TextBraTS: Text-Guided Volumetric Brain Tumor Segmentation with Innovative Dataset Development and Fusion Module Exploration

Xiaoyu Shi[1], Rahul Kumar Jain[1], Yinhao Li[1], Ruibo Hou[1], Jingliang Cheng[2], Jie Bai[2], Guohua Zhao[2], Lanfen Lin[3], Rui Xu[4][5], and Yen-wei Chen[1]⋆

[1] Ritsumeikan University, Osaka, Japan
`chen@is.ritsumei.ac.jp`
[2] Department of Magnetic Resonance Imaging, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China
[3] Zhejiang University, Hangzhou, China
[4] DUT School of Software Technology & DUT-RU International School of Information Science and Engineering, Dalian University of Technology, Dalian, China
[5] DUT-RU Co-Research Center of Advanced ICT for Active Life, Dalian, China

Abstract. Deep learning has achieved great success in medical image segmentation and computer-aided diagnosis, with many advanced methods reaching state-of-the-art performance in brain tumor segmentation from MRI. While studies in other medical domains show that integrating textual reports with images can enhance segmentation, there is no comprehensive brain tumor dataset pairing radiological images with textual annotations. This gap has limited the development of multimodal approaches. To address this, we introduce TextBraTS, the first publicly available, volume-level multimodal dataset with paired MRI volumes and textual annotations, derived from the BraTS2020 benchmark. Based on this dataset, we propose a baseline framework and a sequential cross-attention method for text-guided volumetric segmentation. Extensive experiments with various text-image fusion strategies and templated text demonstrate clear improvements in segmentation accuracy and provide insights into effective multimodal integration. The dataset and model are available at https://github.com/Jupitern52/TextBraTS.

Keywords: Brain tumor · Multimodal fusion · Segmentation.

1 Introduction

Brain tumors are highly dangerous malignancies with low survival rates [14]. In clinical practice, brain tumor diagnosis and treatment are typically guided by the analysis of tumor location, shape, and morphology using brain MRI scans [1]. Tumor segmentation plays a critical role in computer-aided diagnosis systems, as the accurate delineation of tumor regions is essential for downstream

---

⋆ is the corresponding author

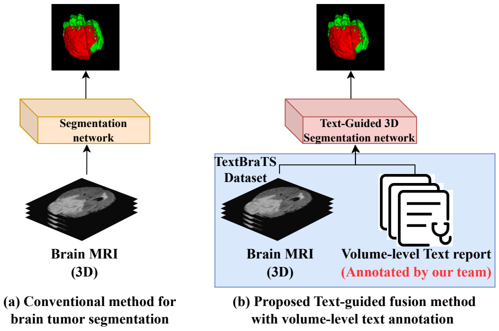


Fig. 1. Comparison between conventional brain tumor segmentation methods and our proposed method: (a) Segmentation network for tumor segmentation with image modality input only; (b) Our proposed text-guided multimodal segmentation network with a novel volume-level text report dataset.

tasks such as tumor grading and genetic marker prediction [23][26][28]. With advancements in deep learning, several slice-based methods [22][24][25][29] and volume-based methods [6][7][10][21][27] have been developed to segment tumor regions based on MRI scans. However, a significant limitation of these approaches is their reliance on imaging data alone, without incorporating complementary multimodal information. In practical diagnosis, doctors often integrate radiological reports with imaging data to achieve a more comprehensive diagnostic evaluation [20]. The absence of multimodal integration in existing methods underscores a critical gap in their applicability and effectiveness. Furthermore, the majority of publicly available brain tumor segmentation datasets are restricted to a single imaging modality. This limitation, which impedes the development of advanced methodologies that could leverage the synergistic potential of integrating textual radiological reports with imaging data. In other medical image segmentation domains, several methods have integrated textual reports with imaging data to achieve segmentation performance superior to that of single-modality approaches. Several works [3][8][12] have introduced different effective text-image fusion strategies on datasets with slice-level textual annotations, such as QaTa-COV19 (X-ray) [4] and MosMedData+ (CT) [16], significantly improving segmentation accuracy. However, in the field of brain tumor segmentation, no similar text-image dataset and text-image multimodal segmentation methods.

To address the aforementioned challenges, we develop TextBraTS, the first publicly available volume-level text-image brain tumor segmentation dataset. This dataset is based on the BraTS20 segmentation challenge training set [15], comprising 369 multimodal brain MRI scans, with expert textual annotations. Furthermore, we propose a volume-level text-guided image segmentation method, that utilizes sequential cross-attention (SeqCA) mechanism to fuse high-level textual information with image features, thereby improving segmentation accuracy for multiple brain tumor regions. We also compare various template-based inputs for text reports and propose a template processing method for brain tumor reports. As shown in Fig. 1, compared to existing methods, this work provides a novel dataset and a foundational approach for volume-level text-guided medical image segmentation. Our key contributions are as follows:

1. We create the first publicly available volume-level text-image brain MRI tumor segmentation dataset, which includes multimodal MRI volumes and their corresponding detailed textual descriptions.
2. We propose a text-guided image segmentation model that incorporates a sequential cross-attention (SeqCA) fusion mechanism, significantly improving segmentation accuracy.
3. Through extensive experiments, we explore the impact of different text template inputs and present comprehensive experimental results on our dataset.

2 TextBraTS Dataset Creation

In this paper, we create the TextBraTS dataset, a novel volume-level text-image pair dataset. This dataset is an extension of the BraTS2020 segmentation challenge dataset [15], comprising four imaging modalities and one textual modality.

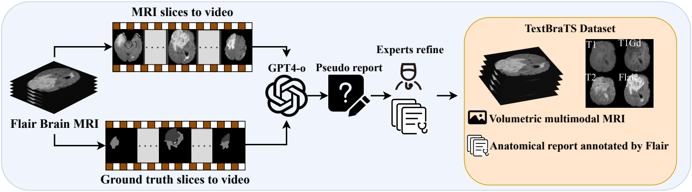
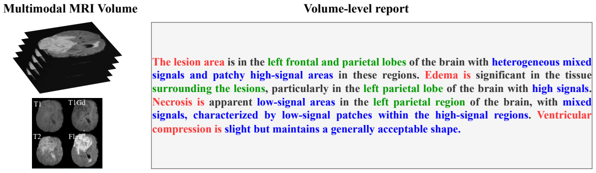


Fig. 2. The TextBraTS dataset creation pipeline involves an initial annotation phase using the GPT-4o model to generate preliminary pseudo-reports. These reports are then refined by expert radiologists to ensure accuracy, resulting in the final TextBraTS dataset.

Dataset Creation The methodology for constructing the TextBraTS dataset is illustrated in Fig. 2 . We created corresponding volume-level textual annotations for each case in the training set of the multimodal brain MRI dataset from the BraTS20 segmentation challenge [15]. Benefiting from the rapid advancements in multimodal large models, current models have demonstrated significant capabilities in analyzing and processing medical images. To reduce time costs, we leveraged the GPT-4o model [17] as a pre-annotation assistant. Each case in the BraTS20 segmentation challenge dataset includes four modalities: T1, T1Gd, T2, and Flair. According to the official dataset description [15], tumor segmentation annotations are primarily based on the Flair modality. Therefore, our



The lesion area is in the left frontal and parietal lobes of the brain with heterogeneous mixed signals and patchy high-signal areas in these regions. Edema is significant in the tissue surrounding the lesions, particularly in the left parietal lobe of the brain with high signals. Necrosis is apparent low-signal areas in the left parietal region of the brain, with mixed signals, characterized by low-signal patches within the high-signal regions. Ventricular compression is slight but maintains a generally acceptable shape.

Fig. 3. A sample in the TextBraTS dataset contains MRI data from four modalities and one text modality. The red part of the report represents the template words, the green part represents the location descriptions, and the blue part represents the features.

textual descriptions are also derived from the Flair modality. For each case, we sliced the 3D Flair images into individual frames and converted them into video format. Similarly, the corresponding tumor ground truth labels were sliced and converted into videos. These were then input into the GPT-4o model, guided by a carefully designed template prompt to generate analyses of the images. The analysis is divided into four parts: Overall lesion location and signal characteristics; Edema regions and features; Necrotic regions and features; Descriptions of midline shift and ventricular compression caused by the tumor's impact. To maintain high data quality and consistency, we developed an automated quality control pipeline. Each report was automatically checked against standardized templates and predefined keywords, with any inconsistencies triggering report regeneration. Owing to space limitations, only selected prompt samples are shown; complete versions are provided in the released dataset. After generation, all reports underwent expert review, with radiologists cross-checking the textual content against the images and segmentation labels. Corrections were made to ensure accuracy, and final descriptions were confirmed by human annotators. For each case, two radiologists independently annotated the data, with disagreements resolved by a third specialist. The GPT-4o model was employed solely to generate preliminary template drafts, streamlining the annotation process.

Dataset Analysis Compared to existing medical text report datasets [4][16], the key distinction of our dataset is featured by comprehensive anatomical information and radiological description. To account for natural variations in human anatomy, radiologists describe locations using standardized anatomical nomenclature rather than fixed spatial coordinates, as per established clinical protocols. Therefore, the locations are described using the 3D spatial coordinates of the brain, such as the frontal lobe, temporal lobe, and other anatomical structures. The incorporated features include details such as signal intensities, the presence of discrete abnormal signal clusters, and other radiological characteristics. We adopted a simplified, template-based approach for the reports, guided by our experimental results. The details of this approach will be discussed in the Experiments section. An example of the TextBraTS textual annotations is shown in Fig. 3.

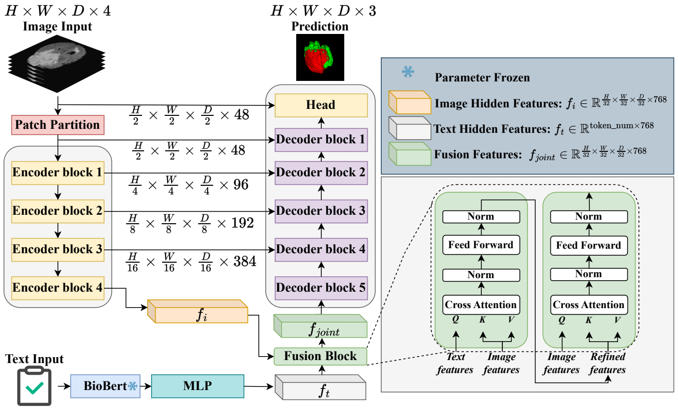


Fig. 4. An overview of our text-guided brain tumor segmentation approach is presented. Our approach is built on a Transformer-based segmentation model as the backbone. To effectively integrate text and image information, we introduce a sequential cross-attention (SeqCA) fusion block, enabling enhanced feature interaction between the two modalities.

3 Text-Guided Volumetric Brain Tumor Segmentation

Based on the TextBraTS dataset, we propose a text-guided integration segmentation model that effectively improves segmentation performance. Additionally, our approach provides foundational strategies to guide the fusion of 3D imaging and textual information for related tasks.

Overview of the model The structure of our framework is illustrated in Fig. 4. We utilize the pre-trained language model BioBERT [11] to extract text embeddings. BioBERT is pretrained on a vast amount of biomedical knowledge giving it strong feature extraction capabilities for medical reports. Since BioBERT is based on a Transformer architecture, we adopt SwinUNETR [6], also a Transformer-based model, as our image feature extractor, incorporating an encoder and decoder design. This architectural alignment facilitates better fusion of text and image modalities by ensuring compatibility in feature map sizes. First, the model extracts text features using the frozen BioBERT, then employs a multi-layer perceptron (MLP) [19] to map the text features into the image feature space for alignment. This process is described by the following formula:

$$f_t = MLP(BioBERT(t)). \tag{1}$$

Let t be the text input, and $f_t \in \mathbb{R}^{\text{token\_num} \times 768}$ is the text features. The token number is set as 128, based on the average length of text data. The output features of the image encoder can be described as $f_i \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times \frac{D}{32} \times 768}$, where the H, W, D are the input size of image encoder. Then, the f_t and f_i are sent to fusion block:

$$f_{joint} = Fusion(f_t, f_i). \tag{2}$$

Given that the initial layers possess significantly higher spatial resolution compared to the bottleneck layer (e.g., $64^{\mathbf{3}}$ vs. $4^{\mathbf{3}}$), integrating features at lower levels is not feasible. Consequently, we implement fusion at the bottleneck layer. Finally, the fused features $f_{joint} \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times \frac{D}{32} \times 768}$, which maintain the same size as f_i are passed to the decoder to complete the segmentation of the target regions.

Sequential Cross-attention fusion For the fusion strategy, we propose a **Sequential Cross-attention (SeqCA)** method. To further align the text and image modalities, we first use text information to refine the image features. The f_i is reshaped as $f_i \in \mathbb{R}^{\frac{H}{32}\frac{W}{32}\frac{D}{32} \times 768}$. The first cross-attention triplet of (Query, Key, Value) is computed as:

$$Q = f_t W_q, K = f_i W_k, V = f_i W_v. \tag{3}$$

The text-guided refined image features f_i' is computed as:

$$f_i' = M((Softmax(\frac{QK^T}{\sqrt{d}})V), \tag{4}$$

where the M is a module consisting of two normalization layers with a linear layer in between. Through the first cross-attention mechanism, the refined features $f_i' \in \mathbb{R}^{\text{token\_num} \times 768}$ are updated to capture shared information from text modality. These refined features are used to guide the original features, further improving the segmentation performance. The second cross-attention can be expressed as:

$$Q' = f_i W_q', K' = f_i' W_k', V' = f_i' W_v', \tag{5}$$

$$f_{joint} = M'((Softmax(\frac{Q'K'^T}{\sqrt{d}})V'), \tag{6}$$

where the M' has the same structure as M. The text-guided image features f_{joint} are sent to the first decoder layer for segmentation prediction.

4 Experiment

Dataset and Evaluation Metrics The experimental dataset used in this study is the TextBraTS dataset proposed in this paper. The dataset consists of 369 brain MRI cases with brain tumors, where each case includes four imaging modalities (T1, T1Gd, T2, and Flair) and one text modality (radiology report). The segmentation task focuses on three regions: whole tumor (WT), enhancing tumor (ET), and tumor core (TC). The dataset was randomly split into training (220 cases), validation (55 cases), and testing (94 cases) sets. During testing, the checkpoint with the best performance on the validation set was used.The evaluation metrics used in this study are the Dice Similarity Coefficient (Dice) [5] and Hausdorff Distance at the 95th Percentile (Hd95) [9].

Table 1. Comparison of Segmentation Performance (Dice and HD95 Metrics)

Method	**Dice (%) ↑**				**HD95 ↓**			
	ET	WT	TC	Avg.	ET	WT	TC	Avg.
3D-UNet [21]	80.4	87.3	81.6	83.1	6.11	10.51	8.93	8.17
nnU-Net [10]	82.2	87.5	82.6	84.1	**4.27**	11.90	8.52	8.23
SegResNet [7]	80.9	88.4	82.3	83.8	6.18	7.28	7.41	6.95
SwinUNETR [6]	81.0	89.5	80.8	83.8	5.95	8.23	7.03	7.07
Nestedformer [27]	82.6	89.5	80.2	84.1	5.08	10.51	8.93	8.17
Our work	**83.3**	**89.9**	**82.8**	**85.3**	4.58	**5.48**	**5.34**	**5.13**

Implementation Details Our text-guided Brain Tumor Segmentation model was implemented using PyTorch [18] and MONAI [2] and trained on two NVIDIA RTX A6000 GPUs. We used a batch size of 2 per GPU, with a learning rate of 0.0001, and the optimizer was AdamW [13]. For initialization, the weights of the SwinUNETR Transformer encoder were loaded from NVIDIA's providing model. A warm-up training strategy was applied for the first 50 epochs, with a total of 200 epochs of training. Since text data cannot undergo the same augmentation and cropping operations as images, we resampled all data to a fixed size of 128×128×128 for both training and testing, ensuring consistency across modalities.

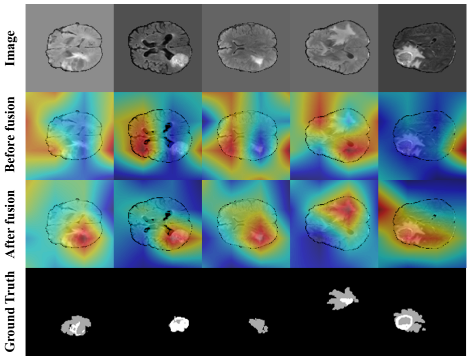


Fig. 5. Visualization of features before fusion block and after fusion block

Comparison with SOTA methods We compared several state-of-the-art (SOTA) methods based on CNN and Transformer architectures. The results, as shown in Table 1, demonstrate that our text-guided image segmentation model, developed using the newly proposed TextBraTS dataset, outperforms existing SOTA methods across all segmentation regions. To evaluate the perfor-

Table 2. Ablaion study on different input text reports

	Location	**Feature**	**Dice (%) ↑**				**HD95 ↓**			
			ET	**WT**	**TC**	**Avg.**	**ET**	**WT**	**TC**	**Avg.**
Raw text	✓	✓	83.0	89.6	82.1	84.6	5.96	7.11	6.32	6.46
Templated	✓		82.5	89.9	81.5	84.6	5.94	7.21	7.28	6.81
		✓	82.2	89.6	82.0	84.6	5.25	5.83	6.30	5.78
	✓	✓	**83.3**	**89.9**	**82.8**	**85.3**	**4.58**	**5.48**	**5.34**	**5.13**

Table 3. Ablation study of different fusion modules

Method	**T2I**	**I2T**	**Dice (%) ↑**				**HD95 ↓**			
			ET	**WT**	**TC**	**Avg.**	**ET**	**WT**	**TC**	**Avg.**
Dot Sum			81.6	87.0	78.8	81.6	5.96	11.90	7.41	8.57
Cross-attention	✓		82.5	89.9	82.1	84.8	5.43	7.18	6.18	6.26
Sequential Cross-attention	✓	✓	**83.3**	**89.9**	**82.8**	**85.3**	**4.58**	**5.48**	**5.34**	**5.13**

mance difference between our model and SOTA methods [27], we conducted 10 independent training and testing runs. A t-test revealed a statistically significant difference with a p-value of 0.0077 (< 0.05). Compared to the split in [27] (315/17/37), our partitioning (220/55/94) achieves better performance, highlighting the generalizability of our model. We also visualized the feature maps before and after the fusion module, with some slice examples shown in Fig. 5. In some cases, it can be observed that the image features guided by text exhibit a significantly improved perception of tumor regions. This indicates that the text effectively assists in image segmentation.

Ablation study Excessive and highly diverse text descriptions in small-scale medical image datasets may hinder model generalization. To address this, we explored different text formats and content representations. First, we standardized a template format to enhance consistency. Then, we evaluated four types of text inputs: raw text, location-only templated, features-only templated, and fully templated, to determine the most effective approach for guiding segmentation. The results, presented in Table 2, show that fully templated text inputs significantly improve segmentation performance for both the tumor interior and its edges. By analyzing the results, we observed that location-only inputs achieved better overall tumor segmentation, as indicated by higher Dice scores. In contrast, feature-only inputs enhanced edge segmentation accuracy, as reflected in lower HD95 values. We believe this is because location information is more effective for identifying the overall tumor position, whereas feature information provides a detailed description of abnormal signals, which better aids edge segmentation. These results suggest that structured and templated text inputs play a crucial role in optimizing segmentation performance by providing more consistent and interpretable guidance for the model. The ablation study for the fusion module presented in Table 3, compares three different methods: the dot sum of two modalities, one-step cross-attention for image features, and sequential cross-attention (SeqCA) (T2I: Text to Image and I2T: Image to Text). The results demonstrate that the SeqCA approach achieves the best performance, highlighting its effectiveness in integrating text and image features for improved segmentation accuracy. Through the processing of SeqCA, the two modalities are better aligned in the same space, leading to superior results compared to using cross-attention only once.

5 Conclusion

We present **TextBraTS**, a novel text-image brain tumor segmentation dataset and a text-guided segmentation method based on sequential cross attention, which outperforms state-of-the-art approaches. While text is essential in clinical diagnosis, existing methods often neglect it due to technical limitations. Our work addresses this gap and offers a foundation for future multimodal medical diagnostics, with plans to explore advanced fusion and segmentation techniques.

Acknowledgments. This work was supported in part by MEXT Grants-in-Aid for Scientific Research (20KK0234, 21H03470, 20K21821), the National Key R&D Project of China (2022YFC2504605), and JST SPRING, Japan (JPMJSP2101).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bakas, S., Akbari, H., Sotiras, A., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
2. Cardoso, M.J., Li, W., Brown, R., et al.: MONAI: An open-source framework for deep learning in healthcare (Nov 2022). `https://doi.org/10.48550/arXiv.2211.02701`
3. Chai, S., Jain, R.K., Mo, S., et al.: A Novel Adaptive Hypergraph Neural Network for Enhancing Medical Image Segmentation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, vol. 15009, pp. 23–33. Cham (2024). `https://doi.org/10.1007/978-3-031-72114-4_3`
4. Degerli, A., Kiranyaz, S., Chowdhury, M.E., et al.: Osegnet: Operational segmentation network for covid-19 detection using chest x-ray images. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 2306–2310 (2022)
5. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
6. Hatamizadeh, A., Nath, V., Tang, Y., et al.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, vol. 12962, pp. 272–284. Cham (2022). `https://doi.org/10.1007/978-3-031-08999-2_22`

7. Hsu, C., Chang, C., Chen, T.W., et al.: Brain Tumor Segmentation (BraTS) Challenge Short Paper: Improving Three-Dimensional Brain Tumor Segmentation Using SegResnet and Hybrid Boundary-Dice Loss. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, vol. 12963, pp. 334–344. Cham (2022). `https://doi.org/10.1007/978-3-031-09002-8_30`
8. Hu, J., Li, Y., Sun, H., et al.: LGA: A Language Guide Adapter for Advancing the SAM Model's Capabilities in Medical Image Segmentation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, vol. 15012, pp. 610–620. Cham (2024). `https://doi.org/10.1007/978-3-031-72390-2_57`
9. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance. IEEE Transactions on pattern analysis and machine intelligence **15**(9), 850–863 (1993)
10. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
11. Lee, J., Yoon, W., Kim, S., et al.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics **36**(4), 1234–1240 (2020)
12. Li, Z., Li, Y., Li, Q., et al.: Lvit: language meets vision transformer in medical image segmentation. IEEE transactions on medical imaging (2023)
13. Loshchilov, I.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Louis, D.N., Perry, A., Wesseling, P., et al.: The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro-oncology **23**(8), 1231–1251 (2021)
15. Menze, B.H., Jakab, A., Bauer, S., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
16. Morozov, S.P., Andreychenko, A.E., Pavlov, N.A., et al.: MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset (May 2020). `https://doi.org/10.48550/arXiv.2005.06465`
17. OpenAI: Hello GPT-4o. Tech. rep. (2024)
18. Paszke, A., Gross, S., Chintala, S., et al.: Automatic differentiation in pytorch (2017)
19. Popescu, M.C., Balas, V.E., Perescu-Popescu, L., et al.: Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems **8**(7), 579–588 (2009)
20. Reiner, B.I., Knight, N., Siegel, E.L.: Radiology reporting, past, present, and future: the radiologist's perspective. Journal of the American College of Radiology **4**(5), 313–319 (2007)
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, pp. 234–241. Cham (2015). `https://doi.org/10.1007/978-3-319-24574-4_28`
22. Shi, X., Jain, R.K., Li, Y., Chai, S., Cheng, J., Bai, J., Zhao, G., Lin, L., Chen, Y.W.: Multi-modal medical sam: An adaptation method of segment anything model (sam) for glioma segmentation using multi-modal mr images. ACM Transactions on Computing for Healthcare **6**(2), 1–21 (2025)

23. Shi, X., Li, Y., Chen, Y.W., et al.: An intra-and inter-modality fusion model with invariant-and specific-constraints using mr images for prediction of glioma isocitrate dehydrogenase mutation status. Journal of Image and Graphics **11**(4) (2023)
24. Shi, X., Li, Y., Cheng, J., Bai, J., Zhao, G., Chen, Y.W.: Knowledge distillation using segment anything to u-net model for lightweight high accuracy medical image segmentation. In: 2024 IEEE 13th Global Conference on Consumer Electronics (GCCE). pp. 1073–1076. IEEE (2024)
25. Shi, X., Li, Y., Cheng, J., Bai, J., Zhao, G., Chen, Y.W.: Medical sam: A glioma segmentation fine-tuning method for sam using brain mr images. In: 2024 IEEE International Conference on Consumer Electronics (ICCE). pp. 1–4. IEEE (2024)
26. Shi, X., Zhang, X., Iwamoto, Y., Cheng, J., Bai, J., Zhao, G., Chen, Y.W.: An intra- and inter-modality fusion model using mr images for prediction of glioma isocitrate dehydrogenase (idh) mutation. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 198–202. IEEE (2022)
27. Xing, Z., Yu, L., Wan, L., et al.: NestedFormer: Nested Modality-Aware Transformer for Brain Tumor Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, vol. 13435, pp. 140–150. Cham (2022)
28. Zhang, X., Shi, X., Iwamoto, Y., et al.: Idh mutation status prediction by a radiomics associated modality attention network. The Visual Computer **39**(6), 2367–2379 (2023)
29. Zhao, X., Wu, Y., Song, G., et al.: A deep learning model integrating fcnns and crfs for brain tumor segmentation. Medical image analysis **43**, 98–111 (2018)