



OPEN ACCESS

EDITED BY

Jose R. Pineda,
University of the Basque Country, Spain

REVIEWED BY

Haixian Zhang,
Sichuan University, China
Zekun Jiang,
Sichuan University, China
Kavita Kundal,
Indian Institute of Technology Hyderabad,
India

*CORRESPONDENCE

Wenbin Li
✉ liwenbin@ccmu.edu.cn

RECEIVED 30 August 2024

ACCEPTED 29 October 2024

PUBLISHED 15 November 2024

CITATION

Guo X, Zhang B, Peng Y, Chen F
and Li W (2024) Segmentation of
glioblastomas via 3D FusionNet.
Front. Oncol. 14:1488616.
doi: 10.3389/fonc.2024.1488616

COPYRIGHT

© 2024 Guo, Zhang, Peng, Chen and Li. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Segmentation of glioblastomas via 3D FusionNet

Xiangyu Guo, Botao Zhang, Yue Peng, Feng Chen
and Wenbin Li *

Department of Neuro-Oncology, Cancer Center, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

Introduction: This study presented an end-to-end 3D deep learning model for the automatic segmentation of brain tumors.

Methods: The MRI data used in this study were obtained from a cohort of 630 GBM patients from the University of Pennsylvania Health System (UPENN-GBM). Data augmentation techniques such as flip and rotations were employed to further increase the sample size of the training set. The segmentation performance of models was evaluated by recall, precision, dice score, Lesion False Positive Rate (LFPR), Average Volume Difference (AVD) and Average Symmetric Surface Distance (ASSD).

Results: When applying FLAIR, T1, ceT1, and T2 MRI modalities, FusionNet-A and FusionNet-C the best-performing model overall, with FusionNet-A particularly excelling in the enhancing tumor areas, while FusionNet-C demonstrates strong performance in the necrotic core and peritumoral edema regions. FusionNet-A excels in the enhancing tumor areas across all metrics (0.75 for recall, 0.83 for precision and 0.74 for dice scores) and also performs well in the peritumoral edema regions (0.77 for recall, 0.77 for precision and 0.75 for dice scores). Combinations including FLAIR and ceT1 tend to have better segmentation performance, especially for necrotic core regions. Using only FLAIR achieves a recall of 0.73 for peritumoral edema regions. Visualization results also indicate that our model generally achieves segmentation results similar to the ground truth.

Discussion: FusionNet combines the benefits of U-Net and SegNet, outperforming the tumor segmentation performance of both. Although our model effectively segments brain tumors with competitive accuracy, we plan to extend the framework to achieve even better segmentation performance.

KEYWORDS

brain tumor segmentation, MRI, U-net, SegNet, 3D deep learning model

Introduction

Brain is an organ that functions as the central hub of the nervous system, composed of various cell types and a complex microenvironment. According to the Global Burden of Disease Study 2019, brain and central nervous system tumors are among the most common types of cancers, with an estimated 350,639 cases and 252,814 cancer-related deaths in 2019 (1). Due to intricate biology and microenvironment of the brain, each type of tumor has its own characteristics, posing a challenge for clinicians to accurately describe the classification, location, and size of tumors (2, 3). Medical experts usually detect brain tumors through neurological exams, imaging practices, and biopsies. Among imaging tests, Magnetic Resonance Imaging (MRI) plays a crucial role in identifying the lesion location, extent of tissue involvement, and the resultant mass effect on the brain, ventricular system, and vasculature, due to its superior performance in visualizing organs and soft tissues (4, 5). Compared to regular X-rays and CT scans, MRI provides clearer images of non-bony parts such as the brain, spinal cord, nerves, and muscles. Since MRI does not use X-rays or other radiation, it is the preferred imaging test when frequent imaging is required for diagnosis or therapy, especially for brain imaging.

Gliomas, one of the most common types of brain cancer, often intermingle with healthy brain tissue and develop within the substance of the brain. Glioblastoma (GBM) constitutes the majority of WHO grade 4 gliomas and is one of the most lethal and recurrence-prone malignant solid tumors, accounting for 57% of all gliomas and 48% of primary central nervous system malignant tumors (6). Because of its unclear morphological structures, it is challenging for physicians to accurately identify the lesion location, extent of tissue involvement, and level of malignancy. Therefore, a series of computer-aided diagnostic (CAD) tools have been applied to assist in the more accurate diagnosis of cancer (7–9).

Tumor segmentation, the process of accurately separating tumors from their background, is one of the crucial steps in radiomics, diagnosis and treatments of brain tumors (10, 11). With the growing demands of clinical applications and scientific research, image segmentation has become increasingly important in the field of medical imaging processing. However, fully manual medical image segmentation is time-consuming and requires expertise. The advancement of CAD techniques has made it easier to complete tasks more efficiently. Image segmentation can be performed using numerous techniques, ranging from conventional methods to advanced deep learning approaches. In traditional techniques, the most common types are thresholding, region-based segmentation, edge-based segmentation, and clustering (12–18). Various machine learning methods, such as support vector machines, have also been effectively applied to medical imaging segmentation (19, 20). The emergence of deep learning has markedly boosted the precision and speed of image segmentation processes. These deep learning methods include generative adversarial networks (GANs), recurrent neural networks (RNNs), diffusion models and convolutional neural

networks (CNNs) (21–25). Of the different deep learning techniques, CNNs continue to be the most widely used method for brain tumor segmentation (26). A study utilized a cascaded 3D Fully Convolutional Network (FCN) to automatically detect and segment brain metastases with high accuracy. This method was also effective in distinguishing brain metastases from high-grade gliomas (27). U-Net and SegNet are two commonly used network architectures for semantic segmentation. Skip connections in U-Net aid in better recovering details and boundary information in segmentation tasks (28). Conversely, SegNet uses pooling indices for upsampling, which may lose some details but offers lower memory consumption and higher computational efficiency (29). U-Net 3+ incorporates multiple improvements over the U-Net structure, enabling it to better capture image details and contextual information, making it suitable for more complex and diverse segmentation tasks (30, 31).

In this study, to fully utilize the advantages of both networks, we developed a 3D FusionNet that combines the features of U-Net and SegNet. This end-to-end trainable model aims to achieve better segmentation performance in medical images.

Methods

Data collection

The MRI data used in this study were obtained from a cohort of 630 GBM patients from the University of Pennsylvania Health System (UPENN-GBM) (32). This dataset, which includes imaging data, clinical data, and radiomic data, is available through the Cancer Imaging Archive (TCIA) at the National Cancer Institute (33, 34). All MRI data were obtained prior to surgery using 3T MRI scanners and included T1, T2, ceT1, FLAIR, and related segmentation labels. Of the 630 patients, 611 with complete MRI data were used in this study. The MRI data consist of Glioblastoma patient scans, meticulously annotated by clinical experts to highlight sub-regions such as the necrotic core (NC), peritumoral edema (ED), and enhancing tumor (ET).

Data augmentation

The MRI data from UPENN-GBM were already well-preprocessed, so no additional preprocessing steps were performed. To better adapt to the network structure, we applied zero-padding to the images before data augmentation, expanding their sizes from the original 240*240*155 voxels to 256*256*256 voxels. We selected 80 samples from the 611 as the test set, 31 samples as the validation set, and the remaining samples as the training set. Subsequently, we employed a series of data augmentation techniques, including horizontal flip, vertical flip, and 90-, 180-, and 270-degree rotations, to further increase the sample size of the training set to 3,000 samples.

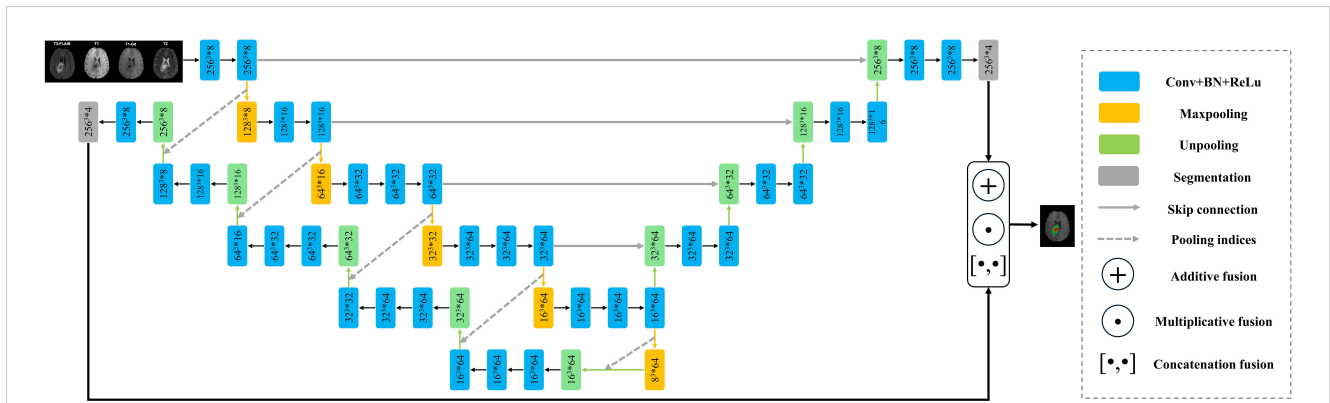


FIGURE 1
The network structure of FusionNet. Note: Here, for example, 2563*8 indicates that the feature map size is 256, and the channel count is 8. The symbol * represents multiplication.

Proposed network

The proposed end-to-end network architecture is shown in Figure 1. It comprises the following components: a shared encoder with 13 convolutional layers, mirroring the convolutional layers of the VGG16 network; two distinct decoders, one for U-Net and another for SegNet; and an optional joint output layer for the two decoders, offering additive (A), elementwise multiplicative (M), and concatenation (C) operations. Unlike traditional ensemble learning methods, our approach employs a shared encoder that enhances parameter efficiency, feature reusability, and model generalization. This design simplifies the architecture and promotes stable training by enabling both decoders to leverage the same robust feature extraction process. The additive (A) operation sums the outputs from the two decoders, harnessing the strengths of both architectures to enhance overall feature representation. The elementwise multiplicative (M) operation multiplies the outputs from the decoders elementwise, allowing for a more nuanced interaction between their outputs and effectively emphasizing areas of consensus. Finally, the concatenation (C) operation merges the outputs along the channel dimension, preserving the unique features from both decoders and enabling the model to leverage diverse representations.

Given the 3D structure of both the volumetric data and the network, which demands substantial computational memory, we had to reduce the number of channels in the convolutional layers. The specific sizes and channels of the feature maps at various stages are illustrated in Figure 1.

The loss function employed in this network is cross-entropy loss, which measures voxel-wise similarity between the prediction and ground truth. Each voxel in the image is treated as an independent classification task, and the cross-entropy loss is computed across all voxels in the image, as show in the Equation 1:

$$L = -\frac{1}{H \times W \times D} \sum_{h=1}^H \sum_{w=1}^W \sum_{d=1}^D \sum_{c=1}^C y_{h,w,d,c} \log(\hat{y}_{h,w,d,c}) \quad (1)$$

where $y_{h,w,d,c}$ is the true label for the pixel at position (h, w, d) for class c , $\hat{y}_{h,w,d,c}$ is the predicted probability for voxel (h, w, d) belonging to class c , given by the model. The loss is averaged over

all voxels to obtain a single scalar value representing the loss for the entire image.

Performance metrics

In this study, the segmentation performance of models was evaluated by recall, precision, dice score, Lesion False Positive Rate (LFPR), Average Volume Difference (AVD) and Average Symmetric Surface Distance (ASSD) (35–37). Recall measures the ability of the model to correctly identify positive instances as show in the Equation 2.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Precision measures the accuracy of the positive predictions as show in the Equation 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Dice Score is the harmonic mean of precision and recall, providing a single metric to evaluate the balance between the two as show in the Equation 4.

$$Dice = \frac{2 \times TP}{TP+FN+TP+FP} \quad (4)$$

Lesion False Positive Rate (LFPR) measures the proportion of non-lesion areas mistakenly identified as lesions by the model as show in the Equation 5. The LFPR values are small because the background occupies a large proportion. Due to the excessively low LFPR values, we performed normalization.

$$LFPR = \frac{FP}{FP+TN} \quad (5)$$

Average Volume Difference (AVD) assesses the difference in volume between the predicted lesions and the ground truth lesions as show in the Equation 6.

$$AVD = \frac{1}{N} \sum_{i=1}^N |V_{pred,i} - V_{true,i}| / V_{true,i} \quad (6)$$

Average Symmetric Surface Distance (ASSD) quantifies the average distance between the surfaces of the predicted and ground truth lesions

as show in the Equation 7.

$$ASSD = \frac{1}{2} \left(\frac{1}{N_{pred}} \sum_{p \in P} d(p, G) + \frac{1}{N_{true}} \sum_{g \in G} d(g, P) \right) \quad (7)$$

True Positive (TP) means the number of instances correctly predicted as positive; false positive (FP) means the number of instances incorrectly predicted as positive; true negative (TN) means the number of instances correctly predicted as negative; false negative (FN) means the number of instances incorrectly predicted as negative; N represents the number of lesion regions; $V_{pred,i}$ denotes the volume of the i -th predicted region; $V_{pred,i}$ indicates the volume of the i -th ground truth region; N_{pred} represents the number of points on the predicted surface; N_{true} represents the number of points on the ground truth surface; $d(p, G)$ refers to the distance from each point on the predicted surface to the nearest point on the ground truth surface; while $d(g, P)$ signifies the distance from each point on the ground truth surface to the nearest point on the predicted surface.

Results

Implementation details

In this study, we utilized four standard MRI modalities along with their corresponding expert-annotated segmentation maps to train the network. Both training and testing were conducted on an 80 GB NVIDIA A100 Tensor Core GPU using the Torch backend. The model's initial parameters were obtained through PyTorch's default initialization, and the network parameters were updated using the stochastic gradient descent optimization algorithm. We divided the 611 samples from UPENN-GBM into a training set (500 samples, 82%), which was augmented to reach a total of 3,000 samples, a validation set (31 samples, 5%), and a test set (80 samples, 13%). Specifically, the learning rate decays exponentially, with an initial value set to 0.01 and a decay factor of 0.8. The training was conducted over 20 epochs with a batch size of 5.

Segmentation performance with different methods

First, we compared the segmentation performance of the proposed network with U-Net, SegNet, and U-Net3+ using FLAIR, T1, ceT1, and T2 MRI modalities. Due to limited computational memory, U-Net3+ with reduced channels performed poorly and was not included in the results table. Overall, the models were more likely to accurately segment peritumoral edema regions and enhancing tumor areas compared to necrotic core regions. In terms of FusionNet, recall, precision, and dice scores for peritumoral edema regions and enhancing tumor areas generally achieve around 0.75 to 0.80. However, the metrics for necrotic core regions were only around 0.50. As shown in Table 1A, our proposed method, which combines the benefits of U-Net and SegNet, achieved comparable tumor segmentation performance to both U-Net and SegNet in terms of recall,

precision, and dice scores. U-Net performs best in recall (0.66) for the necrotic core regions but fails in the enhancing tumor areas. SegNet does not perform best in any region or metric. FusionNet-A excels in the enhancing tumor areas across all metrics (0.75 for recall, 0.83 for precision and 0.74 for dice scores) and also performs well in the peritumoral edema regions (0.77 for recall, 0.77 for precision and 0.75 for dice scores). FusionNet-M has good performance in the peritumoral edema regions (0.77 for recall, 0.70 for precision and 0.70 for dice scores) but does not lead in any specific metric. FusionNet-C has balanced performance, achieving the best precision and dice scores in the necrotic core regions (0.60 for precision and 0.54 for dice scores) and the best precision in the peritumoral edema regions (0.81). In summary, when applying FLAIR, T1, ceT1, and T2 MRI modalities, FusionNet-A and FusionNet-C emerged as the top-performing models overall. FusionNet-A particularly excelled in the segmentation of enhancing tumor areas, while FusionNet-C demonstrated strong performance in segmenting the necrotic core and peritumoral edema regions. U-Net has the best recall for the necrotic core region but lacks in other areas. SegNet only performs reasonably well in the segmentation of peritumoral edema regions but is significantly worse than the other models overall.

On the other hand, we compared different models based on LFPR, AVD, and ASSD (Table 1B). In terms of LFPR, FusionNet models significantly outperformed U-Net and SegNet across all three sub-regions. Among the FusionNet models, FusionNet-C excelled in segmenting the necrotic core, peritumoral edema regions, and enhancing tumor areas, achieving values of 0.01, 0.02, and 0.01, respectively. Additionally, FusionNet-C demonstrated the best AVD index, with a value of 0.34. Moreover, it achieved the best ASSD scores across all three sub-regions, with 1.31 for the necrotic core, 0.94 for peritumoral edema, and 0.85 for enhancing tumor regions, highlighting its superior overall segmentation performance.

Segmentation performance with different combinations of MRI modalities

Next, we assessed the effectiveness of segmentation results from various combinations of MRI modalities based on FusionNet-A. As shown in Table 2, reasonable segmentation results can be achieved using various combinations of MRI modalities. Notably, we found that FLAIR and ceT1 are crucial components in segmentation tasks. Combinations including FLAIR and ceT1 tend to have better segmentation performance, especially for necrotic core regions. Additionally, using either FLAIR or ceT1 alone can still yield somewhat satisfactory results, although slightly less effective compared to combinations of multiple modalities. From Table 2, we can see that using only FLAIR achieves a recall of 0.73 for peritumoral edema regions, and using only ceT1 achieves a recall of 0.50 for necrotic core regions (the best recall for necrotic core regions among all combinations of MRI modalities is 0.51). On the other hand, some combinations of two or three MRI modalities also achieve better performance in specific metrics than using all four MRI modalities. Using FLAIR, T1 and ceT1 results in the best recall

TABLE 1A Summary of commonality performance metrics for different methods.

Methods	Recall			Precision			Dice		
	NC	ED	ET	NC	ED	ET	NC	ED	ET
U-Net	0.66 ± 0.30	0.75 ± 0.17	0	0.28 ± 0.18	0.71 ± 0.21	0.12 ± 0.24	0.33 ± 0.18	0.72 ± 0.19	0
SegNet	0	0.36 ± 0.16	0	0	0.51 ± 0.24	0	0	0.40 ± 0.18	0
FusionNet-A	0.43 ± 0.25	0.77 ± 0.18	0.75 ± 0.26	0.57 ± 0.24	0.77 ± 0.17	0.83 ± 0.19	0.45 ± 0.23	0.75 ± 0.17	0.74 ± 0.24
FusionNet-M	0.05 ± 0.04	0.77 ± 0.22	0.71 ± 0.29	0.40 ± 0.27	0.70 ± 0.23	0.78 ± 0.13	0.08 ± 0.06	0.70 ± 0.23	0.70 ± 0.24
FusionNet-C	0.54 ± 0.22	0.71 ± 0.21	0.70 ± 0.25	0.60 ± 0.24	0.81 ± 0.17	0.75 ± 0.16	0.54 ± 0.21	0.74 ± 0.19	0.69 ± 0.20

The evaluated metrics are presented as mean ± standard deviation. NC, Necrotic Core; ED, Peritumoral Edema; ET, Enhancing Tumor. Bold values indicate the best value for each specific metric within each category.

TABLE 1B Summary of differences in performance metrics for different methods.

Methods	LFPR			AVD	ASSD		
	NC	ED	ET		NC	ED	ET
U-Net	0.04 ± 0.03	0.03 ± 0.03	0	1.35 ± 0.94	1.81 ± 1.43	1.24 ± 1.81	13.07 ± 4.53
SegNet	0	0.04 ± 0.02	0	0.80 ± 0.08	Nan	2.36 ± 2.18	Nan
FusionNet-A	0.01 ± 0.02	0.02 ± 0.02	0.02 ± 0.12	0.46 ± 1.01	2.02 ± 2.91	1.22 ± 2.19	1.68 ± 5.82
FusionNet-M	0	0.05 ± 0.12	0.01 ± 0.01	0.55 ± 0.38	2.95 ± 3.26	2.56 ± 6.09	1.15 ± 2.52
FusionNet-C	0.01 ± 0.01	0.02 ± 0.02	0.01 ± 0.01	0.34 ± 0.31	1.31 ± 1.55	0.94 ± 1.72	0.85 ± 1.21

The evaluated metrics are presented as mean ± standard deviation. LFPR, Lesion False Positive Rate; AVD, Average Volume Difference; ASSD, Average Symmetric Surface Distance; NC, Necrotic Core; ED, Peritumoral Edema; ET, Enhancing Tumor. Bold values indicate the best value for each specific metric within each category.

of 0.81 for enhancing tumor regions, the best precision of 0.81 for peritumoral edema regions and the best dice scores of 0.77 for peritumoral edema regions among all combinations of MRI modalities. The combination of FLAIR, ceT1, and T2 achieves the best recall of 0.51 for necrotic core regions and the best precision of 0.89 for enhancing tumor regions. The combination of T1, ceT1, and T2 achieves the best precision of 0.63 for necrotic core regions. Using FLAIR and ceT1 achieves the best recall of 0.84 for peritumoral edema regions among all combinations of MRI modalities. Another interesting finding was that combinations with T1 might reduce segmentation performance in necrotic core and enhancing tumor regions. Using T1 alone or in combinations without ceT1, metrics for necrotic core and enhancing tumor regions were significantly lower than others, sometimes even falling below 0.10.

Visualization of brain tumor segmentation

Moreover, we visualized the brain tumor segmentation results from various models, excluding SegNet due to its comparatively poor performance relative to the other models. We randomly selected a subject's T2 MRI modality with various colors representing different tumor regions. Figures 2–4 show T2 MRI modality slices from the axial, coronal, and sagittal planes, respectively. In these figures, red indicates necrotic core regions, yellow represents enhancing tumor regions, and green denotes peritumoral edema regions. From left to right, the images

display the original T2 image, the ground truth, and the segmentation results from U-Net, FusionNet-A, FusionNet-M and FusionNet-C. U-Net shows poor segmentation performance in the enhancing tumor areas, mistakenly classifying some of these regions as part of the necrotic core. This misclassification contributes to U-Net achieving the best recall for the necrotic core regions. Conversely, FusionNet-M tends to mistakenly identify necrotic core regions as peritumoral edema or enhancing tumor areas, leading to less precise segmentation in those regions. FusionNet-A and FusionNet-C generally achieve segmentation results that closely resemble the ground truth, although there are slight discrepancies in the segmentation of necrotic core regions compared to the manual annotations.

Discussion

Brain tumors are among the most lethal diseases in humans. MRI images are a crucial diagnostic tool for identifying tumors by providing detailed pictures of soft tissues. Advanced methods for identifying and quantifying tumor lesions are vital in cancer treatment. Image analysis, one of these methods, offers critical information, such as lesion location, volume, count, extent of tissue involvement, and the resultant mass effect on the brain. A variety of state-of-art machine learning and deep learning models have been developed to automate the process of segmentation and classification of tumors. Many deep learning methods primarily focus on handling two-dimensional slices, which require less GPU

TABLE 2 Summary of performance metrics for different combinations of MRI modalities based on FusionNet-A.

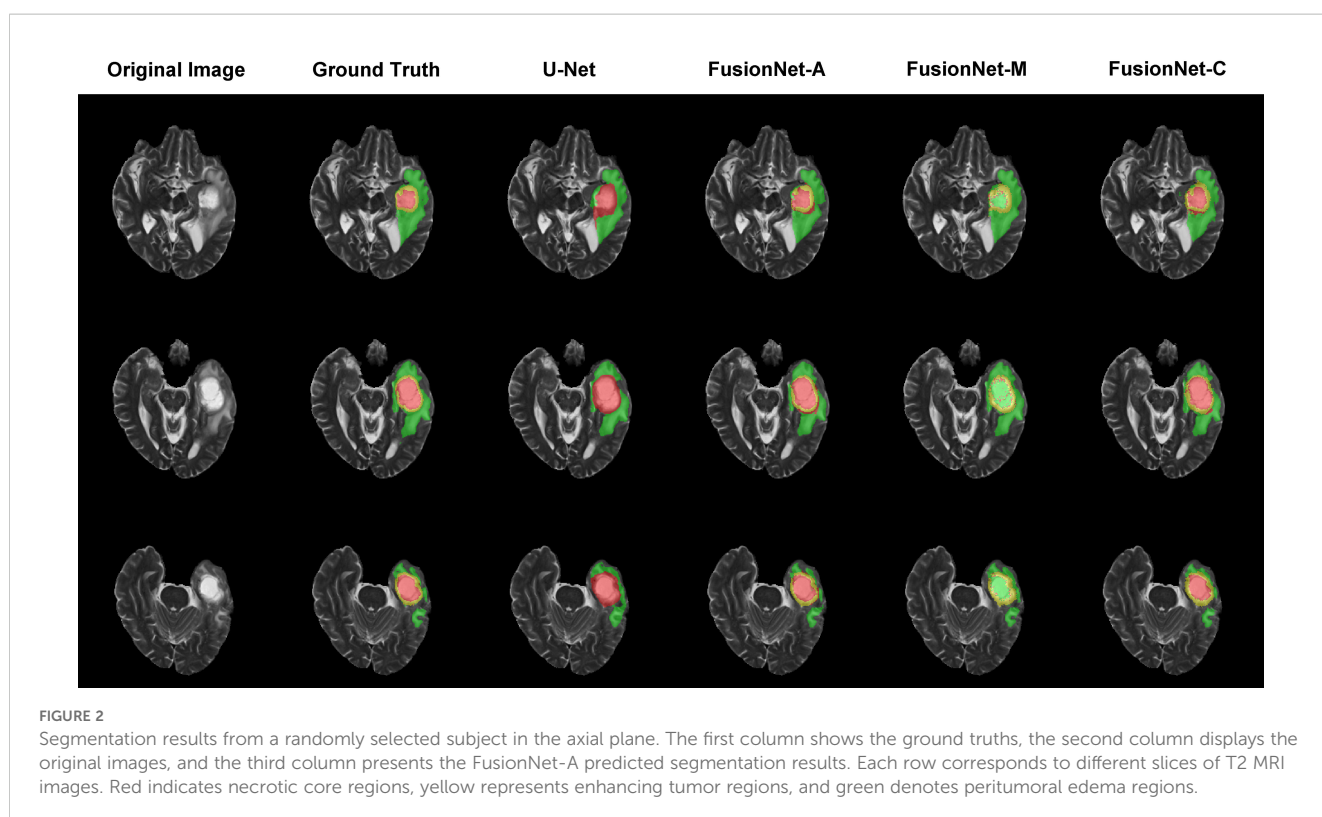
Modalities	Recall			Precision			Dice		
	NC	ED	ET	NC	ED	ET	NC	ED	ET
FLAIR/T1/T1c/T2	0.43 ± 0.25	0.77 ± 0.18	0.75 ± 0.26	0.57 ± 0.24	0.77 ± 0.17	0.83 ± 0.19	0.45 ± 0.23	0.75 ± 0.17	0.74 ± 0.24
FLAIR/T1/T1c	0.42 ± 0.25	0.71 ± 0.20	0.81 ± 0.23	0.58 ± 0.26	0.81 ± 0.18	0.80 ± 0.18	0.44 ± 0.24	0.73 ± 0.18	0.77 ± 0.23
FLAIR/T1/T2	0.12 ± 0.17	0.83 ± 0.15	0.07 ± 0.10	0.56 ± 0.33	0.53 ± 0.19	0.39 ± 0.24	0.17 ± 0.20	0.63 ± 0.19	0.10 ± 0.11
FLAIR/T1c/T2	0.51 ± 0.27	0.70 ± 0.17	0.49 ± 0.29	0.48 ± 0.25	0.72 ± 0.21	0.89 ± 0.15	0.44 ± 0.25	0.70 ± 0.18	0.57 ± 0.29
T1/T1c/T2	0.35 ± 0.25	0.68 ± 0.22	0.74 ± 0.25	0.63 ± 0.26	0.69 ± 0.19	0.80 ± 0.21	0.40 ± 0.25	0.66 ± 0.19	0.73 ± 0.24
FLAIR/T1	0.04 ± 0.06	0.74 ± 0.18	0.14 ± 0.16	0.38 ± 0.27	0.56 ± 0.22	0.44 ± 0.25	0.07 ± 0.09	0.62 ± 0.21	0.17 ± 0.18
FLAIR/T1c	0.34 ± 0.25	0.84 ± 0.18	0.58 ± 0.30	0.53 ± 0.25	0.64 ± 0.18	0.84 ± 0.21	0.35 ± 0.23	0.70 ± 0.17	0.61 ± 0.27
FLAIR/T2	0.39 ± 0.23	0.65 ± 0.17	0.33 ± 0.19	0.38 ± 0.22	0.64 ± 0.22	0.45 ± 0.19	0.34 ± 0.21	0.63 ± 0.18	0.32 ± 0.16
T1/T1c	0.35 ± 0.26	0.33 ± 0.21	0.69 ± 0.29	0.50 ± 0.26	0.54 ± 0.27	0.79 ± 0.26	0.36 ± 0.23	0.37 ± 0.21	0.70 ± 0.27
T1/T2	0.09 ± 0.16	0.66 ± 0.18	0.18 ± 0.17	0.30 ± 0.32	0.50 ± 0.21	0.32 ± 0.22	0.11 ± 0.18	0.55 ± 0.20	0.19 ± 0.15
T1c/T2	0.45 ± 0.26	0.55 ± 0.19	0.76 ± 0.21	0.61 ± 0.27	0.75 ± 0.21	0.73 ± 0.21	0.47 ± 0.24	0.61 ± 0.20	0.70 ± 0.18
FLAIR	0.23 ± 0.16	0.73 ± 0.17	0.28 ± 0.18	0.31 ± 0.17	0.63 ± 0.22	0.50 ± 0.21	0.22 ± 0.14	0.66 ± 0.19	0.31 ± 0.18
T1	0	0.15 ± 0.14	0	0.24 ± 0.37	0.34 ± 0.31	0.14 ± 0.25	0.01 ± 0.01	0.19 ± 0.18	0.01 ± 0.01
T1c	0.50 ± 0.30	0.33 ± 0.18	0.57 ± 0.30	0.46 ± 0.26	0.58 ± 0.27	0.83 ± 0.25	0.40 ± 0.23	0.38 ± 0.21	0.62 ± 0.27
T2	0.30 ± 0.28	0.54 ± 0.23	0.08 ± 0.09	0.37 ± 0.31	0.52 ± 0.24	0.32 ± 0.23	0.26 ± 0.24	0.51 ± 0.23	0.11 ± 0.10

The evaluated metrics are presented as mean ± standard deviation. NC, Necrotic Core; ED, Peritumoral Edema; ET, Enhancing Tumor. Bold values indicate the best value for each specific metric within each category.

memory and computational resources, and benefit from abundant available pre-trained weights (38–41). Compared to 2D models, 3D models use entire 3D MRI volumes as inputs, fully utilize 3D spatial information, better maintain the integrity of 3D structures, and

provide more coherent segmentation results, but they require more GPU memory, computational resources, and longer training times.

In this study, we proposed a 3D end-to-end FusionNet for brain tumor segmentation using multiple MRI modalities, capable of



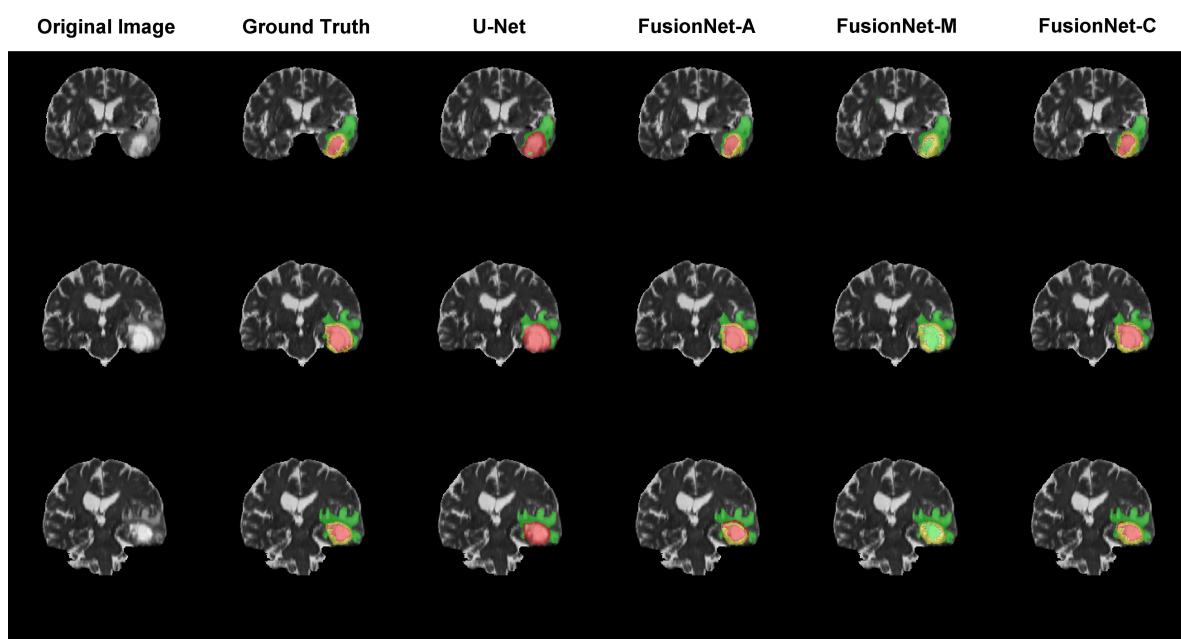


FIGURE 3 Segmentation results from a randomly selected subject in the coronal plane. The first column shows the ground truths, the second column displays the original images, and the third column presents the FusionNet-A predicted segmentation results. Each row corresponds to different slices of T2 MRI images. Red indicates necrotic core regions, yellow represents enhancing tumor regions, and green denotes peritumoral edema regions.

adaptively handling an arbitrary number of modalities. We extensively explored the segmentation performance of our model compared to other segmentation methods using MRI data from UPENN-GBM. We utilize recall, precision, Dice score, LFPR, AVD,

and ASSD to evaluate the segmentation performance, capturing both commonalities and differences between the predicted segmentation and the ground truth from two dimensions: overlap-based and distance-based metrics. FusionNet combines

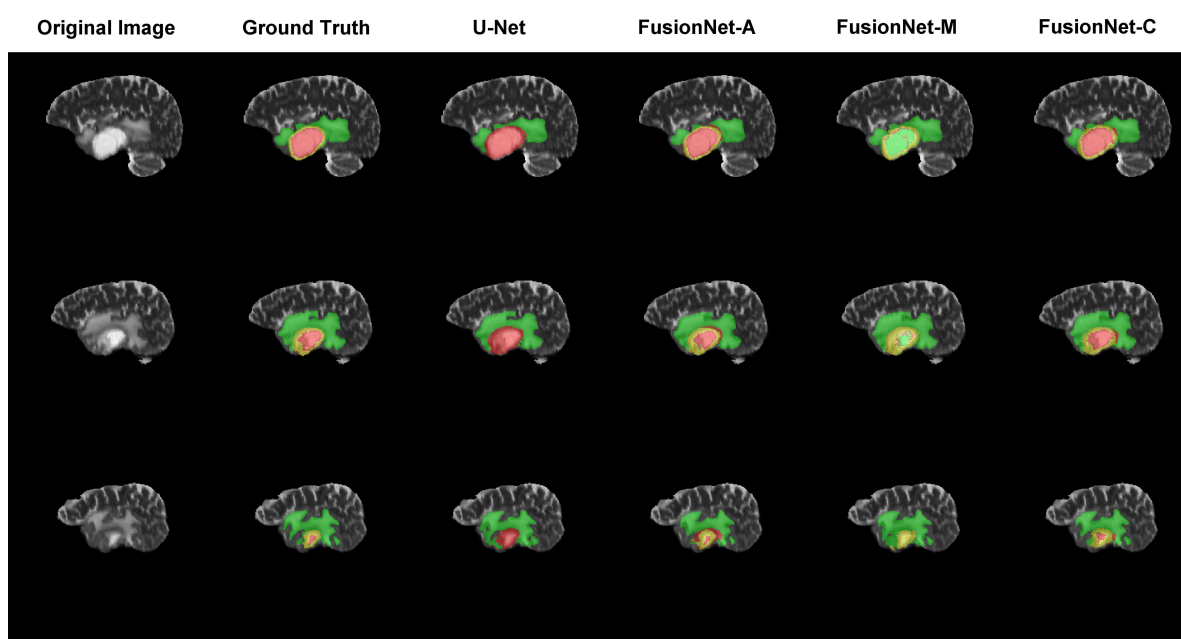


FIGURE 4 Segmentation results from a randomly selected subject in the sagittal plane. The first column shows the ground truths, the second column displays the original images, and the third column presents the FusionNet-A predicted segmentation results. Each row corresponds to different slices of T2 MRI images. Red indicates necrotic core regions, yellow represents enhancing tumor regions, and green denotes peritumoral edema regions.

the benefits of U-Net and SegNet, achieving comparable tumor segmentation performance to both in terms of recall, precision, and dice scores with three different joint strategies. Overall, FusionNet-A and FusionNet-C are the best-performing models in overlap-based metrics, with FusionNet-A particularly excelling in the enhancing tumor areas, while FusionNet-C demonstrates strong performance in the necrotic core and peritumoral edema regions. Additionally, FusionNet-A and FusionNet-C also outperform in distance-based metrics, with FusionNet-C showing a slight edge over FusionNet-A. We observed that some models with poor performance in recall or Dice score had LFP values approaching zero. This indicates that these models were unable to effectively segment the lesion areas, resulting in almost no false positive outcomes. Although FusionNet achieves comparable tumor segmentation results, its performance is limited due to the constraints on channels and layers. The input image size of $256 \times 256 \times 256$ voxels generates numerous large feature maps during the learning process. Consequently, with limited memory resources available for abundant weights, we had to reduce the number of channels and layers. In future work, we will focus on balancing the feature map size with the number of channels and layers.

Compared to the performance of models on the Brain Tumor Segmentation (BraTS) dataset, our results using the UPENN-GBM dataset show lower index values (42, 43). While UPENN-GBM primarily focuses on glioblastoma (GBM), the BraTS dataset includes a wider variety of brain tumors, such as low-grade and high-grade gliomas. Glioblastoma presents additional challenges due to its indistinct boundaries and the unclear separation between its sub-regions (44). This blending with surrounding brain tissue makes segmentation tasks particularly difficult, which contributes to the lower performance indices observed in our experiments.

Additionally, we investigated segmentation performance with different combinations of MRI modalities. FLAIR and ceT1 are vital modalities for segmentation tasks; using either FLAIR or ceT1 alone or in combination tends to result in better segmentation performance, especially for necrotic core regions. However, combinations involving T1 may reduce segmentation performance in necrotic core and enhancing tumor regions. FLAIR, T1, ceT1, and T2 MRI modalities each have unique characteristics for glioma segmentation tasks. The FLAIR sequence is particularly sensitive to peritumoral edema region, presenting as high signal intensity, and assists in segmenting the necrotic core area. The ceT1 sequence, after contrast agent injection, clearly shows the enhancing tumor region, significantly improving the segmentation of the necrotic core. The T1 sequence, without enhancement, has lower contrast for the necrotic core and enhancing tumor, resulting in poorer segmentation performance. The T2 sequence is sensitive to tissues with high water content, prominently highlighting the edema region and aiding in the segmentation of both the necrotic core and enhancing tumor. We will explore assigning different weights to the four sequences and organically combining them to achieve more accurate segmentation and evaluation of the different subregions in the future work.

To provide a more intuitive understanding of our model's segmentation effect, we visualized the segmentation results of

models alongside the corresponding original image and ground truth segmentation from the axial, coronal, and sagittal planes, respectively. FusionNet-A and FusionNet-C generally achieve segmentation results that closely resemble the ground truth compared to other models, although it performs slightly worse in the necrotic core regions. There are several reasons for this. First, the borders between the necrotic core, enhancing tumor, and peritumoral edema tissues are usually diffused and not clearly defined. Second, the model learns from training samples of original MRI images and their corresponding ground truth annotations, which were made by different people. Due to the unclear borders, each annotator may have a slightly different interpretation of the boundaries between the different tumor subregions. Lastly, global segmentation methods, such as semantic segmentation models, are prone to being influenced by background noise, especially when it comes to segmenting small targets, where they often underperform. This is because, in the process of handling the entire image, the features of small targets can easily be overshadowed or diminished by background information, making it difficult for the model to accurately capture the boundaries and shapes of these regions.

Our study has some limitations. First, as a 3D model, FusionNet requires more GPU memory and computational resources. Due to the higher computational cost and large feature maps, FusionNet must have limited channels and layers compared to its 2D counterpart, which might constrain the performance of the model. Second, the current model has slightly worse segmentation performance on necrotic core regions compared to the other two regions. To address the challenge of limited computing resources and poor performance for small targets, we will explore two approaches in future works. First, we will investigate the impact of different MRI image sizes on segmentation model performance, aiming to strike a reasonable balance between model depth and computational efficiency. Second, we will employ region-based segmentation, which involves initially locating and outlining the entire brain tumor, followed by segmenting the tumor sub-regions within this smaller, specific area.

The necrotic core (NC), enhancing tumor (ET), and peritumoral edema (ED) regions are of significant clinical importance (45, 46). The necrotic core is often associated with high-grade malignancies and poor prognosis. The enhancing tumor region indicates a disrupted blood-brain barrier and neovascularization, making it a key target for surgery and radiotherapy. The peritumoral edema reflects the tumor's pressure on surrounding brain tissue, leading to clinical symptoms and impacting treatment planning. Assessing these regions is crucial for diagnosis, treatment decisions, and prognosis in brain tumor management. Although manual segmentation can yield highly accurate results, it is time-consuming and labor-intensive, especially in large-scale studies. Automated segmentation methods can effectively address these limitations, allowing for manual correction to ensure accuracy while significantly improving segmentation efficiency.

In this paper, we presented an end-to-end deep learning model for the automatic segmentation of brain tumors. Our model effectively segments brain tumors with competitive accuracy.

However, brain tumor segmentation remains a challenging task, and we will further extend the framework of this model to achieve better segmentation performance.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

XG: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. BZ: Writing – review & editing. YP: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – review & editing. FC: Writing – review & editing. WL: Conceptualization, Funding acquisition, Supervision, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. We are grateful for the financial support from Talent Introduction

Foundation of Tiantan Hospital (No. RCYJ-2020–2025-LWB to WL); Clinical Major Specialty Projects of Beijing (2-1-2-038 to WL). The study funders did not participate in the design of the study, data collection, analysis, interpretation, or report writing.

Acknowledgments

We sincerely appreciate The Cancer Imaging Archive (TCIA) and the University of Pennsylvania glioblastoma (UPenn-GBM) cohort for providing the GBM data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. (2020) 396(10258):1223–49. doi: 10.1016/S0140-6736(20)30752-2
- Boire A, Brastianos PK, Garzia L, Valiente M. Brain metastasis. *Nat Rev Cancer*. (2020) 20:4–11. doi: 10.1038/s41568-019-0220-y
- Charles NA, Holland EC, Gilbertson R, Glass R, Kettenmann H. The brain tumor microenvironment. *Glia*. (2012) 60:502–14. doi: 10.1002/glia.21264
- Villanueva-Meyer JE, Mabray MC, Cha S. Current clinical brain tumor imaging. *Neurosurgery*. (2017) 81:397–415. doi: 10.1093/neuros/nyx103
- Alexander BM, Cloughesy TF. Adult glioblastoma. *J Clin Oncol*. (2017) 35:2402–9. doi: 10.1200/JCO.2017.73.0119
- Yang K, Wu Z, Zhang H, Zhang N, Wu W, Wang Z, et al. Glioma targeted therapy: insight into future of molecular approaches. *Mol Cancer*. (2022) 21(1):39. doi: 10.1186/s12943-022-01513-z
- Tang R, Liang H, Guo Y, Li Z, Liu Z, Lin X, et al. Pan-mediastinal neoplasm diagnosis via nationwide federated learning: a multicentre cohort study. *Lancet Digit Health*. (2023) 5(23):e560–e70. doi: 10.1016/S2589-7500(23)00106-1
- Zhou Z, Sanders JW, Johnson JM, Gule-Monroe MK, Chen MM, Briere TM, et al. Computer-aided detection of brain metastases in T1-weighted MRI for stereotactic radiosurgery using deep learning single-shot detectors. *Radiology*. (2020) 295(2):407–15. doi: 10.1148/radiol.2020191479
- Raghavendra U, Gudigar A, Paul A, Goutham TS, Inamdar MA, Hegde A, et al. Brain tumor detection and screening using artificial intelligence techniques: Current trends and future perspectives. *Comput Biol Med*. (2023) 163:107063. doi: 10.1016/j.compbmed.2023.107063
- Maino C, Vernuccio F, Cannella R, Franco PN, Giannini V, Dezio M, et al. Radiomics and liver: Where we are and where we are headed? *Eur J Radiol*. (2024) 171:111297. doi: 10.1016/j.ejrad.2024.111297
- Zhuang Y, Liu H, Song E, Hung CC. A 3D cross-modality feature interaction network with volumetric feature alignment for brain tumor and tissue segmentation. *IEEE J BioMed Health Inform*. (2023) 27:75–86. doi: 10.1109/JBHI.2022.3214999
- Chu Z, Zhang Q, Zhou H, Shi Y, Zheng F, Gregori G, et al. Quantifying choriocapillaris flow deficits using global and localized thresholding methods: a correlation study. *Quant Imaging Med Surg*. (2018) 8(11):1102–12. doi: 10.21037/qims.2018.12.09
- Chen X, Pan L. A survey of graph cuts/graph search based medical image segmentation. *IEEE Rev BioMed Eng*. (2018) 11:112–24. doi: 10.1109/RBME.2018.2798701
- Xu Q, Varadarajan S, Chakrabarti C, Karam LJ. A distributed Canny edge detector: algorithm and FPGA implementation. *IEEE Trans Image Process*. (2014) 23:2944–60. doi: 10.1109/TIP.2014.2311656
- Wang L, Tang D, Guo Y, Do MN. Common visual pattern discovery via nonlinear mean shift clustering. *IEEE Trans Image Process*. (2015) 24:5442–54. doi: 10.1109/TIP.2015.2481701
- Braiki M, Benzinou A, Nasreddine K, Hymery N. Automatic human dendritic cells segmentation using K-means clustering and Chan-Vese active contour model. *Comput Methods Programs BioMed*. (2020) 195:105520. doi: 10.1016/j.cmpb.2020.105520
- Freitas N, Silva D, Mavioso C, Cardoso MJ, Cardoso JS. Deep edge detection methods for the automatic calculation of the breast contour. *Bioengineering (Basel)*. (2023) 10(4):401. doi: 10.3390/bioengineering10040401
- Nonato LG, do Carmo FP, Silva CT. GLoG: laplacian of gaussian for spatial pattern detection in spatio-temporal data. *IEEE Trans Vis Comput Graph*. (2021) 27:3481–92. doi: 10.1109/TVCG.2020.2978847
- Wu T, Bae MH, Zhang M, Pan R, Badaea A. A prior feature SVM-MRF based method for mouse brain segmentation. *Neuroimage*. (2012) 59:2298–306. doi: 10.1016/j.neuroimage.2011.09.053
- Abdullah BA, Younis AA, John NM. Multi-sectional views textural based SVM for MS lesion segmentation in multi-channels MRIs. *Open BioMed Eng J*. (2012) 6:56–72. doi: 10.2174/1874120701206010056
- Tan J, Jing L, Huo Y, Li L, Akin O, Tian Y. LGAN: Lung segmentation in CT scans using generative adversarial network. *Comput Med Imaging Graph*. (2021) 87:101817. doi: 10.1016/j.compmedimag.2020.101817

22. Wang J, Tang Y, Xiao Y, Zhou JT, Fang Z, Yang F. GREnet: gradually REcurrent network with curriculum learning for 2-D medical image segmentation. *IEEE Trans Neural Netw Learn Syst.* (2023) 35(7):10018–32. doi: 10.1109/TNNLS.2023.3238381
23. Fan X, Shan S, Li X, Mi J, Yang J, Zhang Y, et al. Attention-modulated multi-branch convolutional neural networks for neonatal brain tissue segmentation. *Comput Biol Med.* (2022) 146:105522. doi: 10.1016/j.compbiomed.2022.105522
24. Li W, Huang W, Zheng Y. CorrDiff: corrective diffusion model for accurate MRI brain tumor segmentation. *IEEE J BioMed Health Inform.* (2024) 28:1587–98. doi: 10.1109/JBHI.2024.3353272
25. Rondinella A, Guarnera F, Giudice O, Ortis A, Russo G, Crispino E, et al. Enhancing multiple sclerosis lesion segmentation in multimodal MRI scans with diffusion models, in: 2023 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, (2023). 3733–40. doi: 10.1109/BIBM58861.2023.10385334
26. Ranjbarzadeh R, Caputo A, Tirkolaee EB, Jafarzadeh Ghouschi S, Bendechache M. Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. *Comput Biol Med.* (2023) 152:106405. doi: 10.1016/j.compbiomed.2022.106405
27. Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, et al. Deep learning-based detection and segmentation-assisted management of brain metastases. *Neuro Oncol.* (2020) 22(4):505–14. doi: 10.1093/neuonc/noz234
28. Zunair H, Ben Hamza A. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation. *Comput Biol Med.* (2021) 136:104699. doi: 10.1016/j.compbiomed.2021.104699
29. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell.* (2017) 39:2481–95. doi: 10.1109/TPAMI.34
30. Alam T, Shia WC, Hsu FR, Hassan T. Improving breast cancer detection and diagnosis through semantic segmentation using the unet3+ Deep learning framework. *Biomedicines.* (2023) 11(6):1536. doi: 10.3390/biomedicines11061536
31. Yang Y, Chen F, Liang H, Bai Y, Wang Z, Zhao L, et al. CNN-based automatic segmentations and radiomics feature reliability on contrast-enhanced ultrasound images for renal tumors. *Front Oncol.* (2023) 13:1166988. doi: 10.3389/fonc.2023.1166988
32. Bakas S, Sako C, Akbari H, Bilello M, Sotiras A, Shukla G, et al. The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: advanced MRI, clinical, genomics, & radiomics. *Sci Data.* (2022) 9(1):453. doi: 10.1038/s41597-022-01560-7
33. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* (2013) 26(6):1045–57. doi: 10.1007/s10278-013-9622-7
34. Bakas S, Sako C, Akbari H, Bilello M, Sotiras A, Shukla G, et al. Multi-parametric magnetic resonance imaging (mpMRI) scans for de novo Glioblastoma (GBM) patients from the University of Pennsylvania Health System (UPENN-GBM) (Version 2) [Data set]. *The Cancer Imaging Archive* (2022). doi: 10.7937/TCIA.709X-DN49
35. Choi YS, Bae S, Chang JH, Kang SG, Kim SH, Kim J, et al. Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. *Neuro Oncol.* (2021) 23(2):304–13. doi: 10.1093/neuonc/noaa177
36. Xu W, Jia X, Mei Z, Gu X, Lu Y, Fu CC, et al. Generalizability and diagnostic performance of AI models for thyroid US. *Radiology.* (2023) 307(5):e221157. doi: 10.1148/radiol.221157
37. Subramaniam P, Kossen T, Ritter K, Hennemuth A, Hildebrand K, Hilbert A, et al. Generating 3D TOF-MRA volumes and segmentation labels using generative adversarial networks. *Med Image Anal.* (2022) 78:102396. doi: 10.1016/j.media.2022.102396
38. Zhang J, Jiang Z, Dong J, Hou Y, Liu B. Attention gate resU-net for automatic MRI brain tumor segmentation. *IEEE Access.* (2020) 8:58533–45. doi: 10.1109/Access.6287639
39. Sun H, Yang S, Chen L, Liao P, Liu X, Liu Y, et al. Brain tumor image segmentation based on improved FPN. *BMC Med Imaging.* (2023) 23(1):172. doi: 10.1186/s12880-023-01131-1
40. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging.* (2016) 35:1240–51. doi: 10.1109/TMI.2016.2538465
41. Luo Z, Jia Z, Yuan Z, Peng J. HDC-net: hierarchical decoupled convolution network for brain tumor segmentation. *IEEE J BioMed Health Inform.* (2021) 25:737–45. doi: 10.1109/JBHI.6221020
42. Berkley A, Saueressig C, Shukla U, Chowdhury I, Munoz-Gauna A, Shehu O, et al. Clinical capability of modern brain tumor segmentation models. *Med Phys.* (2023) 50(8):4943–59. doi: 10.1002/mp.16321
43. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging.* (2015) 34:1993–2024. doi: 10.1109/TMI.2014.2377694
44. Kong D, Peng W, Zong R, Cui G, Yu X. Morphological and biochemical properties of human astrocytes, microglia, glioma, and glioblastoma cells using fourier transform infrared spectroscopy. *Med Sci Monit.* (2020) 26:e925754. doi: 10.12659/MSM.925754
45. Beig N, Singh S, Bera K, Prasanna P, Singh G, Chen J, et al. Sexually dimorphic radiogenomic models identify distinct imaging and biological pathways that are prognostic of overall survival in glioblastoma. *Neuro Oncol.* (2021) 23(2):251–63. doi: 10.1093/neuonc/noaa231
46. Beig N, Bera K, Prasanna P, et al. Radiogenomic-based survival risk stratification of tumor habitat on gd-T1w MRI is associated with biological processes in glioblastoma. *Clin Cancer Res.* (2020) 26:1866–76. doi: 10.1158/1078-0432.CCR-19-2556