2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

HENRY FORD HEALTH

Stress Testing Cross Domain Generalization of High Field Brain Metastasis Auto Contouring On 0.55T MRI

Joshua Kim, PhD[1]; Amanda Dicarlo, PhD[1]; Ning Jin, PhD[2]; Salim Siddiqui, MD[1]; PhD, Mira Shah[1], MD; Eric A Mellon[1], MD, PhD; Anthony J Doemer[1], MS; Kundan Thind, PhD[1]
[1]Henry Ford Health Department of Radiation Oncology, [2]Siemens Healthineers

ABSTRACT

Purpose: To quantify volumetric reliability limits of a vendor brain metastasis auto contouring prototype trained only on 1.5T and 3T MRI when deployed on 0.55T MRI.

Methods: Twelve patients with 95 physician contoured metastases were imaged on a 0.55T MR simulator. Lesion volumes ranged from 0.014 to 18.9 cc. The locked commercial prototype was applied in a zero shot setting with no exposure to less than 1.5T data. Lesion level detection was defined as any overlap between matched physician and AI contours. Geometric fidelity was summarized using Dice similarity coefficient. Performance was stratified by volume into a gray range from 0.014 to 0.065 cc and a reliable range above 0.065 cc. False positives were summarized across the same thresholds.

Results: Overall sensitivity was 92.6% (88/95). Sensitivity in the reliable range was 94.4% (67/71) with mean Dice 0.83. In the gray range, sensitivity was 87.5% (21/24) with mean Dice 0.58. For small lesions, sensitivity was 87.9% (29/33) for volumes below 0.1 cc and 91.9% (34/37) for 0.1 to 0.5 cc. Seven lesions were missed with mean volume 0.10 cc, including four above 0.065 cc. The model produced 57 false positives, 4.8 per patient, with 86.0% (49/57) below 0.065 cc and 38.6% (22/57) below 0.014 cc.

Conclusion: High field trained auto contouring generalized to 0.55T MRI with high sensitivity for lesions at or above 0.1 cc and stable performance above 0.065 cc. The dominant limitation was increased low volume false positives and reduced Dice for small lesions. However, a post processing volume filter near 0.065 cc can substantially reduce false positives, with the operational threshold guided by institutional tolerance for missing very small disease.

CONTACT

Joshua Kim
Henry Ford Health
2800 W Grand Blvd
jkim8@hfhs.org
(313) 575-4620

INTRODUCTION

Stereotactic radiotherapy (SRT) is a treatment strategy used to deliver a high and highly conformal dose in only 1-5 fractions. It is used in the intracranial treatment of malignant tumors such as brain metastases and increasingly in the treatment of patients with large numbers of brain metastases (>5). Due to the need for highly conformal dose with steep dose fall-off, it is essential that the contours are highly accurate.

Additionally, these lesions are often small with volumes less than 0.1 cc. Therefore, the contouring step for patients with large numbers of brain mets can be very time intensive. In order to expedite the process, Siemens Healthineers has developed an AI-based brain metastasis auto contouring prototype trained on 1.5T and 3T data.

METHODS AND MATERIALS

Our clinic has recently installed a 0.55T MR simulator (MAGNETOM Free.Max, Siemens Healthineers, Forchheim, Germany). In this retrospective study, we evaluated the performance of the vendor-provided auto contouring prototype when using 0.55T MR images from the MAGNETOM Free.Max system. For this study, post-contrast T1 weighted images were used as input for the auto contouring prototype.

12 patients were selected for evaluation where each patient had ≥5 brain metastases and were treated using 1-5 isocenters for treatment. Cumulatively, a total of 95 lesions were contoured by physicians for treatment for these patients. Lesions ranged in size from 0.01 cc to 18.90 cc with a median of 0.16 cc. The majority of lesions were < 0.5cc (N=70, 74%).

Lesion detection accuracy was assessed by comparing overlap of auto-contoured lesions and physician-defined lesions used for treatment. Geometric fidelity was measured using the Dice similarity coefficient (DSC) for overlapping lesions: $DSC = \frac{2|A \cap B|}{|A|+|B|}$

where A and B are physician and auto-generated contours.

In Figure 1, axial and sagittal image are shown of a correctly identified lesion demonstrating good agreement (DSC = 0.90 cc) between the contour drawn by the physician (blue) and the contour generated using the auto-contouring prototype (orange).

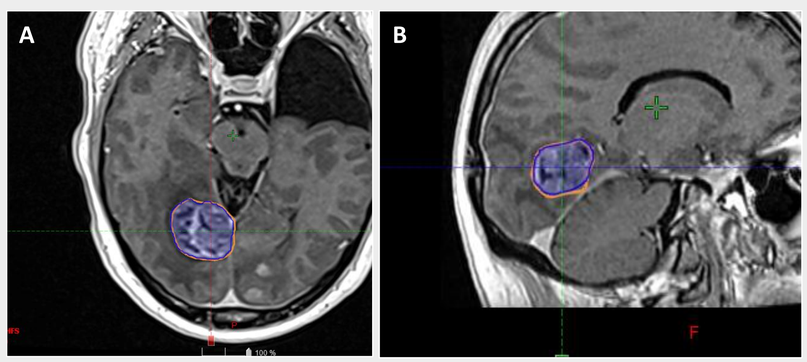

Figure 1. (A) Axial and (B) sagittal scans showing auto-contoured lesion (Blue) overlayed with the corresponding physician contour (Orange).

RESULTS

As shown in Table 1, the vendor provided auto-contouring prototype algorithm was able to identify 88 of the 95 treated lesions providing a true positive rate of 0.93. These results showed a volume dependence where the smallest lesions (< 0.065 cc) showed a lower sensitivity and a more reliable region for volumes > 0.065 cc. This volume dependence was also seen in the DSC results as seen in Figure 2. Using this data, we divided the performance into two regions where the low volume region (< 0.065 cc, "Gray") showed slightly lower performance with a sensitivity of 87.5% and a mean DSC of 0.58 and the more reliable larger volume region with a sensitivity of 94.4% and a mean DSC of 0.83. DSC showed a similar dependance.

All generated lesions that did not coincide with physician defined lesions were counted as false positives. While there were several contours that did not match the physician-defined lesions, the analysis of false positives was complicated by the presence of previously treated lesions in 8 patients of which several were enhancing in post-contrast imaging.

Overall, there appears to be a bias to maximize inclusion of small potential lesions since 86% of false positives were less than 0.065cc in volume. Further steps include physician analysis of non-matching contours to categorize as false positive, previously treated lesion, or missed lesion.

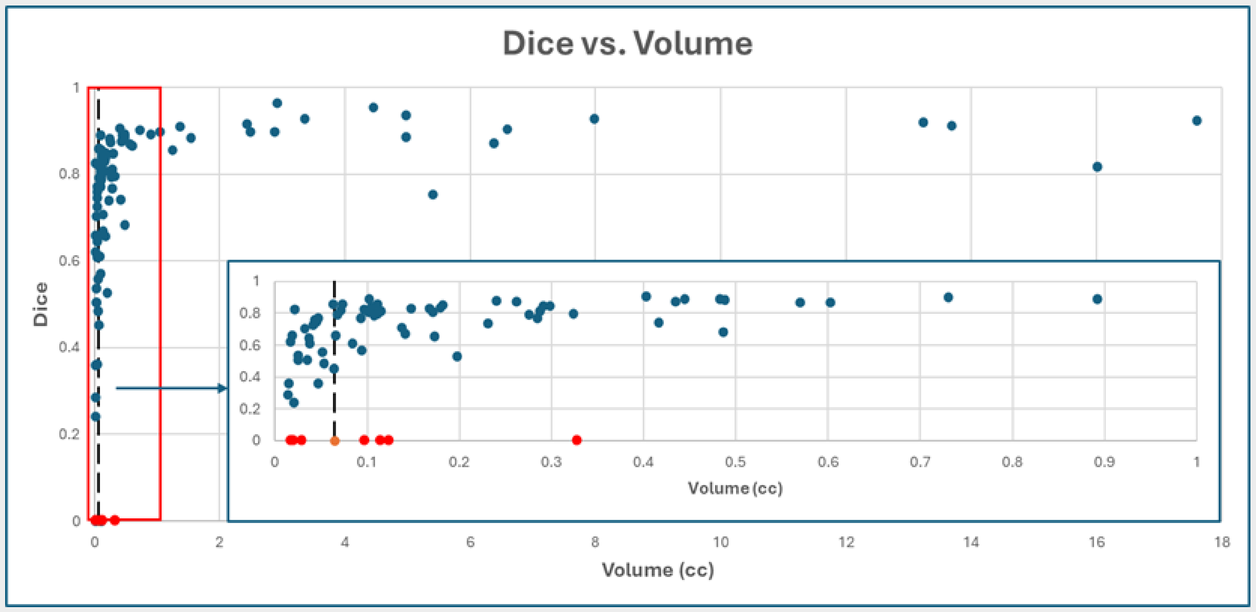

Figure 2. Dice score for corresponding auto-contoured and physician contoured lesions as a function of volume. Dotted line indicates the cutoff between gray and reliable regions (volume = 0.065 cc). False negatives are indicated by red dots.

	Gray (<0.065 cc)	Reliable (≥ 0.065 cc)	Very Small (<0.1 cc)	Small (0.1 - 0.5 cc)	Large (> 0.5 cc)	Total
True Positive	21	67	29	34	25	88
False Negative	3	4	4	3	0	7
Sensitivity	87.5%	94.4%	87.5%	91.9%	100.0%	92.6%
Dice	0.58	0.83	0.55	0.83	0.89	0.75

Table 1. Analysis of lesion detection accuracy and geometric fidelity in patients categorized by volume.

CONCLUSIONS

High field trained auto contouring prototype was generalized to 0.55T MRI and provided high sensitivity for lesions at or above 0.1 cc and stable performance above 0.065 cc.

The main limitation was an increased level of low volume false positives and a reduced Dice for small lesions. The analysis of false positives is further complicated by the presence of previously treated lesions in 8 patients of which several were enhancing in post-contrast imaging. Physician analysis of non-matching contours is planned to distinguish between false positives, previously treated lesions, and possible missed lesions.

Use of a post processing volume filter near 0.065 cc could substantially reduce apparent false positives, with the choice of an operational threshold guided by institutional tolerance used to avoid missing very small disease.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com