

## TECHNICAL NOTE

# Impact of slice thickness, pixel size, and CT dose on the performance of automatic contouring algorithms

Kai Huang<sup>1,2</sup> | Dong Joo Rhee<sup>1,2</sup> | Rachel Ger<sup>2</sup> | Rick Layman<sup>3</sup> | Jinzhong Yang<sup>2</sup>  | Carlos E. Cardenas<sup>2</sup> | Laurence E. Court<sup>2</sup>

<sup>1</sup>The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA

<sup>2</sup>Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

<sup>3</sup>Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Author to whom correspondence should be addressed. Laurence E. Court  
E-mail: LECourt@mdanderson.org.

**Funding information**

National Cancer Institute, Grant/Award Number: UH3CA202665; Siemens Healthineers; Varian Medical Systems

**Present address**

Rachel Ger, Radiation Oncology Department, Mayo Clinic, Phoenix, AZ, USA

**Abstract**

**Purpose:** To investigate the impact of computed tomography (CT) image acquisition and reconstruction parameters, including slice thickness, pixel size, and dose, on automatic contouring algorithms.

**Methods:** Eleven scans from patients with head-and-neck cancer were reconstructed with varying slice thicknesses and pixel sizes. CT dose was varied by adding noise using low-dose simulation software. The impact of these imaging parameters on two in-house auto-contouring algorithms, one convolutional neural network (CNN)-based and one multiatlas-based system (MACS) was investigated for 183 reconstructed scans. For each algorithm, auto-contours for organs-at-risk were compared with auto-contours from scans with 3 mm slice thickness, 0.977 mm pixel size, and 100% CT dose using Dice similarity coefficient (DSC), Hausdorff distance (HD), and mean surface distance (MSD).

**Results:** Increasing the slice thickness from baseline value of 3 mm gave a progressive reduction in DSC and an increase in HD and MSD on average for all structures. Reducing the CT dose only had a relatively minimal effect on DSC and HD. The rate of change with respect to dose for both auto-contouring methods is approximately 0. Changes in pixel size had a small effect on DSC and HD for CNN-based auto-contouring with differences in DSC being within 0.07. Small structures had larger deviations from the baseline values than large structures for DSC. The relative differences in HD and MSD between the large and small structures were small.

**Conclusions:** Auto-contours can deviate substantially with changes in CT acquisition and reconstruction parameters, especially slice thickness and pixel size. The CNN was less sensitive to changes in pixel size, and dose levels than the MACS. The results contraindicated more restrictive values for the parameters should be used than a typical imaging protocol for head-and-neck.

**KEY WORDS**

atlas-based segmentation, auto-contouring, convolutional neural network, CT parameters

## 1 | INTRODUCTION

The advancement of radiation treatment techniques has allowed precise delivery of radiation to a target with minimal toxicity to normal tissue. A crucial step in achieving precise treatment delivery is the accurate delineation of normal structures. This delineation process is prone to inter- and intraobserver variabilities.<sup>1,2</sup> This could contribute substantially to the amount of variation in patient planning using advanced radiation techniques, such as intensity-modulated radiation therapy, which strive for subcentimeter accuracy.<sup>3</sup> Furthermore, manual delineation is time-consuming, and studies have shown that experts need at least 60 min to manually delineate the targets and organs-at-risk structures for an average patient with head-and-neck cancer.<sup>4</sup> Delineation is expected to take longer if disease stage-specific modifications are taken into consideration.

There has been much research aimed at reducing the overhead time and producing more consistent contours.<sup>5</sup> There are two main categories of automatic segmentation systems for normal tissues: atlas-based and deep learning-based segmentation. One example of an atlas-based tool, the multiatlas contouring system (MACS), has been clinically used in our institution for several years.<sup>6–10</sup> More recently, deep learning-based algorithms have been found to outperform atlas-based algorithms in contouring some normal structures from CT images.<sup>5,11–13</sup> In order to take advantage of these advances, Rhee et al.<sup>14</sup> recently developed an auto-contouring algorithm for head-and-neck normal structures using convolutional neural networks (CNNs). Both of these tools (MACS and CNN-based algorithms) are expected to be deployed as part of the radiation planning assistant (RPA) system. The RPA system aims to automate the entire treatment planning process including contouring, planning, and dose calculation with minimal user interactions.<sup>15</sup> In this framework, these tools will likely be used with CT images from a wide range of CT scanners, using a wide range of imaging protocols.

It is known that CT image quality affects the performance of expert contour delineations.<sup>16,17</sup> Thus, the performance of auto-contour algorithms could also depend on the quality of CT images. However, to the best of our knowledge, the impact of CT image quality on auto-contouring algorithm performance has not been assessed. The CT acquisition and reconstruction parameters investigated are slice thickness, pixel size, and CT dose. Previous studies from Ger et al. and Berthelet et al. have shown quantitative image analyses such as radiomics analysis and segmentation are impacted by different choice in slice thickness, pixel size, and CT dose.<sup>18,19</sup> The purpose of this study is to provide reasonable expectations of the deviations in the contours produced by two independent auto-contouring systems providing specific CT acquisition and reconstruction parameters.

## 2 | MATERIALS AND METHODS

### 2.A | Image datasets

This study was performed retrospectively and did not directly involve and pose risk to study participants. The study was compliant

with health insurance portability and accountability act and approved by institutional review board under protocol #PA6-0379. This study used CT sinograms from 11 patients with head-and-neck cancer. Six of the CT sinograms were acquired at a Philips Brilliance Big Bore scanner with  $16 \times 0.75$  mm slice collimation and reconstructed at the console using a range of reconstruction settings (Philips Healthcare, Netherlands). The other five patient scans were acquired at a Siemens Somatom Definition Flash scanner with  $64 \times 0.6$  mm slice collimation and reconstructed using proprietary software ReconCT (v14.1.0.30238, Siemens Healthineers, Forchheim, Germany). Siemens scans with lower doses were simulated by adding noise in the raw projection domain prior to reconstruction using ReconCT. The methodology has been previously described and validated.<sup>20–22</sup>

Computed tomography scans were reconstructed from the sinograms with varying slice thicknesses, pixel sizes, and simulated mAs ranged from 10% to 100% of the original dose level. The slice spacing for all the reconstructed scans had the same value as the slice thickness. The number of patients and CT scans used for each parameter evaluation is summarized in Table 1 (both sets of data). The number of CT scans was not a multiple of the number of patients, because different options were offered by console and ReconCT at the time of reconstruction for each parameter. The range of CTDI<sub>vol</sub> 16 cm for the scans was 48.1 to 61.1 mGy and tube voltage was 120 kVp. In our institution, protocols are designed such that difference from our scanners is minimized for radiological purposes. The standard clinical soft tissue kernel was applied during reconstruction, specifically, Siemens used SAFIRE-Standard iterative reconstruction with J30S kernel and strength 3, and Philips used B or UB. UB is a slightly sharper kernel than B but it would not have a significant effect on soft tissues. Our experience, based on creating, testing and using auto-contouring models with images from multiple scanners is that differences in scanners do not noticeably affect the contouring.

### 2.B | Automated contouring

Normal tissues on all images were contoured using two approaches, MACS and a CNN-based auto-contouring tool. MACS used a set of 12 CT scans as atlases with structures carefully delineated and

**TABLE 1** The number of patients, CT scans and the levels evaluated for each parameter. Due to the availability of levels offered by console and ReconCT, the number of CT scans is not a multiple of the number of patients.

	Number of		Levels evaluated for each parameter
	Patients	CT scans	
Slice thickness	11	105	0.6, 0.75, 0.9, 1, 1.5, 1.8, 2, 3, 4, 5, 6, 7, 8, 10 mm
Pixel size	10	53	0.49, 0.59, 0.68, 0.78, 0.88, 0.98, 1.07, 1.17 mm
Dose	5	25	10%, 25%, 50%, 75%, 100% of original dose
Total		183	

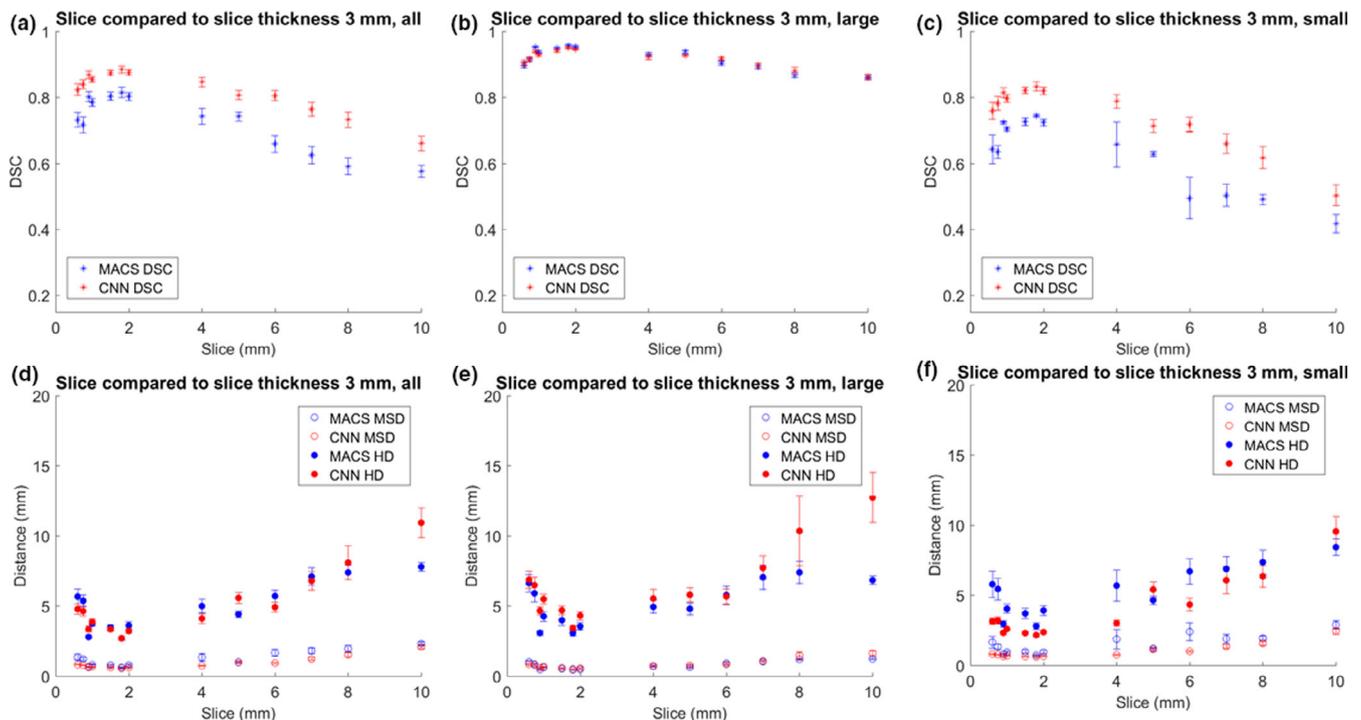
independently validated by physicians.<sup>6</sup> All of the atlases used typical clinical protocols for head and neck with either 2.5 or 3 mm slice thickness, 0.98 to 1.17 mm pixel spacing, and 22.8 to 59.4 mGy CTDI<sub>vol</sub> 16 cm. These values were within the investigated range for each parameter evaluated in this study. For MACS, CT is registered with each atlas using a dual-force “Demons” deformable registration algorithm proposed.<sup>23</sup> The resulting deformation vector fields are then used to deform the contours from each atlas to the new CT scan, creating multiple segmentations of one structure. These multiple segmentations, corresponding to each atlas, are then fused to produce the final contours on the input CT scan using a modified STAPLE algorithm.<sup>7</sup>

The CNN-based tool was trained, validated, and tested on CT scans from 3495 patients with head-and-neck cancers. For all of the CT scans used for training the CNN-based tool, the pixel spacing ranged from 0.53 to 1.37 mm, the slice thickness ranged from 1 to 3 mm, and CTDI<sub>vol</sub> 16 cm ranged from 5.8 to 72.0 mGy. All of the scans used typical clinical protocols for imaging head-and-neck and brain tumors. The CNN-based tool employs a two-step process: (a) a classification model detects the existence of a structure and (b) a segmentation model delineates the detected structures in each CT slice. The classification model uses inception-residual network (Inception-ResNet-v2<sup>24</sup>) architecture and the segmentation model uses both 3D-VNet<sup>25</sup> and a 2D fully convolutional network (FCN-8s<sup>26</sup>) architecture depending on the structure size.<sup>14</sup>

Both the MACS and the CNN-based tool resampled all the input data into the same voxel size of 0.98 mm × 0.98 mm × 2.5 mm with bilinear interpolation before prediction. Sixteen structures were auto-contoured on the available images using both the MACS and the CNN-based tools. These auto-contours including large structures — the brain, brainstem, spinal cord, left and right eyes, mandible, and left and right parotid glands, and small structures — the left and right cochlea, esophagus, optic chiasm, left and right lenses, left and right optic nerves. Structures with predicted average volumes larger than 9 cm<sup>3</sup> were classified as large.

## 2.C | Contour evaluation

The normal tissue contours produced with CT images with varying reconstruction parameters were then compared against the contours produced when using clinical baseline values (i.e., baseline values standard in our clinic) with a resampling voxel of 1 mm<sup>3</sup>. The tool for contour comparison was validated with virtual phantoms such as boxes and spheres with known ground-truth values. The metrics for comparisons were the Dice similarity coefficient (DSC), maximum distance Hausdorff distance (HD), and mean surface distance (MSD).<sup>27</sup> Specifically for investigating the effects of slice thickness variations, auto-contours from scans with varied slice thickness were compared with the corresponding auto-contours from scans with 3 mm slice thickness. The slice thickness of 3 mm is the



**Fig. 1.** Comparison of DSC, MSD, and HD for the auto-contours produced for varying slice thickness and those from a control of slice thickness of 3 mm. Figure parts (a) and (d) show the average values of all patients and structures. Figure parts (b) and (e) show the average values of all patients and large structures. Figure parts (c) and (f) show the average values of all patients and small structures. Large structures included the brain, brainstem, spinal cord, left and right eyes, mandible, and left and right parotid glands, and small structures included the left and right cochlea, esophagus, optic chiasm, left and right lenses, left and right optic nerves.

forementioned baseline value. Similarly for pixel size, baseline value was 0.977 mm. All of the contours from CT images with added noise were compared to the contours from the original CT images with a 100% dose level. The baseline values were selected based on most common clinical values for each parameter used at our institution. The evaluated ranges were the full range of values provided by the ReconCT software and the imaging console for the Philips scanner. Similar values were reported by Kisling et al based on a survey made to the medical physics community.<sup>28</sup>

### 3 | RESULTS

The MACS and CNN-based tool took approximately 26 and 3 min per scan, respectively, to produce contours. Figures 1–3 show the mean DSC, MSD, and HD for auto-contours produced when varying parameter values, with the results shown separately for large, small, and all structures. The figures showed results for scans with parameter values smaller and larger than the baseline values, leaving the lines disconnected at the baseline values. This was because the contours from scans for baseline values compared to themselves had DSC of 1 and distance measures of 0 mm. Connecting the lines through perfect DSC and distance measures would create artificial peaks in figures, thus the regions were left void.

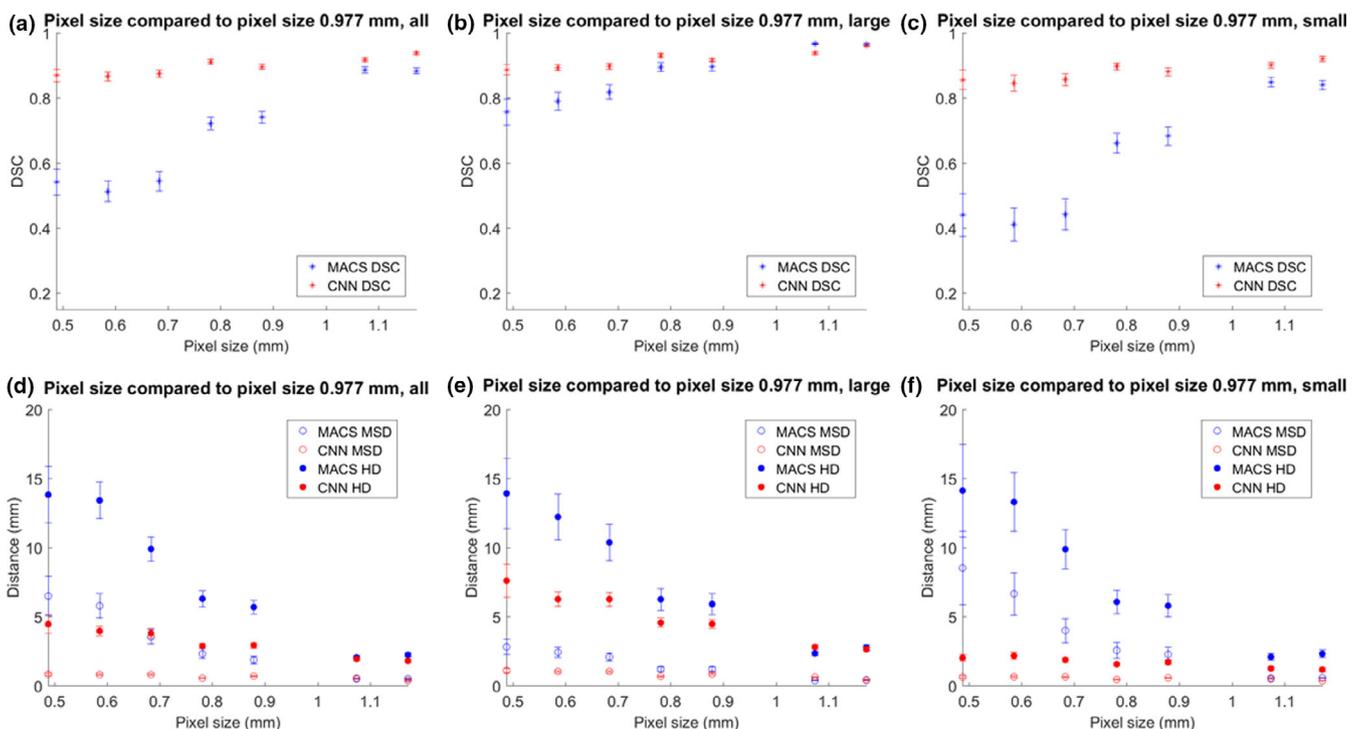
Table 2 shows the slope of a linear fit for averaged values with all structures before and after baseline. The corresponding values

were plotted in (a) and (b) subfigures of Figs. 1–3. The baseline values (perfect DSC and distance measures) were not included in calculating the slope of the linear fit. Based on the slopes, the deviations in dose were smallest amongst the three parameters. Overall, the rates of change for DSC, HD, and MSD were higher for MACS than those for CNN, except for distance measures from slice thickness parameter, in which case the slopes were comparable.

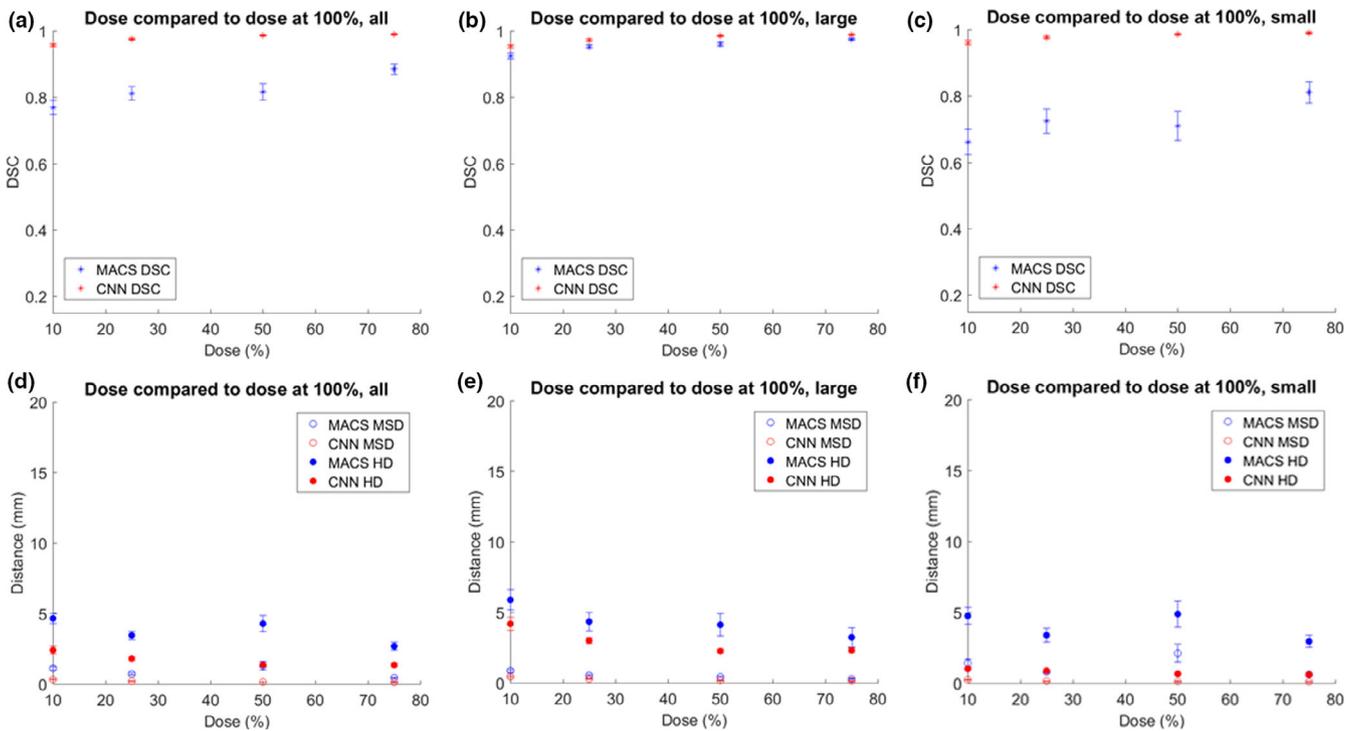
### 4 | DISCUSSION

In this study, we investigated two in-house algorithms: a CNN-based and a MACS-based auto-contouring tools. The exact impact of image resolution on an individual auto-contouring tool might be subject to details of implementation. Overall, for the specific implementation in this study, we have demonstrated that auto-contours can deviate with changes in CT acquisition and reconstruction parameters, especially slice thickness and pixel size. The changes in DSC were more stable for large structure than small structures.

A deviation from the perfect score (in Figs. 1–3) should be carefully interpreted. In this study, the ground truth was the results as auto-contoured on the CT images with standard image parameters (pixel size, etc.). For example, we may expect to see an improvement in contouring quality as slice thickness is reduced, but this may not be realized in the current study because of the use of the 3 mm slice data as the baseline, which may not be perfect. The values



**FIG. 2.** Comparison of DSC, MSD, and HD for the auto-contours produced for varying pixel thickness and those from a control of pixel thickness 0.977 mm. Figure parts (a) and (d) show the average values of all patients and structures. Figure parts (b) and (e) show the average values of all patients and large structures. Figure parts (c) and (f) show the average values of all patients and small structures. Large structures included the brain, brainstem, spinal cord, left and right eyes, mandible, and left and right parotid glands, and small structures included the left and right cochlea, esophagus, optic chiasm, left and right lenses, left and right optic nerves.



**FIG. 3.** Comparison of DSC, MSD, and HD for the auto-contours produced for varying dose and those from a control of 100% of original dose. Figure parts (a) and (d) show the average values of all patients and structures. Figure parts (b) and (e) show the average values of all patients and large structures. Figure parts (c) and (f) show the average values of all patients and small structures. Large structures included the brain, brainstem, spinal cord, left and right eyes, mandible, and left and right parotid glands, and small structures included the left and right cochlea, esophagus, optic chiasm, left and right lenses, left and right optic nerves.

**TABLE 2** The slope of a linear fit for averaged values for all structures smaller and larger than baseline. There were only two data points for calculating the slope for pixel size larger than baseline.

		Pixel size		Slice thickness		Dose
		Smaller than baseline	Larger than baseline	Smaller than baseline	Larger than baseline	
MSD	MACS	-1.301	0.046	-0.026	0.020	-0.001
	CNN	-0.056	-0.124	-0.014	0.022	0.000
HD	MACS	-2.396	0.161	-0.110	0.059	-0.002
	CNN	-0.426	-0.118	-0.116	0.111	-0.002
Dice	MACS	0.623	-0.110	0.037	-0.032	0.002
	CNN	0.103	0.131	0.031	-0.030	0.000

presented in this study demonstrated the amount of deviation from the results of baseline, and were not necessarily indicative of improvements or deteriorations in performance compared with the results of baseline. In future work digital phantoms with a known ground truth could be developed and evaluated.

Both MACS and CNN resampled input data to a fixed resolution. For MACS, input data would be resampled to the resolution of atlases. Therefore, input data with much finer resolution did not have pronounced advantage as the information would be lost in the resampling step. However, even with the same resolution, the

resulting data slices may shift due to different implementations of where the image origin should be. This might contribute in part to the deviation for metrics in the figures due to image resolution.

Additionally, there were uncertainties associated with the contour comparison tool. These uncertainties mainly stem from the volume calculation especially at the beginning and tailing slices of the structures. Therefore, by setting the reference voxel to a standard of  $1 \text{ mm}^3$ , the impact of uncertainties from the comparison tool to the final results was minimal.

Specifically, for slice thickness, large structures could achieve DSC values of more than 80%. Overall, small structures had lower DSC values compared to large structures. Additionally, the trend was amplified in slice thicknesses larger than 3 mm compared to smaller. For head-and-neck CT simulations, slice thickness is recommended to be no more than 3 mm.<sup>29,30</sup> A reference protocol for head CT from manufacturer can have slice thickness range from 0.5 to 6 mm depending on the machine specifications.<sup>31</sup> Both results for DSC and distance measures indicated larger deviations when the slice thickness is larger than 6 mm. The variation for the slice thickness smaller than 3 mm (0.6–3 mm) may be explained in part by image resampling within the auto-contouring algorithm. The slight increase in trend as the slice thickness decreases from 3 to 0.6 mm could result from missing slices in the most cranial and caudate parts of the structures. The variation for values smaller than 3 mm is more stable than the upward trend for values larger than 3 mm. Overall, the results did not indicate a more restrictive

value for slice thickness to be implemented than what has been recommended by guidelines.

In regards to pixel size variations, CNN-based solution produced more stable results compared to the corresponding values from MACS. When the pixel size decreases, the images become noisier and the interpolation in resampling process might be negatively impacted causing deteriorated predictions. The amplified response of MACS compared to CNN due to small pixel sizes may be partly caused by limited range of pixel size in the atlases. Even though all of the input data would be resampled, no atlases used a pixel size smaller than 0.98 mm, whereas CNN-based tool was trained on a much wider range of pixel sizes (from 0.53 to 1.37 mm). Even with the same resolution, more variations in the training data might have contributed to the relatively more robust result from CNN-based tool. Further studies with varying standard resampling grid in the auto-contouring tools are required to systematically evaluate the impact of resampling on contour results.

As for changes in dose levels, CNN-based solution was less sensitive to the change in dose compared to MACS. This may in part due to the wider range of  $CTDI_{vol}$  values in the training dataset for CNN-based tool. The decrease in dose affects DSC values of the small structures more than the large structures, especially for MACS. In terms of distance measures, both HD and MSD increased as dose decreases. The HD values from CNN for large structures were larger than those for the smaller structures. This may be explained by post-processing in CNN algorithm would only retain the largest predicted volume if there are more than one volume predicted for one structure. If there were discontinuities in predictions, postprocessing would remove smaller volumes, making predicted volumes much smaller than the other scans for comparison. The details of the post-processing in algorithms may vary considerably based on the specific implementations.

For CNN-based method, small and large structures used different architectures. For the atlas-based method, both large and small structures used the same set of atlases and registration process. It is possible some of the differences in the evaluation results for CNN-based method may be attributed to the segmentation networks used. However, the results from MACS-based method follow the same trend and are amplified compared to the corresponding data from CNN-based method. Therefore, it is reasonable to conclude the contribution of different networks to the discrepancies between large and small structures are minimal. One of the limitations of this study is the only two specific auto-contouring algorithms were investigated. These two algorithms have been implemented clinically in our institution. However, the specific details of implementation vary between various auto-contouring tools. Thus, the results of this study may not be generalized to other auto-contouring tools without further investigation.

## 5 | CONCLUSIONS

We investigated the impact of CT image quality parameters including slice thickness, pixel size, and dose levels, on auto-contouring

algorithms. The results showed that CNN-based tool was less impacted by the change of pixel size and dose levels as compared to MACS. The results do not indicate more restrictive values for the image resolutions should be used than a typical imaging protocol for head-and-neck. The assessment of the impact of CT image quality parameters is warranted before commissioning auto-contouring algorithms for clinical uses.

## ACKNOWLEDGMENTS

The work was partially funded by National Cancer Institute UH3CA202665 grant. The authors acknowledge the support from Siemens Healthineers for providing the ReconCT software and the Department of Scientific Publications for providing valuable edits.

## CONFLICT OF INTEREST

Other aspects of the Radiation Planning Assistant project are funded by Varian Medical Systems.

## CONTRIBUTION STATEMENTS

All authors contribute substantially to the conception, data acquisition, analysis, and writing of the manuscript.

## REFERENCES

1. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, Calandrino R. Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiother Oncol.* 1998;47:285–292.
2. Collier DC, Burnett SSC, Amin M, et al. Assessment of consistency in contouring of normal-tissue anatomic structures. *J Appl Clin Med Phys.* 2003;4:17–24.
3. Klein EE, Hanley J, Bayouth J, et al. Task Group 142 report: quality assurance of medical accelerators. *Med Phys.* 2009;36:4197–4212.
4. Teguh DN, Levendag PC, Voet PWJ, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. *Int J Radiat Oncol Biol Phys.* 2011;81:950–957.
5. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. *Semin Radiat Oncol.* 2019;29:185–197.
6. 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.
7. Yang J, Haas B, Fang R, et al. Atlas ranking and selection for automatic segmentation of the esophagus from CT scans. *Phys Med Biol.* 2017;62:9140–9158.
8. Yang J, Amini A, Williamson R, et al. Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. *Pract Radiat Oncol.* 2013;3:e139–e147.
9. Yang J, Zhang Y, Zhang L, et al. *Automatic Segmentation of Parotids from CT Scans Using Multiple Atlases.* Medical Image Analysis for the Clinic: A Grand Challenge; 2010:323–330.
10. Zhou R, Liao Z, Pan T, et al. Cardiac atlas development and validation for automatic segmentation of cardiac substructures. *Radiother Oncol.* 2017;122:66–71.

11. 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.
12. Raudaschl PF, Zaffino P, Sharp GC, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015. *Med Phys*. 2017;44:2020–2036.
13. Yang J, Veeraraghavan H, Armato SG 3rd, et al. Autosegmentation for thoracic radiation treatment planning: a grand challenge at AAPM 2017. *Med Phys*. 2018;45:4568–4581.
14. Rhee DJ, Cardenas CE, Elhalawani H, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys*. 2019;46:5086–5097.
15. Court LE, Kisling K, McCarroll R, et al. Radiation planning assistant - a streamlined, fully automated radiotherapy treatment planning system. *J Vis Exp*. 2018;134:1–9.
16. Caldwell CB, Mah K, Ung YC, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys*. 2001;51:923–931.
17. Bellon MR, Siddiqui MS, Ryu S, Chetty IJ, et al. The effect of longitudinal CT resolution and pixel size (FOV) on target delineation and treatment planning in stereotactic radiosurgery. *J Radiosurg SBRT*. 2014;3:149–163.
18. Ger RB, Zhou S, Melinda Chi P-C, et al. Comprehensive investigation on controlling for CT imaging variabilities in radiomics studies. *Sci Rep*. 2018;8:13047.
19. Berthelet E, Liu M, Truong P, et al. CT slice index and thickness: impact on organ contouring in radiation treatment planning for prostate cancer. *J Appl Clin Med Phys*. 2003;4:365–373.
20. Kramer M, Ellmann S, Allmendinger T, et al. Computed tomography angiography of carotid arteries and vertebrobasilar system: a simulation study for radiation dose reduction. *Medicine (Baltimore)*. 2015;94:e1058.
21. Yu L, Shiung M, Jondal S, McCollough CH. Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols. *J Comput Assist Tomogr*. 2012;36:477–487.
22. Ellmann S, Kammerer F, Brand M, et al. A novel pairwise comparison-based method to determine radiation dose reduction potentials of iterative reconstruction algorithms, exemplified through circle of Willis computed tomography angiography. *Invest Radiol*. 2016;51:331–339.
23. Wang H, Dong L, Lii MF, et al. Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. *Int J Radiat Oncol Biol Phys*. 2005;61:725–735.
24. Szegedy C, Ioffe S, Vanhoucke V, et al. *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*. arXiv e-prints; 2016. arXiv:1602.07261.
25. Milletari F, Navab N, Ahmadi SA. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. Proceedings of 2016 Fourth International Conference on 3d Vision (3dv); 2016: 565–571.
26. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39:640–651.
27. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29.
28. Kisling KD, Ger RB, Netherton TJ, et al. A snapshot of medical physics practice patterns. *J Appl Clin Med Phys*. 2018;19:306–315.
29. Radiology, A.C.o. *ACR-ASNR-SPR PRACTICE PARAMETER FOR THE PERFORMANCE OF COMPUTED TOMOGRAPHY (CT) OF THE EXTRACRANIAL HEAD AND NECK*; 2016.
30. Merlotti A, Alterio D, Vigna-Taglianti R, et al. Technical guidelines for head and neck cancer IMRT on behalf of the Italian association of radiation oncology - head and neck working group. *Radiat Oncol*. 2014;9:264.
31. American Association of Physics in Medicine. *Adult Routine Head CT Protocols Version 2.0*; 2016. p. 20.