Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-1858

Reviewer A

The author writes an article on interesting topic Segmentation method of magnetic resonance imaging brain tumor images based on improved U-Net network. However, the paper doesn't have any section for related work. The flow of proposed methodology is not according.

Response: Thanks very much for this suggestion. We had Method section to describe our related work. And we thought our result followed the flow of proposed methodology. Please check them again. Thanks very much again.

Reviewer B

- Abstract should be improved and extended. The authors talk lot about the problem formulation, but novelty of the proposed model is missing. Also provided the general applicability of their model. Please be specific what are the main quantitative results to attract general audiences

Response: Thanks very much for this suggestion. We added "The result showed that we developed a 2D residual block UNet, which can improve the incorporation of glioma segmentation into the clinical process." in the conclusion section.

- The introduction can be improved. The authors should focus on extending the novelty of the current study. Emphasize should be given in improvement of the model (in quantitative sense) compared to existing state-of the art models.

Response: Thanks very much for this suggestion. We added that "In the evaluation outcomes of the Brain Tumor Image Segmentation Benchmark (BraTS) validation set, both Zikic et al. (35) used a single-branch 2D Conv neural network based on pixel-by-pixel segmentation, yet Pereira et al. (36) used a dual-branch pixel-by-pixel segmentation network. Meanwhile, Naceur et al. (37) used 3 full convolutional networks for segmentation of gliomas, including a 2D full convolutional network and a 3D full convolutional network and fused the results by integrated learning techniques. Li et al. (38) conducted quantitative and qualitative experiments on the BraTS 2020 dataset to evaluate the performance of the Evidential Deep Learning model for segmentation and uncertainty estimation. The results show that this method has excellent performance in quantitative segmentation of uncertainty and robust segmentation of tumors (38). The publicly available training dataset provided for the 2021 RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge was used in Boehringer's study, consisting of 1251 multi-institutional, multi-parametric MR images (39). Their result showed that the active learning approach when applied to model training can drastically reduce the time and labor spent on preparation of ground truth training data (39).".

- More details about network architecture and complexity of the model should be provided

Response: Thanks very much for this suggestion. We had a lot of details in ###Improvement of the UNet model and ###Structure of the proposed 2DResUNet and ###Loss function and so on. These are the details about how we built the models.

- What about comparison of the result with current state-of the art models? Did authors perform ablation study to compare with different models?

Response: Thanks very much for this suggestion. We had comparison, please check Discussion section. And we had some description in the introduction. For instance, In the evaluation outcomes of the Brain Tumor Image Segmentation Benchmark (BraTS) validation set, both Zikic et al. (35) used a single-branch 2D Conv neural network based on pixel-by-pixel segmentation, yet Pereira et al. (36) used a dual-branch pixel-by-pixel segmentation network. Meanwhile, Naceur et al. (37) used 3 full convolutional networks for segmentation of gliomas, including a 2D full convolutional network and a 3D full convolutional network and fused the results by integrated learning techniques. Li et al. (38) conducted quantitative and qualitative experiments on the BraTS 2020 dataset to evaluate the performance of the Evidential Deep Learning model for segmentation and uncertainty estimation. The results show that this method has excellent performance in quantitative segmentation of uncertainty and robust segmentation of tumors (38). The publicly available training dataset provided for the 2021 RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge was used in Boehringer's study, consisting of 1251 multi-institutional, multi-parametric MR images (39). Their result showed that the active learning approach when applied to model training can drastically reduce the time and labor spent on preparation of ground truth training data (39). We did not perform ablation study.

- What are the baseline models and benchmark results? The authors can compared the result with existing models evaluated with datasets

Response: Thanks very much for this suggestion. We had it "In the evaluation outcomes of the Brain Tumor Image Segmentation Benchmark (BraTS) validation set, both Zikic et al. (35) used a single-branch 2D Conv neural network based on pixel-by-pixel segmentation, yet Pereira et al. (36) used a dual-branch pixel-by-pixel segmentation network. Meanwhile, Naceur et al. (37) used 3 full convolutional networks for segmentation of gliomas, including a 2D full convolutional network and a 3D full convolutional network and fused the results by integrated learning techniques. Li et al. (38) conducted quantitative and qualitative experiments on the BraTS 2020 dataset to evaluate the performance of the Evidential Deep Learning model for segmentation and uncertainty estimation. The results show that this method has excellent performance in quantitative segmentation of uncertainty and robust segmentation of tumors (38). The publicly available training dataset provided for the 2021 RSNA-ASNR-MICCAI

Brain Tumor Segmentation (BraTS) Challenge was used in Boehringer's study, consisting of 1251 multi-institutional, multi-parametric MR images (39). Their result showed that the active learning approach when applied to model training can drastically reduce the time and labor spent on preparation of ground truth training data (39).".

- Please provide a fair weakness and limitation of the model, and how it can be improved.
Response: Thanks very much for this suggestion. We had it in our manuscript "Despite the novel findings, our study has several limitations that should be considered: The 2D networks still have the problem of losing some of the spatial information of 3D images. For instance, the batch size was limited by Graphics Processing Unit (GPU) memory during training and evaluation. We will continue to ameliorate the web based on the difficulties of the 2D network and the issue of too many 3D network parameters to improve the segmentation effect.".

- Provide flowcharts and pseudo codes of the proposed algorithm in detail.
Response: Thanks very much for this suggestion. Please check Figures 2-5, 7.

Reviewer C

The work is interesting and should be considered by the field experts.
I would liked to find embedded links for the reader and users to reproduce or replicate the results and/or re-apply to new data.
Response: Thanks very much for this suggestion. We will provide the embedded links when someone is interested in our work.

Reviewer D

The article presents a method of semantic segmentation of MRI data in brain tumor cases using 2DResUNet architecture. The article offers a very detailed technical description of the method. However, we feel that from the clinical perspective, the article is not very clear and benefits and limitations of the method are not communicated clearly. In the current version, we don't see how it could benefit clinical practice. We believe that implementing major revisions could improve the quality of the article substantially.

Our main objections is:
The article is very inaccessible for a clinical reader without excessive technical background.
No clinical usefulness or discussion of clinical applicability.
Response: Thanks very much for this suggestion. Our work can offer important reference value

for the degree grading of glioma, provides scientific basis for the accuracy and sensitivity of tumor diagnosis, and is worthy of clinical popularization and application. In addition, the method based on the improved UNet network has obvious advantages in MRI brain tumor image segmentation and has stronger feature extraction ability than the earlier UNet model. Moreover, it can effectively alleviate the class imbalance of glioma data and effectively segment glioma. Our findings demonstrated that the use of generalized Dice loss and subjective cross entropy as loss functions in the training process effectively alleviated the class imbalance of glioma data and effectively segmented glioma.

Comments:

1) ln 72 - "Glioma is a 5-fold cross-validation evaluation results due to the influence of ionizing" not clear.

Response: Thanks very much for this suggestion. We have modified that "According to a 5-fold cross-validation evaluation results (1,2), Glioma is accounting for 27% of primary central nervous system tumors (3,4,5) due to the influence of ionizing radiation and gene mutation."

2) ln 76 - should cite WHO classification, cited (9) Multilevel segmentation of intracranial aneurysms in CT angiography images.".

Response: Thanks very much for this suggestion. We did as your suggestion.

3) ln 78 - HGG with high malignancy - please explain

Response: Thanks very much for this suggestion. We have changed it to "HGG patients with high-grade gliomas".

4) When speaking about classification, explain, why it's important

Response: Thanks very much for this suggestion. We did some changes "The incidence of glioma increases with age (1,2,7,8). With different grades gliomas, there is different incidence rate. According to malignancy, glioma is pathologically classified into Grade I to Grade IV, among which Grade II and below are low-grade gliomas (LGGs), and Grade III and above are high-grade gliomas (HGGs) (9). For instance, the median survival time (MST) of HGG patients is generally less than 2 years, whereas the MST of HGG patients with high-grade gliomas is only 4–9 months.".

5) Magnetic resonance imaging (MRI) is the only magnetic imaging technology that can evaluate the biochemical and metabolic conditions of cells and tissues without trauma - requires explanation, this is not completely wrong, there are MR modalities to evaluate biochemical and metabolic features of tissue, the standard however is to evaluate the structure of tissue

Response: Thanks very much for this suggestion. We did some changes "Magnetic resonance imaging (MRI) is the only magnetic imaging technology that can evaluate the biochemical and metabolic conditions of cells and tissues without trauma by evaluating the structure of tissue.".

6) ln 84 - CT is rarely instrumental in preoperative planning of glioma cases, PET definitely not routinely, I would recommend mentioning of DTI

Response: Thanks very much for this suggestion. We mentioned DTI as your suggestion.

7) “Currently, glioma has been confirmed to consist of tumors in a variety of tissues” - reference? also glioma doesn’t consist of tumors, better expression would be tissue types

Response: Thanks very much for this suggestion. We have modified that “Currently, glioma has been confirmed to consist of tumors in a variety of tissues types”.

8) “Typically, different degrees of glioblastoma” - glioma and glioblastoma are not interchangeable terms, glioblastoma refers to WHO Grade IV glioma

Response: Thanks very much for this suggestion. We changed it to “different degrees of gliomas”.

9) Discuss the role and importance of segmentation (both manual and automatic). It is definitely not required and not a standard, explain, how it is beneficial

Response: Thanks very much for this suggestion. Concerning this question, we added that “Glioma segmentation is believed to be one of the most important stages of treatment management (12,13). Recent developments in MRI protocols have led to a renewed interest in using automatic glioma segmentation with different MRI image weights.”.

10) ln 110 “these techniques are often ineffective for large-scale and multimodal glioma datasets.” - why?

Response: Thanks very much for this suggestion. We modified it to “The reports on large-scale and multimodal glioma dataset using these techniques are rare.”.

11) when speaking of classification and grading, it’s essential to mention the role of histopathology and recently molecular pathology

Response: Thanks very much for this suggestion. We have added that “The presence of IDH1/2 mutation along with ATRX and TP53 mutations is indicative of diffuse astrocytomas, while IDH1/2 mutations combined with 1p19q loss is indicative of oligodendroglioma(43). Focal amplifications of receptor tyrosine kinase genes, TERT promoter mutation, and the loss of chromosomes 10 and 13 with trisomy of chromosome 7 are distinctive features of glioblastoma and can be utilized for diagnostic purposes(43). Additionally, the presence of BRAF gene fusions and mutations in low-grade gliomas and mutations in histone H3 in high-grade gliomas can also serve as diagnostic markers (43).”.

12) “We developed a 2D residual block UNet in order to improve the incorporation of glioma segmentation into the clinical process.” - explain how using a certain model helps with incorporation of a model into clinical workflow. What are the roadblocks for incorporation now?

Response: Thanks very much for this suggestion. For these two questions, we have answered in the manuscript. For first one, please check “The efficiency is seamlessly interrelated with using something called the Brain Tumor Image Segmentation Benchmark (BRATS). Twenty of the most advanced tumor segmentation algorithms were used on sixty-five multi-contrast MR scans of low- and high-grade glioma patients. As many as four raters had personally annotated these scans, and they were compared to sixty-five scans that were created using tumor image simulation software. It was specified that the Dice scores for WT, TC, and ET would be

provided, with expert segmentation serving as the benchmark. XXX Hospital, XXX University in Norway (“XXX” is masked text according to the double-blind peer review policy), provided a distinct clinical dataset that was utilized in the subsequent evaluation of the approach.”. For second one, please check “To address various problems associated with the traditional full convolutional networks such as shallow model depth and inadequate acquisition of image feature information, a full convolutional network model using the residual block (ResBlock) mechanism is proposed, which is based on the UNet model, for glioma segmentation.”

13) Inclusion criteria missing. Specify diagnoses.

Response: Thanks very much for this suggestion. We have added that “Inclusion criteria were followed: All patients with MRI images were glioma patients.”.

14) 156 Improvement of the UNet model - many unexplained terms used - residual block, convolutional network, shallow depth, unet - explain to a clinical nontechnical reader

Response: Thanks very much for this suggestion. Residual block: the residual block transfers information along the depth of the network by introducing skip connections that add the network output of the previous layer directly to the network input of the current layer. Convolutional network: The proposed network is a simple design that employs different heads involving graph convolutions focused on edges and nodes, capturing representations from the input data thoroughly. Shallow model: Single hidden layer neural network and single layer neural network (that is, logistic regression), usually a neural network with few hidden layers, such as only one or two layers, is called shallow model. ResUNet model: UNet with residual block helps to solve the gradient vanishing and gradient exploding problems and train deeper networks while ensuring good performance.

15) 166 single channel modified two four channels - this is not novelty as stated, has been done before, cite

Response: Thanks very much for this suggestion. We did.

16) 179 - explain how this reduces information loss

Response: Thanks very much for this suggestion. We have modified that “the Decoder ResBlock’ structure is used to replace the 2 convolutional layers, which effectively reduces the information loss during the up-sampling process due to this model incorporating a modified atrous spatial pyramid pooling module to learn the location information and to extract multi-level contextual information.”.

17) 182 used as the loss function - consider using more self explanatory terms (i.e. performance measure) to accompany the technical terminology

Response: Thanks very much for this suggestion. We have added that “Loss function is a function that maps the value of a random event or its related random variables to a non-negative real number to represent the "risk" or "loss" of that random event”.

18) 274 splitting approach - the overall workflow described here - in my opinion, this should come before explaining details of the model

Response: Thanks very much for this suggestion. We did.

19) 278 - explain what p

Response: Thanks very much for this suggestion. We have explained in the manuscript. p is the pixel value of the predicted outcome.

20) no clear definition of primary and secondary outcomes

Response: Thanks very much for this suggestion. Please check "##Experimental results". According to the description. We could know what primary and secondary outcomes are.

21) 373 we believe this should serve as results section.

Response: Thanks very much for this suggestion. We did some changes.

22) 376 local dataset divided by the BraTS2018 dataset with the ratio of test set – unclear

Response: Thanks very much for this suggestion. Please check "The Dice score of different loss functions evaluated by the validation set after 10 generations of training in a local dataset divided by the BraTS2018 dataset with the ratio of test set, validation set, and training set, in a ratio of 1:1:8 is shown in Table 1.".

23) results section fails to clearly communicate quantitative results defined in methods.

Response: Thanks very much for this suggestion. We think that results section includes quantitative results defined in methods.

24) Discussion - starts with existing literature, this would be better to move to introduction, unless differences to the proposed method and their influence on performance are discussed.

Response: Thanks very much for this suggestion. We did some changes for the Discussion section.

25) frequent use of terms like "excellent results" without defining criteria for them

Response: Thanks very much for this suggestion. We did not mention "excellent results". Please make it clear.

26) significantly improved in various indicators - what indicators? why are they important for clinical practice?

Response: Thanks very much for this suggestion. Indicators include Dice score, Specificity, Sensitivity, Hausdorff distance. Because when glioma is mentioned, people will mention them.

27) 556 - "this paper explores the potential of magnetic resonance imaging (MRI) brain tumor" - almost every method before did that.

Response: Thanks very much for this suggestion. We changed the description.

28) what are the criteria of a "good enough" model for clinical practice? define them and rate your outputs according to these criteria.

Response: Thanks very much for this suggestion. We did not mention "good enough". Please make it clear.

29) discuss how your model could be used by clinicians? is it an application? command line program? software integrated into already existing tools?

Response: Thanks very much for this suggestion. This is a program. The clinicians just need to add the data and set some parameters to use it.

Reviewer E

1. The information of Ref. 37 in the main text differed from the information in the reference list. Please revise.

Meanwhile, Naceur et al. (37) used 3 full convolutional networks for segmentation of gliomas, including a 2D full convolutional network and a 3D full convolutional network and fused the results by integrated learning techniques.

Ref. 37: Ben NM, Akil M, Saouli R, Kachouri R. Deep Convolutional Neural Networks for Brain tumor segmentation: boosting performance using deep transfer learning: prelim-inary results. En Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 5. Springer International Publishing, 2020, 303-315.

Response: Thanks very much for this suggestion. We have modified it as your suggestion.

2. The information of Ref. 27 in the main text differed from the information in the reference list. Please revise.

A commonly used UNet loss function is Dice loss, which was proposed by Milletari et al. (27).

Ref. 27: Kshirsagar PR, Manoharan H, Meshram P, Alqahtani J, Naveed QN, Islam S, Abebe TG. Recognition of Diabetic Retinopathy with Ground Truth Segmentation Using Fundus Images and Neural Network Algorithm. Comput Intell Neurosci 2022;2022:8356081.

Response: Thanks very much for this suggestion. We have modified it as your suggestion.

3. The authors mentioned "studies...", while no reference was cited. Please revise.

Furthermore, molecular studies have identified features that can enhance diagnosis and provide biomarkers.

Response: Thanks very much for this suggestion. We have cited it as your suggestion in the main text.

4. Tables should be cited **consecutively** in the text and numbered in the order in which they are discussed

Tables 2 and 3 should be cited after Table 1. Please revise.

Response: Thanks very much for this suggestion. We have modified the order of the tables as your suggestion.

5. Table 4

Should it be TC and ET? Please revise.

946

947 **Table 4** Results of the official BraTS2017 validation set

Model structure	Dice score			Specificity			Sensitivity		
	WT	TC	ET	WT	TC	ET	WT	ET	WT

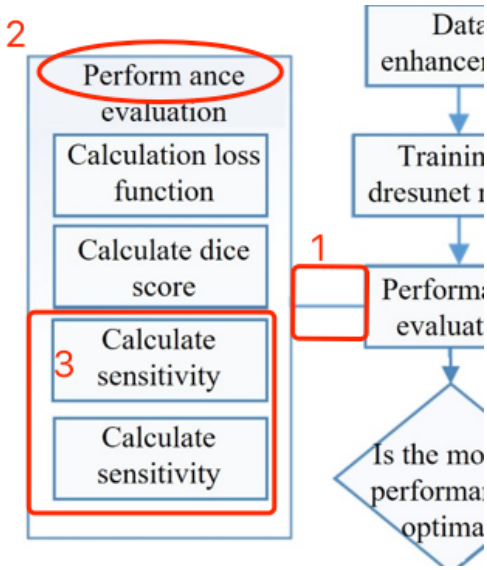
Response: Thanks very much for this suggestion. We have modified it as your suggestion.

6. Figure 7

6.1 Please add an arrow

6.2 Please revise this typo

6.3 The two boxes are the same. Please check and revise.



Response: Thanks very much for this suggestion. We have changed it as your suggestion.