The use of artificial intelligence to auto-segment organs-at-risk in total marrow irradiation treatment

An Liu[1], Richard Li[1], Chunhui Han[1], Jieming Liang[1], Ashwin Shinde[1], Savita Dandapani[1], Arya Amini[1], Scott Glaser[1], Jeffrey Wong[1]

[1] Department of Radiation Oncology, City of Hope Medical Center, Duarte, CA

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

INTRODUCTION

Treatment planning for total marrow irradiation(TMI) is a time-intensive process requiring the contouring of many organs-at-risk(OARs) throughout the entire body. Auto-segmentation using Artificial intelligence can significantly reduce the contouring time and make TMI treatment planning more efficient.

AIM

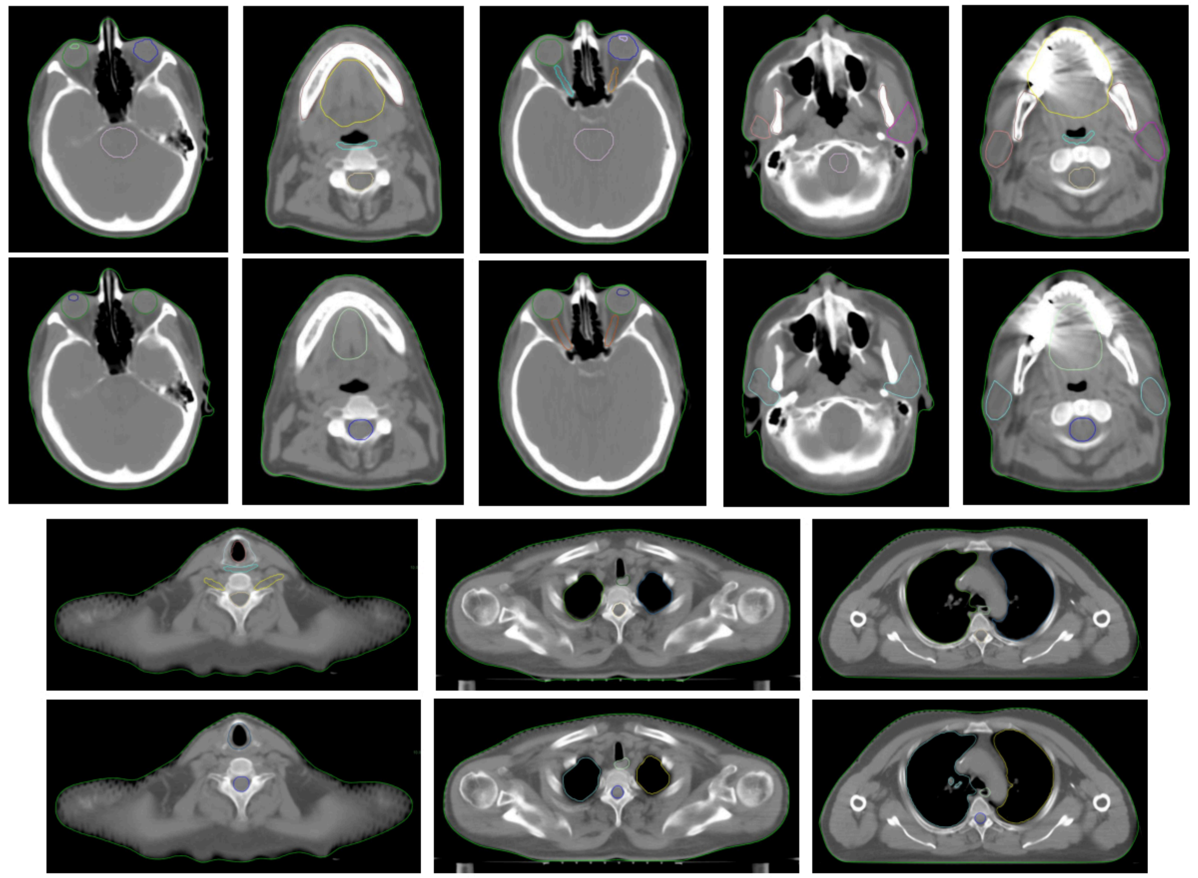
This study evaluated the quality of contours auto-generated by a deep learning (DL) contouring algorithm for OAR volumes in TMI.

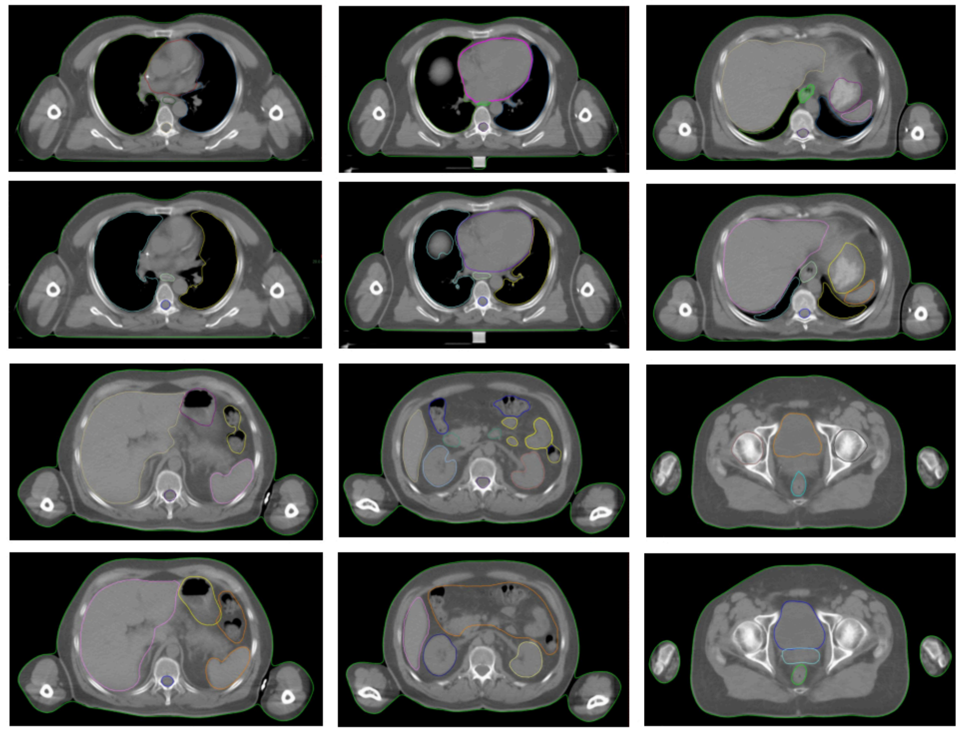
METHOD

1. The first ten patients in a phase II trial treated with TMI were selected for evaluation.
2. Dose prescriptions were 20 Gy to bone/lymph_nodes/spleen and 12 Gy to liver/brain delivered over 5 days, twice daily.
3. Each patient had more than 150 slices/30 structures to contour and took approximately 6-8 hours of dosimetrist time per patient.
4. Clinically used contours drawn by human were used as the reference.
5. A deep machine learning model (Ua-Net, DeepVoxel Inc, Irvine, CA) was used to auto-segment the OARs.
6. We evaluated the performance of this DL model using 3 spatial overlap based metrics (Dice coefficient, Jaccard index(JAC) and True positive rate sensitivity(TPR)), 2 surface distance metrics (95% Hausdorff distance(HD) and average distance(AD)), 1 volume similarity index(VS).
7. Eighteen common OARs were evaluated.

RESULTS

The DL auto-segmentation model was most similar to human generated contours for eyes, parotids, heart, liver, kidneys, spleen and lungs where average Dice, JAC, TPR, HD, AD, VS in DL model were 0.85(range 0.76-0.95), 0.72(0.62-0.91), 0.85(0.74-0.98), 14.4(7.5-24.6), 4.4mm(2.0-7.9) and 0.92(0.88-0.97) respectively. Other OARs still needed improvement. Several factors contributed to the difference. The training CT dataset used for abdomen and pelvis had patients in arms-up position, but TMI patients were simulated with arms on the side. The model was trained to draw the spinal cord in contrast to the reference where spinal canal was drawn. On the right, the DL model generated contours (top rows) and corresponding human contours (bottom rows) are shown for a typical TMI patients at various body levels.





	Dice Coefficient	Jaccard Index	True Positve Rate Sensitivity	95% Hausdorff Distance	Average Distance	Volumetric Similarity
Bowel	0.490	0.329	0.345	58.612	16.431	0.553
Esophagus	0.516	0.360	0.390	20.229	6.170	0.600
Eyes	0.811	0.687	0.748	8.574	2.717	0.892
Heart	0.802	0.677	0.805	24.636	7.498	0.941
Kidney left	0.683	0.528	0.531	18.325	6.641	0.688
Kidney right	0.639	0.490	0.498	21.424	8.428	0.655
Larynx	0.601	0.440	0.854	32.990	10.538	0.713
Lens	0.527	0.371	0.470	14.304	6.579	0.728
Liver	0.825	0.716	0.743	23.931	7.927	0.882
Lung left	0.948	0.902	0.983	7.511	2.032	0.960
Lung right	0.950	0.906	0.975	9.634	2.281	0.968
Optical nerves/chiasm	0.356	0.225	0.254	13.603	6.070	0.555
Oral cavity	0.613	0.448	0.746	19.041	9.165	0.797
Parotids	0.757	0.620	0.850	11.860	3.848	0.884
Spinal cord	0.668	0.504	0.576	154.827	34.250	0.828
Spleen	0.615	0.483	0.504	23.692	10.036	0.675
Stomach	0.479	0.336	0.348	55.042	21.494	0.511
Thyroid	0.609	0.450	0.515	11.443	3.906	0.780

Dice Score: The volumetric Dice similarity coefficient (DSC) measures the volume overlapped between the AS and manual delineations.

$$DSC = \frac{2(M_p \cap M_g)}{M_p + M_g}$$

Hausdorff Distance: HD describes the similarity between two sets of points by measuring the maximum distance of a point in M_p to the nearest point in M_g.

$$HD = max\{h(M_p, M_g), h(M_g, M_p)\}$$

$$h(M_p, M_g) = \max_{a \in M_p} \min_{b \in M_g} \| a - b \|$$

Jaccard Index

$$JAC = \frac{|S_g^1 \cap S_t^1|}{|S_g^1 \cup S_t^1|} = \frac{TP}{TP + FP + FN}$$

Sensitivity

$$Recall = Sensitivity = TPR = \frac{TP}{TP + FN}$$

Volumetric Similarity

$$VS = 1 - \frac{||S_t^1| - |S_g^1||}{|S_t^1| + |S_g^1|} = 1 - \frac{|FN + FP|}{2TP + FP + FN}$$

CONCLUSIONS

DL auto-generated contours from a convolutional neural network model showed promise to replace human generated ones for many OARs for TMI planning, with potential to be adopted in routine clinical practice and significantly reduce the lengthy contouring time. Future models may allow for auto-contouring of more organs to further reduce dosimetrist time.

ACKNOWLEDGEMENTS

Special thanks to Professor Xiaohui Xie, Department of Computer Science, University of California, Irvine and Dr. Narisu Bai, DeepVoxel, Inc, Irvine, CA for help with the AI model.

CONTACT INFORMATION

An Liu, Ph.D.
City of Hope National Medical Center
aliu@coh.org