Plan Quality-Driven Evaluation of Automated Segmentation for Radiotherapy

Ji Zhu[1*], Xinyuan Chen[1*], Nan Bi[1], Tao Zhang[1], Kuo Men[1#] and Jianrong Dai[1#]

1 National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

JULY 12–16 2020 VIRTUAL JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

INTRODUCTION

Automatic segmentation model is proven to be efficient in delineation of organs at risk (OARs) in radiotherapy; its performance is usually evaluated with geometric differences between automatic and manual delineations. However, dosimetric differences attract more interests than geometric differences in clinics. Therefore, this study evaluates the performance of automatic segmentation with dosimetric metrics for volumetric modulated arc therapy of esophageal cancer patients.

AIM

This study introduces a dosimetric evaluation system to substitute the geometric evaluations on automatic delineation for esophageal cancer volumetric modulated arc therapy radiotherapy.

METHOD

Nineteen esophageal cancer cases were assessed in this study. Physicians manually delineated the target volumes and the OARs for each case. Another set of OARs was automatically generated using convolutional neural network models. The radiotherapy plans were optimized with the manually delineated targets and the automatically delineated OARs. Segmentation accuracy was evaluated by Dice similarity coefficient (DSC) and mean distance to agreement (MDA). Dosimetric metrics of manually and automatically delineated OARs were obtained and compared. The clinically acceptable dose difference and volume difference of OARs between manual and automatic delineations are supposed to be within 1 Gy and 1%, respectively.

RESULTS

Average DSC values were above 0.92 except for the spinal cord (0.82), and average MDA values were below 0.90 mm except for the heart (1.74 mm). 11 of the 20 dosimetric metrics of the OARs were not significant ($p > 0.05$). Although there were significant differences ($p < 0.05$) for the spinal cord (D2%), left lung (V10, V20, V30 and mean dose), and bilateral lung (V10, V20, V30 and mean dose), their absolute differences were small and acceptable for the clinic.

The maximum dosimetric metrics differences of OARs between manual and automatic delineations were ΔD2% = 0.35 Gy for the spinal cord and ΔV30 = 0.4% for the bilateral lung, which were within the clinical criteria in this study.

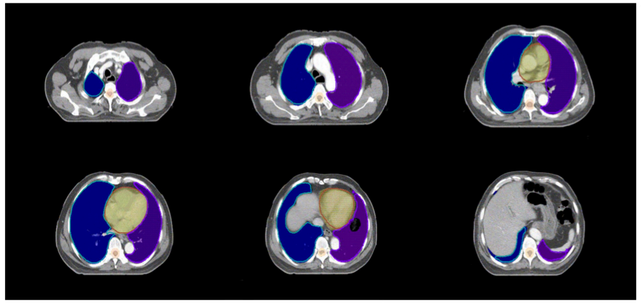

Figure 1. Examples of the segmentation. Colourwash: Manual segmentation; Line: Automatic segmentation

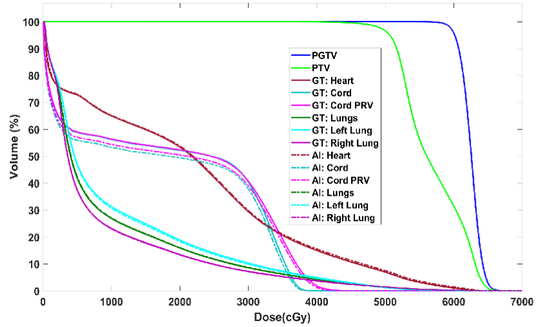

Figure 2. Comparison of mean DVH of all the 19 cases. Solid line: Manual segmentation; Dash line: Automatic segmentation

Table 1 The paired t-test outcome of the dosimetric characteristic of OARs between manual and deep learning automatic delineation-based plan

Dosimetric metrics		GT	AI	P-value
Spinal Cord	D2% (Gy)	36.08±0.41	35.73±0.41	<0.01
Spinal Cord PRV	D2% (Gy)	40.25±0.43	40.42±0.30	0.55
Heart	V30 (%)	28.60±4.06	28.70±4.09	0.87
	V40 (%)	14.68±2.19	15.00±2.28	0.48
	Mean (Gy)	20.54±2.71	20.64±2.76	0.65
Lung all	V30 (%)	8.63±2.69	8.23±2.73	0.02
	V20 (%)	15.81±4.95	15.63±4.99	<0.01
	V10 (%)	26.47±8.27	26.28±8.28	0.04
	V5 (%)	41.05±12.76	41.48±13.18	0.44
	Mean (Gy)	9.26±2.55	9.21±2.57	0.04
Lung L	V30 (%)	10.24±4.86	10.01±4.95	<0.01
	V20 (%)	18.55±8.28	18.28±8.38	<0.01
	V10 (%)	30.53±12.02	30.31±12.07	0.04
	V5 (%)	45.84±16.26	45.88±16.45	0.73
	Mean (Gy)	10.31±3.68	10.21±3.71	<0.01
Lung R	V30 (%)	7.31±3.75	7.32±3.79	0.90
	V20 (%)	13.55±5.35	13.45±5.28	0.41
	V10 (%)	23.11±7.83	22.97±1.76	0.42
	V5 (%)	37.04±11.71	37.01±11.74	0.89
	Mean (Gy)	8.40±2.51	8.38±2.51	0.75

CONCLUSIONS

The findings of this study showed that the geometric evaluation between manual and automatic delineations was not enough in clinical applications. Dosimetric metrics were proposed to assess the automatic delineation in radiotherapy planning of esophageal cancer. Based on the dosimetric evaluation in this study, the manual delineation for esophageal cancer radiotherapy can be substituted by automatic delineation.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (11875320, 11975313), the Beijing Municipal Science & Technology Commission (Z181100001918002), and the Beijing Hope Run Special Fund of Cancer Foundation of China (LC2019B06, LC2018A14).

REFERENCES

1. Bi, N et al. Deep Learning Improved Clinical Target Volume Contouring Quality and Efficiency for Postoperative Radiation Therapy in Non-small Cell Lung Cancer. Front Oncol, 2019. 9: p. 1192.
2. Fung, N.T.C. et al., Automatic segmentation for adaptive planning in nasopharyngeal carcinoma IMRT: Time, geometrical, and dosimetric analysis. Medical Dosimetry, 2019.
3. Yang, J et al. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. Med Phys, 2018. 45(10): p. 4568-4581.
4. Lustberg, T et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiotherapy & Oncology, 2017. 126(2): p. 312-317.
5. Gregoire, V et al. Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). J ICRU, 2010. 10(1): p. 1-106.

CONTACT INFORMATION

Ji Zhu, 278zhuji@163.com
Xinyuan Chen, cinya126chen@163.com
Nan Bi, binan_email@163.com
Tao Zhang, zhangt10@126.com
Jianrong Dai: dai_jianrong@cicams.ac.cn
Kuo Men: menkuo@cicams.ac.cn