Multi-level Feature Exploration and Fusion Network for Prediction of IDH Status in Gliomas from MRI

Jiawei Zhang, Jianyun Cao, Fan Tang, Tao Xie, Qianjin Feng, *Member, IEEE*, and Meiyan Huang

This supplementary material provides details related to the three datasets, performance of the proposed method with different backbones, configurations of eight compared methods, segmentation performance of the proposed network on BraTS dataset, and the survival analysis of the three independent testing sets.

I. DATASET RELATED DETAILS

The parameters of MRI sequences obtained from three datasets are listed in Tables SI and SII. The scanning parameters of four sequences are not fixed in three datasets due to the different scanning devices. Therefore, value range of each scanning parameter is listed in Tables SI and SII, where TR, TE and TI are repetition time, echo time and inversion time, respectively.

II. EXPERIMENTS

A. Performance of the Proposed Method with Different Backbones

1) Different backbones used in SFE module: Experiments were performed to investigate the effects of using different backbones in the SFE module on the performance of IDH prediction. Particularly, ResNet50, DenseNet169 [1], and CBAM-ResNet50 were applied to replace SE-ResNet50 in the SFE module. Then, the CBAM-ResNet50 was constructed by replacing the SE attention module with CBAM [2] attention module, which can be used to calculate the channel attention and spatial attention successively to improve network performance. Moreover, the AMF and DFF modules remained unchanged for a fair comparison. The highest AUC, ACC, and SEN values are achieved by using SE-ResNet50 as the

TABLE SI
MRI SEQUENCE PARAMETERS IN NANFANG AND ZHUJIANG HOSPITALS

	Nanfang				Zhujiang	
Equipment (num)	GE (90)	Siemens (31)	Philips (132)	UIH (1)	GE (2)	Philips (296)
TR(ms)	FLAIR(7502–9602) T1(1585–3441) T1CE(360–2898) T2(3640–5940)	FLAIR(8000) T1(550) T1CE(450–663) T2(5000)	FLAIR(9000–11000) T1(2000–2065) T1CE(4–2000) T2(2719–3108)	FLAIR(8000) T1(2274) T1CE(180) T2(4500)	FLAIR(8600) T1(1738–1739) T1CE(2435) T2(4329–4355)	FLAIR(9000–11000) T1(15–2000) T1CE(8–303) T2(2734–15000)
TE(ms)	FLAIR(110–165) T1(19–24) T1CE(9–28) T2(97–137)	FLAIR(92) T1(10) T1CE(17) T2(89)	FLAIR(110–160) T1(20) T1CE(1–20) T2(80–105)	FLAIR(107) T1(10) T1CE(3) T2(107)	FLAIR(147–148) T1(20) T1CE(24) T2(112–113)	FLAIR(120–149) T1(4–20) T1CE(2–4) T2(77–100)
TI(ms)	FLAIR(2100–2400) T1(720–860) T1CE(0–860) T2(0)	FLAIR(2371) T1(Unknown) T1CE(Unknown) T2(Unknown)	FLAIR(2500–2800) T1(800–850) T1CE(Unknown) T2(Unknown)	FLAIR(2425) T1(950) T1CE(Unknown) T2(Unknown)	FLAIR(2302–2308) T1(720) T1CE(720) T2(Unknown)	FLAIR(2500–2850) T1(800) T1CE(Unknown) T2(Unknown)
Flip angle (degree)	FLAIR(90) T1(90) T1CE(90) T2(90)	FLAIR(150) T1(90) T1CE(90) T2(150)	FLAIR(90) T1(90) T1CE(8–90) T2(90)	FLAIR(135) T1(135) T1CE(70) T2(150)	FLAIR(160) T1(160) T1CE(160) T2(160)	FLAIR(90) T1(15–90) T1CE(15–80) T2(90)
Pixel spacing (mm^2)	FLAIR(0.4297–0.5469) T1(0.4297–0.5469) T1CE(0.4297–0.5469) T2(0.4297–0.5469)	FLAIR(0.3594) T1(0.4492) T1CE(0.4492–0.4590) T2(0.2669–0.2995)	FLAIR(0.4271–0.4792) T1(0.2995–0.4492) T1CE(0.4102–0.4688) T2(0.2396–0.4492)	FLAIR(0.5047) T1(0.4357) T1CE(0.4357) T2(0.3832)	FLAIR(0.4688) T1(0.4688) T1CE(0.4688) T2(0.4688)	FLAIR(0.4217–0.8984) T1(0.4229–0.4492) T1CE(0.4141–0.5324) T2(0.2658–0.4492)
Slice thickness(mm)	FLAIR(5–6) T1(5–6.5) T1CE(5–6.5) T2(5–6.5)	FLAIR(5–5.5) T1(5–5.5) T1CE(5–5.5) T2(5–5.5)	FLAIR(3–7) T1(5–7.5) T1CE(5–7.5) T2(5–7)	FLAIR(5) T1(5) T1CE(5) T2(5)	FLAIR(6–6.5) T1(6–6.5) T1CE(6) T2(6–6.5)	FLAIR(3–6.5) T1(3–6.5) T1CE(1.1–6.5) T2(2–6.5)
Slice number(num)	FLAIR(19–26) T1(19–26) T1CE(19–26) T2(19–26)	FLAIR(20–24) T1(20–24) T1CE(20–24) T2(20–24)	FLAIR(18–45) T1(18–32) T1CE(14–28) T2(18–32)	FLAIR(23) T1(23) T1CE(23) T2(21)	FLAIR(18) T1(18) T1CE(18) T2(18)	FLAIR(18–36) T1(12–36) T1CE(15–150) T2(9–60)
Magnetic field strength(num)	1.5T(31) 3.0T(59)	1.5T(31)	3.0T(132)	3.0T(1)	1.5T(2)	3.0T(296)

TABLE SII
MRI SEQUENCE PARAMETERS IN TCIA DATASET.

	TCIA				
Equipment (num)	GE (105)	Siemens (38)	Philips (28)	Hitachi (1)	Unknown (35)
TR(ms)	FLAIR(9502–10004) T1(9–3099) T1CE(6–3256) T2(3000–6000)	FLAIR(7000–10000) T1(440–2200) T1CE(39–5500) T2(3500–6000)	FLAIR(6000–11000) T1(8–2000) T1CE(8–2450) T2(2555–6251)	FLAIR(6600) T1(479) T1CE(479) T2(4829)	FLAIR(8002–10004) T1(500–3379) T1CE(500–3285) T2(3000–5780)
TE(ms)	FLAIR(120–157) T1(2–17) T1CE(2–17) T2(59–116)	FLAIR(94–134) T1(2–17) T1CE(2–17) T2(90–120)	FLAIR(120–140) T1(3–20) T1CE(3–20) T2(80–110)	FLAIR(100) T1(12) T1CE(12) T2(96)	FLAIR(123–155) T1(6–15) T1CE(6–15) T2(103–110)
TI(ms)	FLAIR(2000–2500) T1(0–1238) T1CE(0–1238) T2(0)	FLAIR(2200–2700) T1(900–1100) T1CE(900–1100) T2(Unknown)	FLAIR(2000–2850) T1(800–900) T1CE(750–900) T2(Unknown)	FLAIR(2000) T1(0) T1CE(0) T2(0)	FLAIR(2200–2800) T1(0–1238) T1CE(0–1238) T2(0)
Flip angle (degree)	FLAIR(90) T1(8–90) T1CE(8–90) T2(90)	FLAIR(125–180) T1(8–90) T1CE(8–90) T2(120–180)	FLAIR(90) T1(8–90) T1CE(8–90) T2(90)	FLAIR(90) T1(90) T1CE(90) T2(90)	FLAIR(90) T1(62–90) T1CE(62–90) T2(90)
Pixel spacing (mm^2)	FLAIR(0.4297–1.1016) T1(0.4297–1.0156) T1CE(0.4297–1.0156) T2(0.4297–1.0156)	FLAIR(0.4101–1) T1(0.4492–1) T1CE(0.4492–1) T2(0.3593–1)	FLAIR(0.625–0.9375) T1(0.3906–0.9375) T1CE(0.4921–0.9375) T2(0.3906–0.9375)	FLAIR(0.3906) T1(0.3906) T1CE(0.3906) T2(0.3906)	FLAIR(0.4687–0.938) T1(0.4687–0.938) T1CE(0.4687–0.94) T2(0.4687–0.7187)
Slice thickness(mm)	FLAIR(2–5) T1(1.2–5) T1CE(1.2–5) T2(2–5.5)	FLAIR(3–5) T1(1–5) T1CE(1–5) T2(2–5)	FLAIR(2.5–5) T1(1–5) T1CE(1–5) T2(2.5–5)	FLAIR(5) T1(5) T1CE(5) T2(23)	FLAIR(2.5–5) T1(3–5) T1CE(2.5–5) T2(5)
Slice number(num)	FLAIR(22–88) T1(20–212) T1CE(20–212) T2(20–90)	FLAIR(19–64) T1(20–192) T1CE(20–192) T2(20–92)	FLAIR(24–79) T1(20–170) T1CE(20–170) T2(22–79)	FLAIR(24) T1(23) T1CE(23) T2(23)	FLAIR(25–78) T1(26–60) T1CE(26–144) T2(25–40)
Magnetic field strength(num)	1.5T(54) 3.0T(42) Unknown(9)	1.5T(29) 3.0T(9)	1.5T(8) 3.0T(20)	1.16T(1)	1.0T(1) 1.5T(15) 3.0T(13) Unknown(6)

backbone of the SFE module, as shown in Table SIII, which indicates that exploring the relationship among channels and assigning channel weights adaptively may benefit the feature extraction. Moreover, the performance of IDH prediction slightly decreased when the SE module was replaced by CBAM. Given that spatial attention is added in CBAM, more parameters are added in CBAM than in the SE module, which may increase the training difficulty of the network and decline the performance of the proposed method.

2) Different backbones used in AMF module: ResNet34 and ResNet50 were used to replace ResNet18 as the backbone in the AMF module to evaluate the influence of using different backbones in the AMF module on the performance of the proposed method. The results are shown in Table SIV. The best performance is achieved when ResNet18 is set as the backbone of the AMF module. Moreover, the performance of IDH prediction decreased when the complexity of the model increased, which indicates that increased parameter size may add difficulty in model training and lead to model overfitting, thereby resulting in worse performance.

Two independent ResNet18 were used to replace the Siamese architecture to evaluate the influence of using Siamese architecture in the AMF module on the proposed method. The features f_{T2} and f_{FLAIR} extracted from two independent ResNet18 were concatenated with f_{tumor} and f_{mis}. Then, the concatenated features were sent into the intra-slice feature

TABLE SIII
PREDICTION PERFORMANCE OF USING DIFFERENT BACKBONES IN SFE MODULE.

Backbone	AUC	ACC	SPE	SEN
ResNet50	0.8385	0.7795	**0.8881**	0.6396
DenseNet169	0.8381	0.7717	0.7343	0.8198
CBAM-ResNet50	0.8552	0.7835	0.7483	0.8288
SE-ResNet50	**0.8564**	**0.8032**	0.7762	**0.8378**

TABLE SIV
PREDICTION PERFORMANCE OF USING DIFFERENT BACKBONES IN AMF MODULE.

Backbone	AUC	ACC	SPE	SEN
ResNet50	0.8423	0.7874	0.8531	0.7027
ResNet34	0.8504	0.8031	**0.8951**	0.6847
ResNet18	**0.8564**	**0.8032**	0.7762	**0.8378**

fusion part to perform feature fusion. Results are shown in Table SV. Herein, the performance of IDH prediction decreased when the Siamese architecture is replaced by the independent network, which indicates that the Siamese network can project the T2 and FLAIR sequences into a common latent space and benefit from the extraction of T2-FLAIR mismatch features, thereby improving the performance of IDH prediction.

TABLE SV

USING DIFFERENT ARCHITECTURES IN THE FEATURE-LEVEL MISMATCH PART OF THE AMF MODULE, WHERE "INDEPENDENT" REPRESENTS USING TWO INDEPENDENT RESNET18 NETWORKS.

Architecture	AUC	ACC	SPE	SEN
Independent	0.8385	0.7717	**0.8392**	0.6847
Siamese	**0.8564**	**0.8032**	0.7762	**0.8378**

B. The Configurations of the Eight Compared Methods

Four state-of-the-art IDH prediction methods, two 3D segmentation methods, and two multi-task methods, including DCNN [3], RCNN [4], FAD [5], MTDL [6], Swin UNETR [7], nnUnet_3D [8], SGPNet [9], and PS-Net [10], were compared with the proposed method to evaluate the effectiveness of the proposed method on tumor segmentation and IDH prediction. The configurations of the eight compared methods are as follows:

- DCNN [3]: Tumor masks were used to generate cropped slice-by-slice images of the tumor on all MRI sequences, and each image was subsequently resized to be an 32×32 input. A residual network containing four residual blocks was used. For the testing stage, mean scores were calculated across all image slices of a patient to evaluate overall per-patient accuracy for final IDH prediction.
- RCNN [4]: Axial slice with the largest tumor area was chosen, and the tumor region was cropped and resized to 142×142 for training. The preprocessed 2D slices were fed into a 2D ResNet34 for IDH prediction. The first two layers of the original residual network architecture were discarded because the input size was smaller than that of the original residual net.
- FAD [5]: The axial slice with the maximum tumor area was selected as the "maximum tumor image." The 4 lower, 2 lower, 2 upper, and 4 upper slices from the maximum tumor image were also selected as training samples. Each selected slice concatenated with its corresponding tumor mask was resized to 128×128 and used as input of 2D ResNet34. For the testing stage, the final result was obtained by averaging the results of the five selected slices.
- MTDL [6]: The biggest size of all tumor bounding boxes was first searched, and all MRI images with the biggest box were cropped to obtain inputs with the same size. A multi-task CNN was used to predict tumor genotype and overall survival time. Based on the datasets used in our study, we calculated the tumor bounding box and cropped the inputs to $160 \times 160 \times 16$. Only the IDH prediction task used in [6] was retained for fair comparison.
- Swin UNETR [7]: It adopted a self-attention mechanism to use multi-scale contextual information effectively, which was implemented using MONAI toolkit. The configuration of the Swin UNETR was set by default as described in [7] and the input size was set as $96 \times 96 \times 96$.
- nnUnet_3D [8]: The MRI sequences were initially performed N4 algorithm and registration. Then, all MRI images were skull-stripped and sent into the nnUnet framework to fulfill the remaining pre-processing. Moreover, the network configuration details can be inferred by the nnUnet pipeline.
- SGPNet [9]: It modified the hyperparameters of 3D U-Net, added more skip connections, and replaced the ReLU activation function with the LeakyReLU activation function to improve the performance of IDH prediction and tumor segmentation. The MRI sequences were registered, cropped, and padded, which resulted in volumes with the size of $160 \times 160 \times 160$.
- PS-Net [10]: It added global max-pooling layers at each depth of the 3D U-Net to extract helpful imaging features for IDH prediction and tumor segmentation. The four MRI sequences were registered and interpolated to a resolution of $1 \times 1 \times 1 \ mm^3$. Subsequently, the four MRI sequences were cropped and padded to achieve volumes with the size of $160 \times 160 \times 160$, and these volumes were concatenated along channel dimensions as network inputs.

C. Segmentation Performance of the Proposed Network on BraTS Dataset

The performance of whole tumor segmentation of the proposed method was validated on BraTS 2021 [11] validation dataset to further illustrate the effectiveness of the proposed method. Considering that the IDH status was unavailable in BraTS 2021 dataset, the IDH prediction task was removed (AMF and DFF modules were removed) from the proposed method. With this process, the proposed method degenerated into a 2D SE-ResUNet. Moreover, the implementation details of the SE-ResUNet were based on the nnUnet framework as described in [8]. Table SVI lists the whole tumor segmentation results of the proposed method and five 3D segmentation methods provided by the MICCAI workshop, such as GNN-CNN [12], ResUnet-L [13], Swin UNETR [7], nnUnet_3D, and nnSegNet [14]. The Swin UNETR and nnSegNet were the top-ranking methods for the BraTS 2021 dataset among the compared segmentation methods. From Table SVI, the proposed method achieved good segmentation results on the online validation dataset, which demonstrates the ability of the whole tumor segmentation of the proposed method.

D. The Survival Analysis of the Three Independent Testing Sets

Given that gliomas with IDH-mutant show better prognosis than gliomas with IDH-wild, the IDH-related risk scores generated by the proposed method were used for survival analysis. Moreover, the IDH-related risk scores were regarded as a hazard ratio. Particularly, the Yorden index in the receiver operating characteristic (ROC) curve was initially used to identify the threshold, and thus, high- and low-risk cohorts can be divided. Then, the survival analysis on three independent testing sets was performed through Kaplan-Meier analysis along with a log-rank test [15]. Furthermore, the concordance index (C-index) and time-dependent ROC (tROC) curves were used to assess the prediction performance and the effectiveness of the IDH-related risk scores in survival analysis, respectively.

TABLE SVI

PERFORMANCE OF THE WHOLE TUMOR SEGMENTATION ON BRATS 2021 VALIDATION DATASET IN TERMS OF MEAN DICE SCORE AND HAUSDORFF DISTANCE (95%) VALUES.

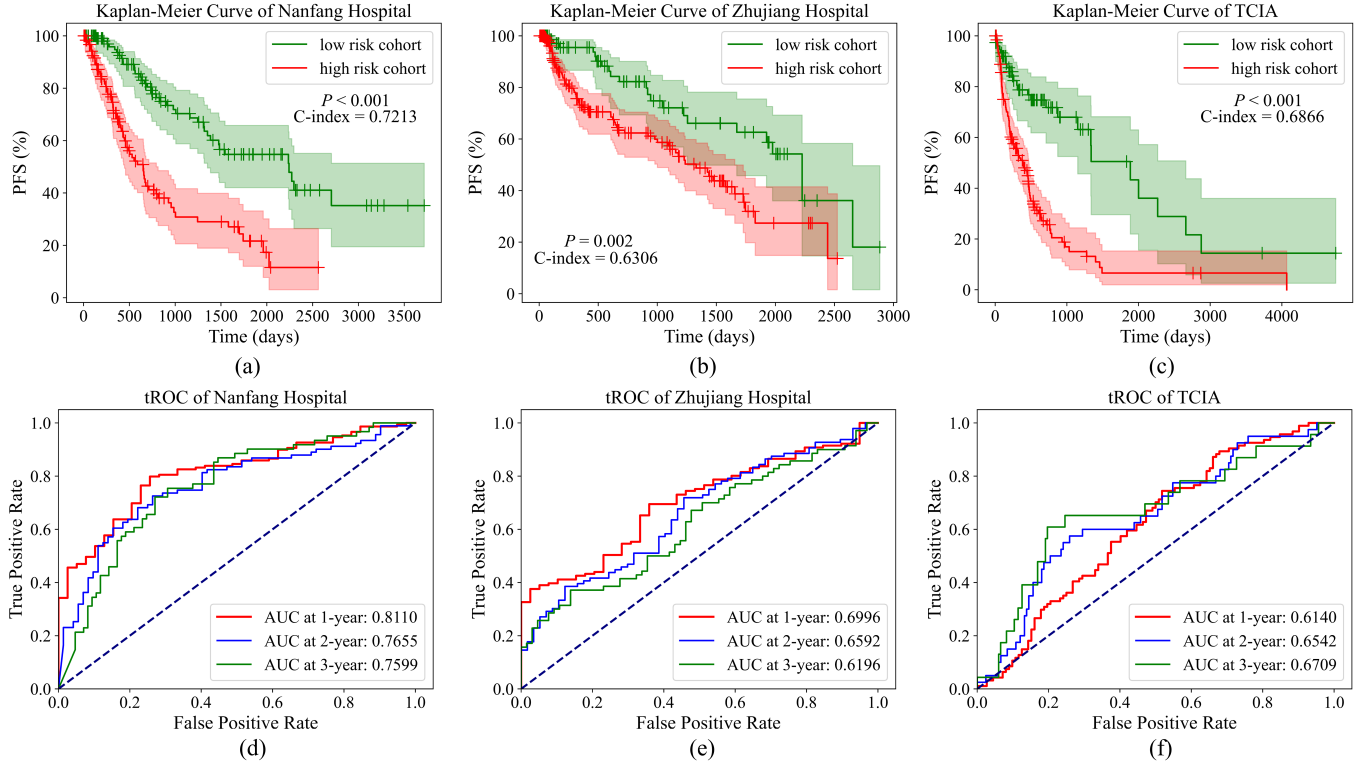| Method	GNN-CNN	ResUnet-L	Swin UNETR	nnUnet_3D	nnSegNet	SE-ResUNet
Dice	0.894	0.916	0.926	0.926	**0.928**	0.919
HD95	6.790	4.350	5.831	3.793	**3.483**	4.439

Fig. S1. Kaplan-Meier and tROC curves of three independent testing sets.

Fig. S1 shows the Kaplan-Meier curves and tROC curves of three independent testing sets. Kaplan-Meier curves show that the probability of PFS is significantly worse in the high-risk cohort than in the low-risk cohort in three independent testing sets [Figs. S1 (a), (b), and (c)], which illustrates that IDH-related risk scores are highly related to the PFS of patients. In addition, the C-index [Figs. S1 (a), (b), and (c)] shows the good discriminative ability of the proposed method on three different testing sets. Moreover, tROC curves show the AUC of 1-, 2-, and 3-year, which demonstrates that the IDH-related risk scores generated by the proposed method could be used for the prediction of patients' survival [Figs. S1 (d), (e), and (f)].

REFERENCE

[1] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.

[2] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European conference on computer vision*, 2018, pp. 3-19.

[3] P. Chang, J. Grinband, B. D. Weinberg, M. Bardis, M. Khy, G. Cadena, M. Y. Su, S. Cha, C. G. Filippi, and D. Bota, "Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas," *American Journal of Neuroradiology*, vol. 39, no. 7, pp. 1201-1207, 2018.

[4] K. Chang, H. X. Bai, H. Zhou, C. Su, W. L. Bi, E. Agbodza, V. K. Kavouridis, J. T. Senders, A. Boaro, and A. Beers, "Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging," *Clinical Cancer Research*, vol. 24, no. 5, pp. 1073-1081, 2018.

[5] Y. S. Choi, S. Bae, J. H. Chang, S.-G. Kang, S. H. Kim, J. Kim, T. H. Rim, S. H. Choi, R. Jain, and S.-K. Lee, "Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics," *Neuro-Oncology*, vol. 23, no. 2, pp. 304-313, 2021.

[6] Z. Tang, Y. Xu, L. Jin, A. Aibaidula, J. Lu, Z. Jiao, J. Wu, H. Zhang, and D. Shen, "Deep learning of imaging phenotype and genotype for predicting overall survival time of glioblastoma patients," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2100-2109, 2020.

[7] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop*, 2022, pp.272-284.

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203-211, 2021.

[9] Y. Wang, Y. Wang, C. Guo, S. Zhang, and L. Yang, "SGPNet: A three-dimensional multitask residual framework for segmentation and IDH genotype prediction of gliomas," *Computational Intelligence and*

Neuroscience, vol. 2021, pp. 1-9, 2021.

[10] S. R. van der Voort, F. Incekara, M. M. J. Wijnenga, G. Kapsas, R. Gahrmann, J. W. Schouten, R. Nandoe Tewarie, G. J. Lycklama, P. C. De Witt Hamer, and R. S. Eijgelaar, "Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning," *Neuro-oncology*, vol. 25, no. 2, pp. 279–289, 2023.

[11] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, and S. Pati, "The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv:* 2107.02314, 2021.

[12] C. Saueressig, A. Berkley, R. Munbodh, and R. Singh, "A joint graph and image convolution network for automatic brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop*, 2022, pp. 356-365.

[13] M. Demoustier, I. Khemir, Q. D. Nguyen, L. Martin-Gaffé, and N. Boutry, "Residual 3D U-Net with Localization for Brain Tumor Segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop*, 2022, pp. 389-399.

[14] N. Jabareen, and S. Lukassen, "Segmenting brain tumors in multi-modal MRI scans using a 3D SegNet architecture," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop*, 2022, pp. 377-388.

[15] Z. Zuo, W. Liu, Y. Zeng, X. Fan, L. Li, J. Chen, X. Zhou, Y. Jiang, X. Yang, and Y. Feng, "Multiparametric magnetic resonance imaging-derived deep learning network to determine ferroptosis-related gene signatures in gliomas," *Frontiers in Neuroscience*, vol. 16, pp. 1082867, 2022.