Development and validation of a fully automatic tissue delineation model for brain metastasis using a deep neural network

Jie-Yi Zhao[1#], Qi Cao[2#], Jing Chen[1], Wei Chen[3], Si-Yu Du[4], Jie Yu[5], Yi-Miao Zeng[4], Shu-Min Wang[4], Jing-Yu Peng[4], Chao You[1], Jian-Guo Xu[1], Xiao-Yu Wang[1]

[1]Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China; [2]Department of Reproductive Medical Center, West China Second University Hospital, Sichuan University, Chengdu, China; [3]Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China; [4]West China School of Medicine, Sichuan University, Chengdu, China; [5]West China School of Public Health, Sichuan University, Chengdu, China

Contributions: (I) Conception and design: JY Zhao, Q Cao, XY Wang; (II) Administrative support: C You, XY Wang, JG Xu; (III) Provision of study materials or patients: YM Zeng, J Yu, SY Du; (IV) Collection and assembly of data: SM Wang, JY Peng, JG Xu; (V) Data analysis and interpretation: J Chen, XY Wang, W Chen, JY Zhao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

Correspondence to: Xiao-Yu Wang, PhD; Jian-Guo Xu, PhD. Department of Neurosurgery, West China Hospital, Sichuan University, No. 37 Guoxue Lane, Wuhou District, Chengdu 610041, China. Email: yuxixi1052006@126.com; jianguo_1229@126.com.

Background: Stereotactic radiosurgery (SRS) treatment planning requires accurate delineation of brain metastases, a task that can be tedious and time-consuming. Although studies have explored the use of convolutional neural networks (CNNs) in magnetic resonance imaging (MRI) for automatic brain metastases delineation, none of these studies have performed clinical evaluation, raising concerns about clinical applicability. This study aimed to develop an artificial intelligence (AI) tool for the automatic delineation of single brain metastasis that could be integrated into clinical practice.

Methods: Data from 426 patients with postcontrast T1-weighted MRIs who underwent SRS between March 2007 and August 2019 were retrospectively collected and divided into training, validation, and testing cohorts of 299, 42, and 85 patients, respectively. Two Gamma Knife (GK) surgeons contoured the brain metastases as the ground truth. A novel 2.5D CNN network was developed for single brain metastasis delineation. The mean Dice similarity coefficient (DSC) and average surface distance (ASD) were used to assess the performance of this method.

Results: The mean DSC and ASD values were 88.34%±5.00% and 0.35±0.21 mm, respectively, for the contours generated with the AI tool based on the testing set. The DSC measure of the AI tool's performance was dependent on metastatic shape, reinforcement shape, and the existence of peritumoral edema (all P values <0.05). The clinical experts' subjective assessments showed that 415 out of 572 slices (72.6%) in the testing cohort were acceptable for clinical usage without revision. The average time spent editing an AI-generated contour compared with time spent with manual contouring was 74 *vs.* 196 seconds, respectively (P<0.01).

Conclusions: The contours delineated with the AI tool for single brain metastasis were in close agreement with the ground truth. The developed AI tool can effectively reduce contouring time and aid in GK treatment planning of single brain metastasis in clinical practice.

Keywords: Brain metastases; deep neural network; tissue segmentation; magnetic resonance imaging (MRI); Gamma Knife plan

Introduction

Brain metastases, the most common intracranial tumors in adults, are neoplasms that originate in tissues outside the brain and then spread secondarily to the brain (1). An estimated 20–40% of patients with cancer will develop brain metastases, and the true incidence is likely higher, as such estimates are often limited to patients who are considered for treatment (2,3). Among all patients with newly diagnosed brain metastases, 49–53% have single metastasis (4,5). Stereotactic radiosurgery (SRS) is now a primary treatment option for patients with brain metastases, particularly those with a limited number of lesions (i.e., 1–3) (6,7). Various treatment units can perform SRS, including Gamma Knife (GK), CyberKnife, linear accelerator (LINAC)-based radiosurgery, and TomoTherapy (8). Additionally, certain types of SRS can also be used for a higher number of lesions. For instance, up to 15 lesions are widely acknowledged as being acceptable for treatment with SRS and GK as a primary management option.

Accurate tumor delineation based on magnetic resonance imaging (MRI) is the first step toward a precise radiation prescription in GK-SRS treatment planning (9). Currently, GK surgeons manually delineate brain metastases on each axial slice using GK-SRS planning software. This is a time-consuming, laborious, and subjective task (10). An automated brain metastases delineation tool could improve the efficiency and reliability of GK-SRS treatment planning.

In recent years, deep convolution neural networks (CNNs) have been widely applied in medical image segmentation, and many successes have been achieved (11-14). Nevertheless, the clinical evaluation of learning contouring quality has been limited (15). Multiple studies have explored automatic brain metastasis delineation based on CNNs in MRI (16-20). These studies reported an average Dice similarity coefficient (DSC) ranging from 0.67 to 0.85 with the average false positives (FPs) per patient ranging from 3 to 20. A recent study reported a DSC of 0.85±0.12 with FPs of 3±3 per patient but also a relatively low positive predictive value (PPV) of 67±3. None of these studies performed a clinical evaluation, leaving the clinical applicability of CNNs in this setting unclear (16).

This study investigated the use of the deep neural network for automatic single brain metastasis. We first constructed an artificial intelligence (AI) contouring tool based on a deep CNN (DCNN). The AI tool was trained on cohort of 341 patients, and its performance was validated on a separated testing cohort of 85 patients. Next, the clinical usability of the AI tool was evaluated by 2 experienced GK surgeons. Finally, the AI tool was compared with 3 other qualified GK surgeons using 40 randomly selected patients from the testing cohort. We present this article in accordance with the TRIPOD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1216/rc).

Methods

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of West China Hospital of Sichuan University. The study was registered at clinical trial registration URL: https://www.chictr.org.cn (unique identifier: ChiCTR2100046265). Individual consent for this retrospective analysis was waived since patients were selected retrospectively and the MRI were completely anonymized before analysis.

Study population

We retrospectively collected the postcontrast T1-weighted MRI examinations of all patients who underwent GK-SRS for brain metastases at West China Hospital of Sichuan University between March 2007 and August 2019. Planning MRI axial sequences were acquired using a 1.5-T scanner (Siemens Healthineers, Erlangen, Germany) under the following parameters: in-plane resolution 0.449 mm × 0.449 mm, in-plane matrix 512×512, slice thickness 1 mm, and voxel size 0.449 mm × 0.449 mm × 1.000 mm.

A total of 1,899 examinations were collected, and the exclusion criteria for collection are presented in *Figure 1*. The final data set comprised 426 patients (247 males, 179 females; mean age 57 years), who were then randomly separated into 3 nonoverlapping cohorts at a ratio of 7:1:2: (I) a training cohort of 299 patients for AI model construction; (II) a validation cohort of 42 patients for optimization of the AI model hyperparameters; and (III) a testing cohort of 85 patients to test the performance of the AI model.

Manual delineation of contours

For all the 426 patients, the brain metastasis volume
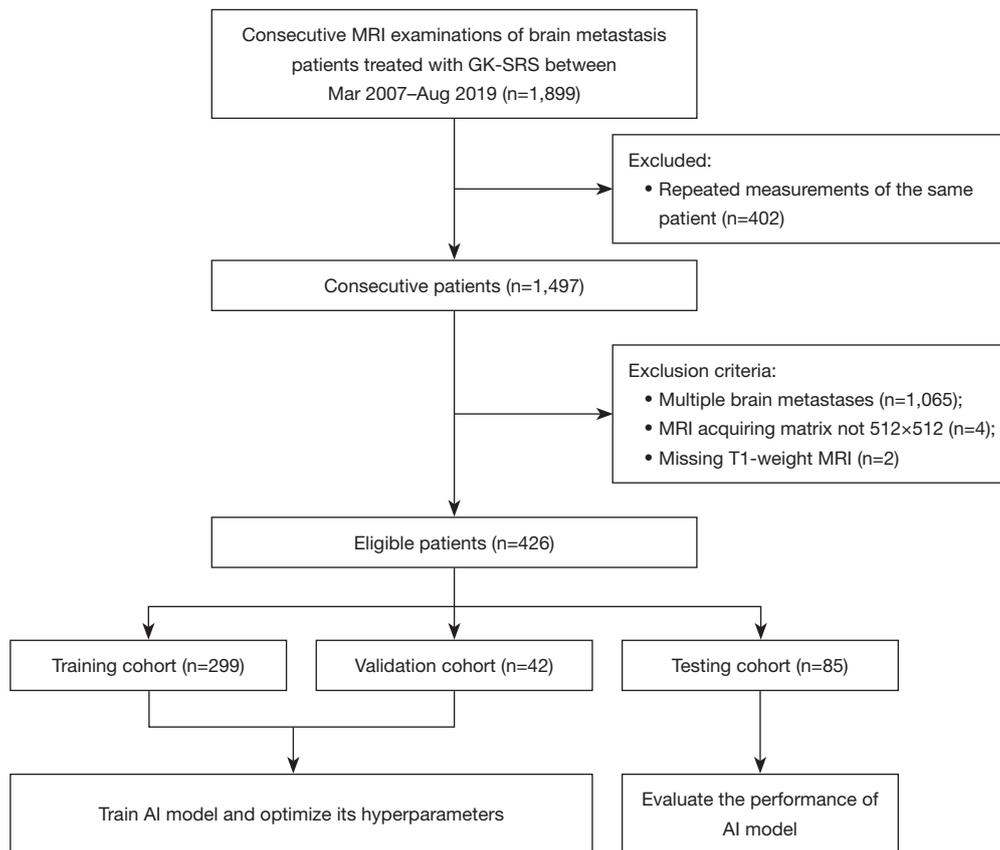
Figure 1 Study flowchart. MRI, magnetic resonance imaging; GK-SRS, Gamma Knife stereotactic radiosurgery; AI, artificial intelligence.

on T1-weighted MRI was first manually delineated by 2 experienced GK surgeons (Chen J with 20 years of experience in GK planning and Wang XY with 15 years in experience of GK planning) via consensus. We chose 3 adjacent slices showing the largest area of metastasis from each patient. All manual segmentations were performed using GK planning software (Leksell Gamma Plan, Elekta Group, Stockholm, Sweden), and each axial slice was manually segmented in turn. After manual segmentation, these MRI examinations were used to train and test the AI model.

Automated delineation

We developed a high-order attention (HA)-based CNN to extract representative features for the brain metastases in postcontrast T1-weighted MRI. Specifically, our network is based on the 2-dimensional (2D) CNN architecture of DeepLabV3+, which follows a typical encoder-decoder design (21), with a HA module inserted between the encoder

and decoder (22). The HA module has adaptive receptive fields with dynamic weights, which can be applied to efficiently capture in-plane features. Since through-plane features are also crucial for contouring, our network was designed as a 2.5-dimensional (2.5D) architecture via the stacking of 3 adjacent slices into the 3 channels as its input. The output was the delineation result of the middle slice. The detailed architecture of our network is shown in *Figure 2*.

Network training

The minibatch size was fixed to 16, Adam optimizer was used to train the network (23), and the initial learning rate was set to 0.001. Before being input into the model, the original images were center-cropped to a size of 420×420. Data augmentation is a useful means to alleviating the problem of overfitting. During the training process, random rotation with a range of 0 to 45° was applied to the input image to augment the data. The training process was terminated if the validation loss did not improve after
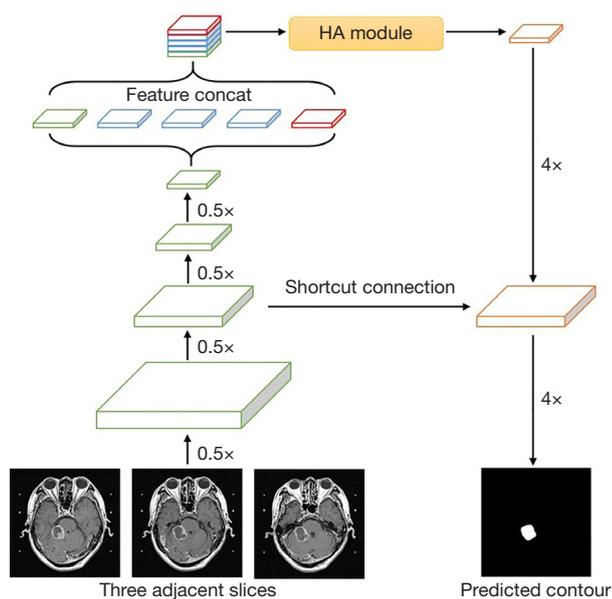
Figure 2 The proposed 2.5D CNN with high-order attention module for single brain metastasis segmentation from postcontrast T1-weighted MRI. The green and orange boxes denote the feature maps output by the convolutional layer in the encoder and decoder, respectively. The blue box represents the feature maps output by the dilated convolutional layer, and the red box is the feature maps output by the global average pooling layer. The numbers beside the vertical arrows denote the up and down scales of the sizes of the features. HA, high order attention; concat, concatenation; 2.5D, 2.5-dimensional; CNN, convolutional neural network; MRI, magnetic resonance imaging.

10 epochs. The developed AI model was configured based on the PyTorch deep learning framework using Python (Python Software Foundation, Wilmington, DE, USA) (24). The backbone was determined via a heuristic search. All experiments were performed on a Linux operating system workstation with a CPU Intel Xeon E5-2620 v3 at 2.4 GHz, 3 Nvidia Tesla P100 GPUs, and 64 GB of RAM.

AI model performance evaluation

The performance of the developed AI model was evaluated in the testing cohort (n=85). For objective evaluation, we adopted 2 quantitative evaluation metrics: (I) DSC was used to evaluate the volume of overlap between 2 contours according to the following formula (25): $DSC = 2 \times \dfrac{|V_S \cap V_G|}{V_S \cup V_G}$, where V_S and V_G denote the volume of the model's

segmentation results and ground truth, respectively. (II) The average surface distance (ASD) was used as a measure of the average distance between the surfaces of 2 contours according to the following formula (26):

$$ASD = \frac{1}{2}\left(\underset{i\in Gt}{mean}\ \underset{i\in Auto}{min}\ d(i,j) + \underset{i\in Auto}{mean}\ \underset{i\in Gt}{min}\ d(i,j)\right),\ where\ d(i,j)$$

denotes the Euclidean distance between voxel i and voxel j measured in millimeters.

Additionally, we compared the following subgroups according to the above-described indices: metastasis shape (regular or irregular), metastasis location (supratentorial or infratentorial), ring-shaped reinforcement (present or absent), and peritumoral edema (present or absent) (27).

Statistical analyses

All statistical comparisons were performed using scikit-learn in Python (28). Descriptive statistics were analyzed using the independent samples *t*-test and the Mann-Whitney test (if data were not normally distributed). Pearson chi-squared test was applied for categorical data. A 2-tailed P value <0.05 was considered statistically significant.

Target delineation evaluation

In addition to quantitative evaluation, we further evaluated the applicability of the AI tool in clinical practice. Specifically, 2 experienced GK surgeons (Chen J and Wang XY) were assigned with grading each AI-generated contour via consensus according to the following criteria: 1 = no revision (the contour is flawless and completely acceptable for treatment), 2 = minor revision (the contour requires a few minor edits but will no significant clinical impact without correction), and 3 = major revision (the contour requires significant revision before treatment can proceed).

Subsequently, a professional GK surgeon committee consisting of 3 experts (Zhao JY, Zeng YM, and Chen W with 4, 5, and 8 years of experience in GK planning, respectively) was invited to perform further evaluations. First, 40 randomly sampled MRI examinations from the testing cohort were assigned to the 3 qualified surgeons for manual contouring. Then, the AI-generated contours of these 40 MRI examinations were distributed to the surgeons for editing after a minimum interval of 1 month. The GK surgeons were blinded to the ground truth contours, their manual contours, and those performed by their counterparts. DSC and ASD were employed to evaluate the contouring accuracy, and times taken for each step were also reported.

6728

Zhao et al. An automatic tissue delineation model for brain metastasis

Table 1 Patient characteristics and indications on postcontrast T1-weighted MRI

Parameter	Entire cohort	Training/validation cohort	Testing cohort	P value
No. of patients	426	341	85	–
Sex				0.20
Male	247 (58.0)	192 (56.3)	55 (64.7)	
Female	179 (42.0)	149 (43.7)	30 (35.3)	
Age (years)	57.47±11.10	57.60±11.43	56.94±9.71	0.66
Tumor volume (mL)	6.71±9.54	6.90±10.11	5.93±6.80	0.84
Tumor size (mm)	17.93±8.32	17.99±8.59	17.72±7.16	0.86
Metastasis shape				0.36
Regular	338 (79.3)	267 (78.3)	71 (83.5)	
Irregular	88 (20.7)	74 (21.7)	14 (16.5)	
Metastasis location				0.51
Supratentorial	353 (82.9)	280 (82.1)	73 (85.9)	
Infratentorial	73 (17.1)	61 (17.9)	12 (14.1)	
Ring-shaped reinforcement				0.63
Present	223 (52.3)	176 (51.6)	47 (55.3)	
Absent	203 (47.7)	165 (48.4)	38 (44.7)	
Peritumoral edema				0.66
Present	184 (43.2)	145 (42.5)	39 (45.9)	
Absent	242 (56.8)	196 (57.5)	46 (54.1)	
Primary tumor				0.91
Lung	310 (72.8)	244 (71.6)	66 (77.6)	
Breast	35 (8.2)	30 (8.8)	5 (5.9)	
Renal	16 (3.8)	13 (3.8)	3 (3.5)	
Colon	10 (2.3)	8 (2.3)	2 (2.4)	
Other known primary	24 (5.6)	20 (5.9)	4 (4.7)	
Unknown primary	31 (7.3)	26 (7.6)	5 (5.9)	

Data are either number of patients, with the percentage in parentheses, or average, with the standard deviation. P values were calculated using the χ^2 for categorical variables and the Mann-Whitney test for numeric variables. A 2-tailed P<0.05 indicated a significant difference. MRI, magnetic resonance imaging.

Results

A total of 426 patients were included in the study, with the flowchart of patient inclusion being shown in *Figure 1* and the patients' characteristics being shown in *Table 1*. No significant differences were found regarding sex, age, tumor volume, tumor size, metastasis shape, metastasis location, presence of ring-shaped reinforcement, presence of peritumoral edema, and primary tumor between the training-validation cohort and the testing cohort.

The quantitative evaluation results are summarized in *Table 2*. Compared to the ground truth contours, the AI tool had a mean DSC score of 88.34% [standard deviation (SD) 5.00%] and a mean ASD of 0.35 mm (SD 0.21). *Figure 3* shows the illustrative examples of the best, median, and worst segmentation results using the AI tool.

Table 2 Performance of the AI tool in the testing cohort

| Groups	DSC (%)		ASD (mm)	
	Mean ± SD	P value	Mean ± SD	P value
Total (n=85)	88.34±5.00	–	0.35±0.21	–
Metastasis shape		0.01		0.01
Regular (n=71)	88.98±4.64		0.33±0.21	
Irregular (n=14)	85.07±5.45		0.45±0.19	
Metastasis location		0.38		0.32
Supratentorial (n=73)	88.61±4.81		0.33±0.16	
Infratentorial (n=12)	86.69±5.76		0.45±0.38	
Ring-shaped reinforcement		<0.01		0.17
Present (n=47)	90.16±4.26		0.36±0.17	
Absent (n=38)	86.08±4.92		0.33±0.25	
Peritumoral edema		<0.01		0.57
Present (n=39)	90.52±4.10		0.35±0.15	
Absent (n=46)	86.49±4.95		0.35±0.24	

P values were calculated using the Mann-Whitney test. A 2-tailed P<0.05 indicated a significant difference. AI, artificial intelligence; DSC, Dice similarity coefficient; ASD, average surface distance; SD, standard deviation.

In these examples, the autosegmented contours using the AI tool were close to the ground truth contours although inconsistencies were present. These results indicate an excellent concordance between our AI model and human experts for brain metastasis contouring.

In the subgroup analyses (*Table 2*), the AI achieved a comparable performance with the metastasis location in terms of mean DSC (supratentorial area: 88.61%; infratentorial area: 86.69%; P=0.38) and mean ASD (supratentorial area: 0.33 mm; infratentorial area: 0.45 mm; P=0.32). However, significant differences were observed between the type of metastasis shape in terms of mean DSC (regular: 88.98%; irregular: 85.07%; P=0.01) and mean ASD (regular: 0.33 mm; irregular: 0.45 mm; P=0.01). For different enhancement shapes, the AI tool achieved a significantly larger mean DSC in the ring-shaped enhancement tumors than in the none ring-shaped enhancement tumors (90.16% *vs.* 86.08%; P<0.01). There was also a significant difference mean DSC between the tumors with peritumoral edema and tumors without peritumoral edema (90.52% *vs.* 86.49%; P<0.01). In contrast, we did not observe differences for ASD between the different enhancement shapes and the tumors with edema and tumors without peritumoral edema.

The predicted contours of 572 slices from 85 patients from the test set were subjectively evaluated by the GK surgeon committee (Wang XY and Chen J). We chose 3–9 slices from patients depending on the size of metastasis. The majority (514/572, 89.9%) of the AI-generated contours were evaluated as "No revision" (415/572, 72.6%) or "Minor revision" (99/572, 17.3%). Only 58 slices were assessed as requiring "Major revision".

We compared the AI-generated contours with those of 3 qualified GK surgeons (S1: Zhao JY; S2: Zeng YM; S3: Chen W), with the ground truth contours being delineated by the 2 experienced GK surgeons (Chen J and Wang XY). In terms of DSC, the AI tool outperformed 2 of the 3 experts (AI tool: 88.09%; S1, S2 and S3: 85.92%, 86.32% and 86.75%; all P<0.05) and achieved comparable results to the other one. In terms of ASD, our AI tool outperformed 1 of the 3 experts (mean ASD: 0.37 *vs.* 0.45 mm; P=0.03) and achieved comparable results to the other 2. Next, we evaluated the effectiveness and efficiency of our AI tool for assisting manual contouring, and the results are presented in *Table 3* and *Figure 4*. We observed that 2 of 3 radiologists achieved better performance under AI assistance compared with manual contouring only in terms of mean DSC (S1: 88.00% *vs.* 85.92%, P<0.05; S3: 89.18% *vs.* 86.75%,
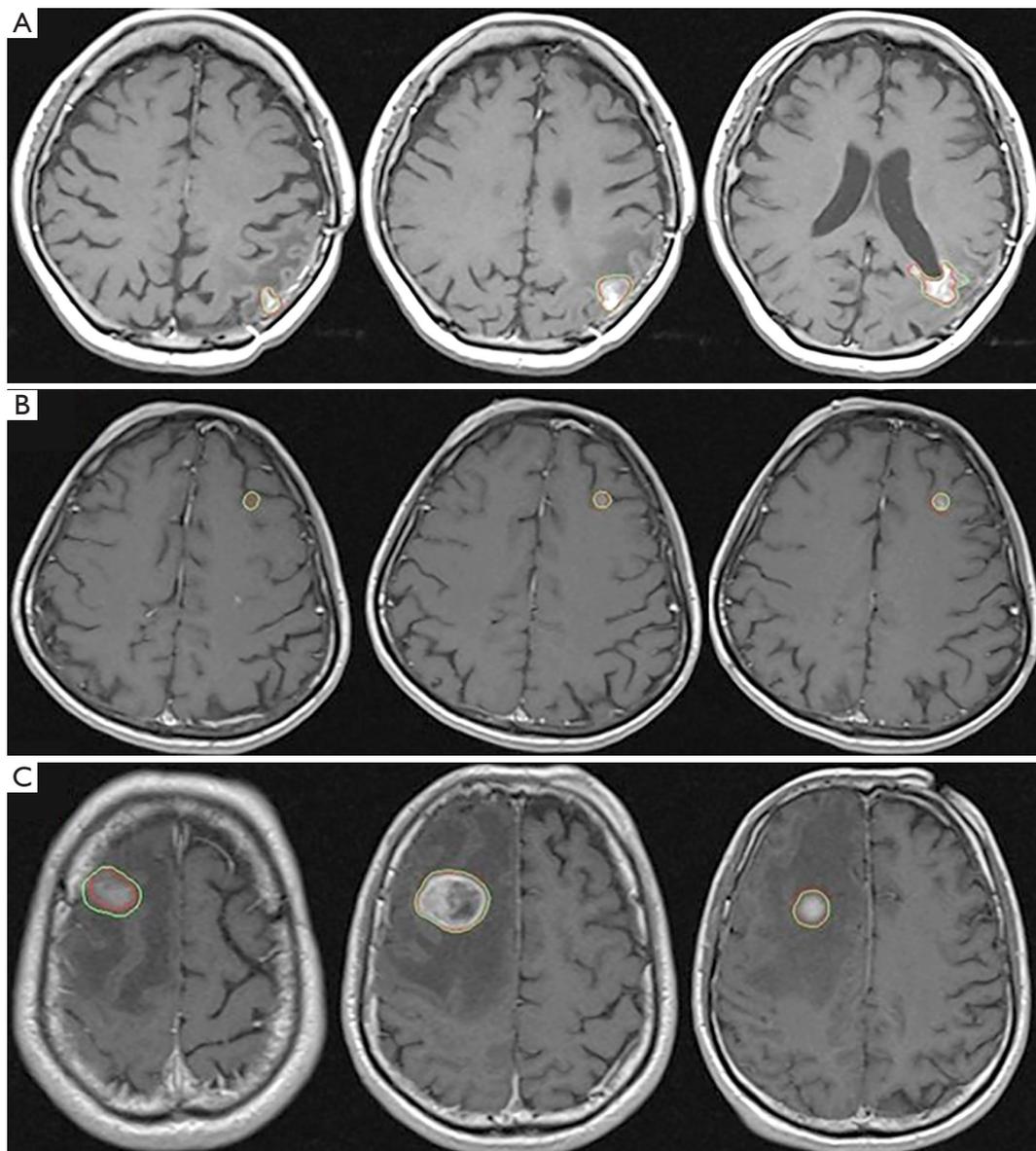
Figure 3 Illustrative contouring examples of the AI tool and human experts through the lower, middle, and upper 3-dimensional sections within the tumor. Red contours denote the human experts delineated ground truth, and the green contours denote the AI-generated contours. MRIs were obtained in patients with Dice similarity coefficients of 0.77 (A), 0.85 (B), and 0.96 (C). MRI, magnetic resonance imaging; AI, artificial intelligence.

$P<0.05$) and mean ASD (S1: 0.38 *vs.* 0.45, $P<0.05$; S3: 0.32 *vs.* 0.40, $P<0.05$). The main advantage of AI assistance is that it can greatly save contouring time. The average time spent editing an AI-generated contour compared with the time spent with manual contouring was 74 *vs.* 196 seconds ($P<0.01$), corresponding to a time savings of 60% work hours.

Discussion

In this study, an AI contouring tool was developed for single brain metastasis segmentation on a large set of postcontrast T1-weighted MRI examinations from 426 patients, and its ability to delineate single brain metastasis was validated via a comparison against qualified GK surgeons.

Table 3 Comparison of manual delineation with AI-assisted delineation on 3 qualified GK surgeons. Forty patients were randomly sampled from the testing cohort for this comparison

Variables	DSC (%)		ASD (mm)		Time (s)	
	Mean ± SD	P value	Mean ± SD	P value	Mean ± SD	P value
AI tool	88.09±5.12	–	0.37±0.16	–	–	–
S1		<0.01		0.01		<0.01
Manual delineation	85.92±6.08		0.45±0.23		212.78±102.65	
AI-assisted delineation	88.00±5.23		0.38±0.23		77.85±54.00	
S2		0.30		0.20		<0.01
Manual delineation	86.32±6.20		0.44±0.24		184.45±65.54	
AI-assisted delineation	87.74±5.17		0.37±0.18		62.80±43.92	
S3		0.03		0.03		<0.01
Manual delineation	86.75±6.65		0.40±0.22		189.93±62.92	
AI-assisted delineation	89.18±3.21		0.32±0.15		82.55±38.33	

P values were calculated using the Wilcoxon signed-rank test. A 2-tailed P<0.05 indicated a significant difference. AI, artificial intelligence; GK, Gamma Knife; DSC, Dice similarity coefficient; ASD, average surface distance; SD, standard deviation.
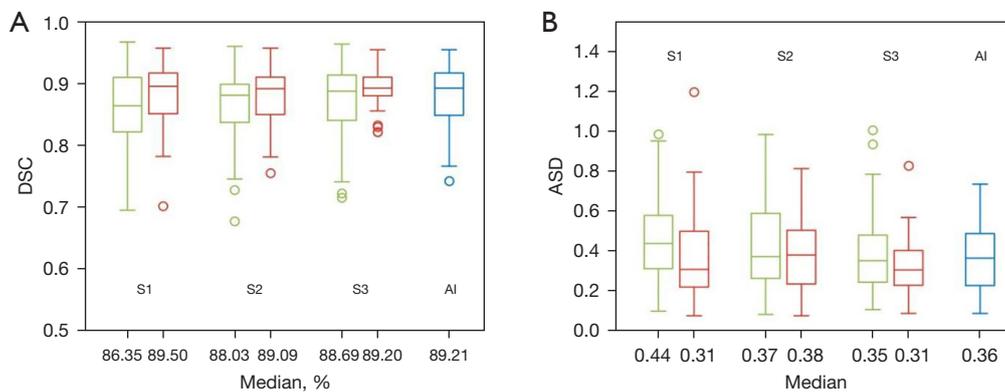
Figure 4 Clinical evaluation results of 3 qualified GK surgeons. (A) DSC distributions of manual contours, post-AI-assisted contours, and AI-generated contours. (B) ASD distributions of manual contours, post-AI-assisted contours, and AI-generated contours. The green box denotes the result between the manual contours of radiologists and the ground truth contours. The red box represents the result between the post-AI-assisted contours of the radiologists and the ground truth contours. The blue box denotes the result between the AI-generated contours of the radiologists and the ground truth contours. DSC, Dice similarity coefficient; ASD, average surface distance; GK, Gamma Knife; AI, artificial intelligence.

The developed AI tool achieved excellent results, showing a mean DSC score of 88.34% (SD 5.00%) and a mean ASD of 0.35 mm (SD 0.21). In the subgroup analysis, we found that the metastasis shape, presence of ring-shaped reinforcement, and presence of peritumoral edema highly associated the performance of the AI tool (all P values <0.05). The AI tool can be easily used to segment metastases with a regular shape (round or oval) but showed limited performance for metastases with an irregular shape in both the DSC and ASD assessment. This can be explained by the fact that metastases with an irregular shape, characterized by rugged borders without order and interlacing with peritumoral edema, are difficult for neural networks to learn features from. Moreover, as the training of neural networks is data driven, the small number of patients with irregular metastasis shapes also

might have contributed to these results. We plan to collect more data related to irregular metastasis shape to validate this finding. Experimental results also indicated that the segmentation performance highly depended on the reinforcement shape and the presence of peritumoral edema in the DSC assessment but not in ASD assessment. This probably occurred because in metastases with ring-shaped enhancement or edema, it is easier for GK surgeons to mark the boundary precisely, leading to less diffusion and much clearer boundaries. With more accurate boundary markers, the DCNN can learn the features more accurately. Although the marking precision between metastasis with and without ring-shaped enhancement or edema differed, the AI tool showed no difference in performance in terms of the ASD assessment. This is likely because the assessment accuracy of ASD was lower than that of DSC for this task.

Brain metastasis contouring is tedious, laborious, and time-consuming for GK planning. Thresholding is an additional tool that may help an experienced user to make the segmentation of brain metastases fairly rapid and relatively uniform depending on the thresholding parameters. However, it is not wildly used in clinical platforms. Several studies have paid examined autosegmenting brain metastases, but there was no clinical evaluation performed in these studies (16-20). Quantitative evaluation metrics can only evaluate the overall performance, treating every pixel with the same importance, which is different from clinical practice. Thus, there remains questions concerning the clinical applicability of these studies in the real world. In this study, the AI tool results in the testing cohort were subjectively evaluated by 2 experienced physicians via consensus. Approximately 89.9% of predicted contours were evaluated as "No revision" or "Minor revision". In addition, our AI tool outperformed 2 of the 3 qualified physicians (both P values <0.05). By allowing the 3 physicians to delineate brain metastases based on the contours generated initially with our AI tool, the contouring time was reduced by 60%, with the contouring accuracy being improved. The resulting time savings could also be useful for streamlining therapeutic strategies that require timely contouring interventions, such as in neoadjuvant SRS (29). These findings suggest that our AI tool can be useful in performing single brain metastasis contouring for GK planning in clinical practice.

This study presents several unique aspects compared with previous studies for automatic brain metastasis delineation. First, on this study, we analyzed the effect of the peritumoral edema on segmentation performance, and our dataset contained not only regular (round or oval) but also irregular metastasis shapes, which is more reflective of real-world patients. Experimental results showed that both metastasis shape and peritumoral edema significantly associate segmentation performance (all P values <0.05).

Several limitations to our study should be noted. First, the dataset used was collected from a single institution, and thus additional data from multiple institutions and multiple manufacturers should be collected to validate the developed AI tool. Second, as shown in *Table 1*, the majority of brain metastases included in our study were regular in shape, located in the supratentorial region, and originated from lung tumors. This may limit the external realism and generalizability of our AI tool to brain metastases with different shapes, locations, and primary tumors. Future studies with larger and more diverse datasets are needed to validate the performance of our AI tool and expand its application to a wider range of brain metastases. Furthermore, this study was performed on single brain metastasis, and the development of an autosegmenting tool for multiple brain metastases is still required. A fairly recent study reported a PPV of 58%, a sensitivity of 85%, and a DSC of 0.85 of the entire segmentation mask for each patient, which may not be helpful in clinical practice (16). We will focus on validating the utility of AI for multiple brain metastasis contouring in clinical applications, especially given the increasingly common trend of SRS being preferred over whole brain radiotherapy even in this scenario (30,31).

Conclusions

In this study, we developed an AI model to automate tissue delineation for single brain metastasis of the GK plan. The AI model demonstrated a high contouring accuracy comparable to that of human experts. The contouring results of AI model were evaluated by human experts, and most were deemed to be directly usable in clinical practice without need of revision. Thus, the AI model can effectively reduce contouring time and aid in treatment planning of single brain metastasis for SRS. Future work will focus on the development and validation of the automatic contouring tool for multiple brain metastases.

Acknowledgments

Innovation Research and Development Projects in Chengdu (No. 2019-YF05-00333-SN) and the Clinical Research Incubation Project of Subject Excellence Development in West China Hospital of Sichuan University (No. 18HXFH010).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-22-1216/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1216/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committee of West China Hospital of Sichuan University. The study was registered at clinical trial registration URL: https://www.chictr.org.cn (unique identifier: ChiCTR2100046265). Individual consent for this retrospective analysis was waived since patients were selected retrospectively and the MRI were completely anonymized before analysis.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Patchell RA. The management of brain metastases. Cancer Treat Rev 2003;29:533-40.
2. Yamamoto M, Serizawa T, Shuto T, Akabane A, Higuchi Y, Kawagishi J, et al. Stereotactic radiosurgery for patients with multiple brain metastases (JLGK0901): a multi-institutional prospective observational study. Lancet Oncol 2014;15:387-95.
3. Perlow HK, Dibs K, Liu K, Jiang W, Rajappa P, Blakaj DM, Palmer J, Raval RR. Whole-Brain Radiation Therapy Versus Stereotactic Radiosurgery for Cerebral Metastases. Neurosurg Clin N Am 2020;31:565-73.
4. Nussbaum ES, Djalilian HR, Cho KH, Hall WA. Brain metastases. Histology, multiplicity, surgery, and survival. Cancer 1996;78:1781-8.
5. Delattre JY, Krol G, Thaler HT, Posner JB. Distribution of brain metastases. Arch Neurol 1988;45:741-4.
6. Kocher M, Soffietti R, Abacioglu U, Villà S, Fauchon F, Baumert BG, Fariselli L, Tzuk-Shina T, Kortmann RD, Carrie C, Ben Hassel M, Kouri M, Valeinis E, van den Berge D, Collette S, Collette L, Mueller RP. Adjuvant whole-brain radiotherapy versus observation after radiosurgery or surgical resection of one to three cerebral metastases: results of the EORTC 22952-26001 study. J Clin Oncol 2011;29:134-41.
7. Hu J, Xie X, Zhou W, Hu X, Sun X. The emerging potential of quantitative MRI biomarkers for the early prediction of brain metastasis response after stereotactic radiosurgery: a scoping review. Quant Imaging Med Surg 2023;13:1174-89.
8. Levivier M, Gevaert T, Negretti L. Gamma Knife, CyberKnife, TomoTherapy: gadgets or useful tools? Curr Opin Neurol 2011;24:616-25.
9. Topkan E, Kucuk A, Senyurek S, Sezen D, Durankus NK, Akdemir EY, Akdemir EY, Saglam Y, Bolukbasi Y, Pehlivan B, Selek U. Radiosurgery Techniques for Brain Metastases. Journal of Cancer and Tumor International 2020;10:1-14..
10. McGrath H, Li P, Dorent R, Bradford R, Saeed S, Bisdas S, Ourselin S, Shapey J, Vercauteren T. Manual segmentation versus semi-automated segmentation for quantifying vestibular schwannoma volume on MRI. Int J Comput Assist Radiol Surg 2020;15:1445-55.
11. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 2017;19:221-48.
12. Shapey J, Wang G, Dorent R, Dimitriadis A, Li W, Paddick I, Kitchen N, Bisdas S, Saeed SR, Ourselin S, Bradford R, Vercauteren T. An artificial intelligence framework for automatic segmentation and volumetry of vestibular schwannomas from contrast-enhanced T1-weighted and high-resolution T2-weighted MRI. J Neurosurg 2019;134:171-9.
13. Bello GA, Dawes TJW, Duan J, Biffi C, de Marvao A, Howard LSGE, Gibbs JSR, Wilkins MR, Cook SA, Rueckert D, O'Regan DP. Deep learning cardiac motion analysis for human survival prediction. Nat Mach Intell 2019;1:95-104.

14. Wu B, Zhang F, Xu L, Shen S, Shao P, Sun M, Liu P, Yao P, Xu RX. Modality preserving U-Net for segmentation of multimodal medical images. Quant Imaging Med Surg 2023;13:5242-57.

15. Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, De Fauw J, et al. Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. J Med Internet Res 2021;23:e26151.

16. Zhou Z, Sanders JW, Johnson JM, Gule-Monroe M, Chen M, Briere TM, Wang Y, Son JB, Pagel MD, Ma J, Li J. MetNet: Computer-aided segmentation of brain metastases in post-contrast T1-weighted magnetic resonance imaging. Radiother Oncol 2020;153:189-96.

17. Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. J Magn Reson Imaging 2020;51:175-82.

18. Bousabarah K, Ruge M, Brand JS, Hoevels M, Rueß D, Borggrefe J, Große Hokamp N, Visser-Vandewalle V, Maintz D, Treuer H, Kocher M. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. Radiat Oncol 2020;15:87.

19. Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, Liu X, Chen L, Qu J, Xu S, Tang X, Mao Y, Liu Y, Li D. Deep learning-based detection and segmentation-assisted management of brain metastases. Neuro Oncol 2020;22:505-14.

20. Di Ieva A, Russo C, Liu S, Jian A, Bai MY, Qian Y, Magnussen JS. Application of deep learning for automatic segmentation of brain tumors on magnetic resonance imaging: a heuristic approach in the clinical scenario. Neuroradiology 2021;63:1253-62.

21. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y. editors. Computer Vision – ECCV 2018. Lecture Notes in Computer Science, vol 11211. Cham: Springer; 2018.

22. Ding F, Yang G, Wu J, Ding D, Xv J, Cheng G, Li X. High-Order Attention Networks for Medical Image Segmentation. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, Racoceanu D, Joskowicz L. editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Lecture Notes in Computer Science, vol 12261. Cham: Springer; 2020.

23. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [Preprint]. 2014. Available online: https://arxiv.org/abs/1412.6980

24. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A. Automatic differentiation in PyTorch. 31st Conference on Neural Information Processing Systems (NIPS 2017); Long Beach, CA, USA. 2017.

25. Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology. 1945;26:297-302.

26. Yousefi S, Kehtarnavaz N, Gholipour A. Improved labeling of subcortical brain structures in atlas-based segmentation of magnetic resonance images. IEEE Trans Biomed Eng 2012;59:1808-17.

27. Stummer W. Mechanisms of tumor-related brain edema. Neurosurg Focus 2007;22:E8.

28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. ArXiv. 2011;abs/1201.0490.

29. Palmisciano P, Ferini G, Khan R, Bin-Alamer O, Umana GE, Yu K, Cohen-Gadol AA, El Ahmadieh TY, Haider AS. Neoadjuvant Stereotactic Radiotherapy for Brain Metastases: Systematic Review and Meta-Analysis of the Literature and Ongoing Clinical Trials. Cancers (Basel) 2022;14:4328.

30. Ferini G, Viola A, Valenti V, Tripoli A, Molino L, Marchese VA, Illari SI, Borzì GR, Prestifilippo A, Umana GE, Martorana E, Mortellaro G, Ferrera G, Cacciola A, Lillo S, Pontoriero A, Pergolizzi S, Parisi S. Whole Brain Irradiation or Stereotactic RadioSurgery for five or more brain metastases (WHOBI-STER): A prospective comparative study of neurocognitive outcomes, level of autonomy in daily activities and quality of life. Clin Transl Radiat Oncol 2022;32:52-8.

31. Pergolizzi S, Cacciola A, Parisi S, Lillo S, Tamburella C, Santacaterina A, Ferini G, Cellini F, Draghini L, Trippa F, Arcidiacono F, Maranzano E. An Italian survey on "palliative intent" radiotherapy. Rep Pract Oncol Radiother 2022;27:419-27.