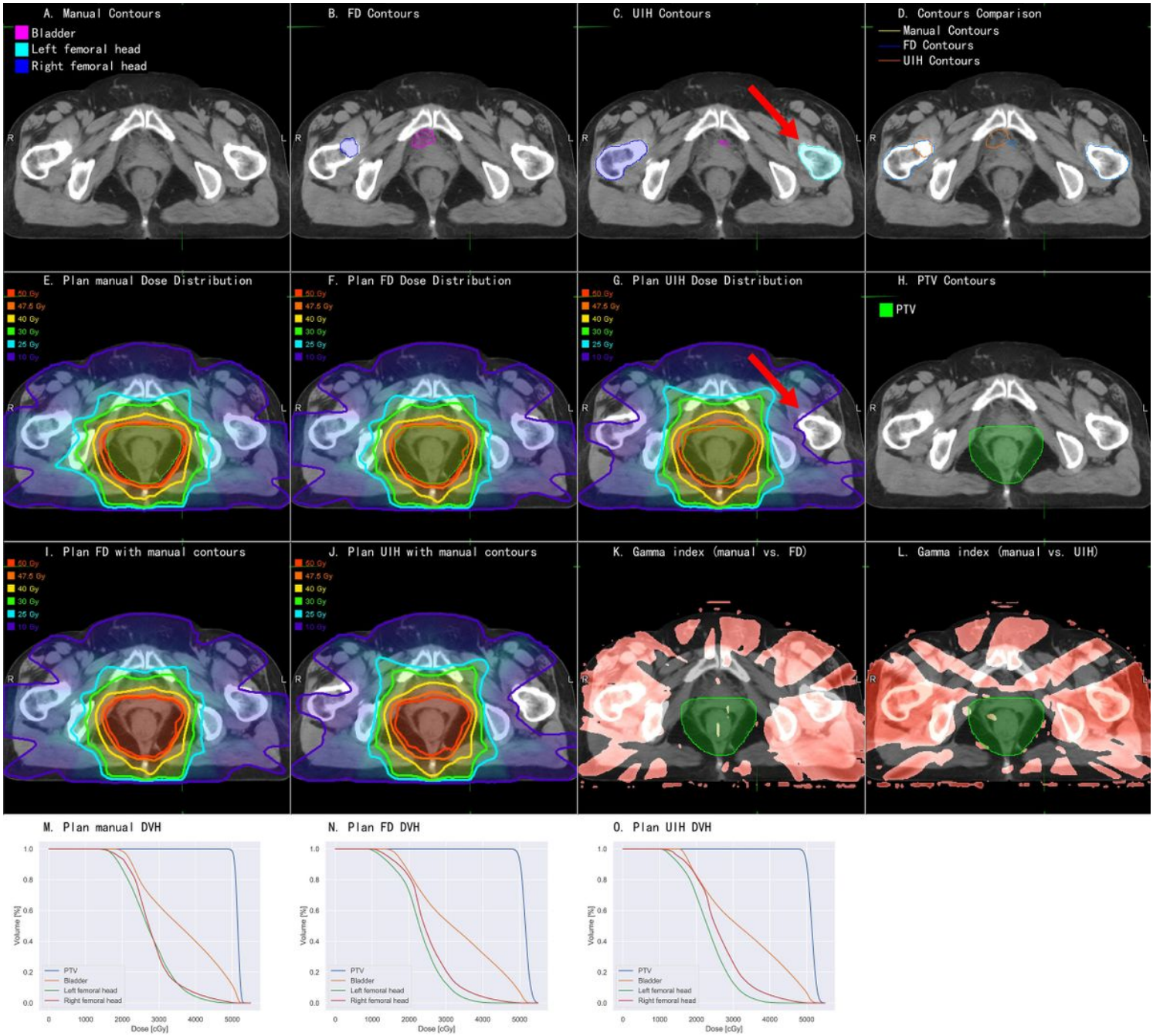

Figure 3

An example of rectal cancer patient. A. Manual OARs; B. FD OARs; C. UIH OARs; D. Contour comparison; E. Plan_Manual dose distribution; F. Plan_FD dose distribution; G. Plan_UIH dose distribution; H. PTV contour; I. Plan_FD with manual OARs; J. Plan_UIH with manual OARs; K. 3D Gamma analysis of Plan_FD; red color represents gamma index > 1; L. 3D Gamma analysis of Plan_UIH, red color represents gamma index > 1; M. Plan_Manual DVH; N. Plan_FD DVH; O. Plan_UIH DVH.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx