Multi-task model for glioma segmentation and isocitrate dehydrogenase status prediction using segmentation boundary

Xiaoyu Shi[1], Yinhao Li[1], Yen-Wei Chen[*,1], Jingliang Cheng[2], Jie Bai[2], and Guohua Zhao[2]

1. Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga, Japan,

2. The Affiliated Hospital of Zhengzhou University, Zhengzhou, China,

is0490sr@ed.ritsumei.ac.jp,
yinli@fc.ritsumei.ac.jp,
chen@is.ritsumei.ac.jp,
fccchengjl@zzu.edu.cn,
baijie113783501377@126.com,
ghzhao@ha.edu.cn

Abstract. According to the 2021 World Health Organization IDH status prediction scheme for gliomas, isocitrate dehydrogenase (IDH) is a particularly important basis for diagnosis. In general, 3D multimodal brain MRI is an effective diagnostic tool for glioma. However, it is difficult for experienced doctors to only use brain MRI to predict the IDH status. Surgery is necessary to be performed for confirming the IDH. Past studies have confirmed that there is information about IDH in the tumor region on brain MRI images. These studies usually need to mark the glioma area in advance to complete the prediction of IDH status, which takes a lot of time and is costly. In this study, we proposed a multi-task deep learning model using 3D multimodal brain MRI images to achieve glioma segmentation and IDH status prediction simultaneously, which improved the accuracy of both tasks effectively. First, we used a segmentation model to segment the tumor region. Second, the segmentation results were then applied to the original input to obtain the glioma image, and another feature extractor was used to complete the prediction of IDH status. The effectiveness of the proposed method was validated via a public glioma dataset from the BraTS2020. Our experimental results show that our proposed method outperformed state-of-the-art methods with an accuracy of 84.2%, average dice of 79.3%. Thus, we predicted the IDH mutation status for glioma treatment with a 9% increase in accuracy and 2% increase in average dice for glioma treatment.

Keywords: Multi-task learning, Isocitrate Dehydrogenase, glioma segmentation, Computer diagnosis.

1 Introduction

1.1 Background

Brain tumors are classified as either primary or secondary, and gliomas are the most prevalent primary brain tumors [20]. Glioblastoma (GBM) is the most aggressive type of glioma worldwide. Less than 5% of glioblastoma patients survive for five years after diagnosis [26]. According to the 2016 and 2021 World Health Organization (WHO) IDH status prediction schemes for gliomas [8,18], isocitrate dehydrogenase (IDH) status is a particularly important basis for diagnosis. In addition, IDH mutation status has a strong correlation with the prognosis of glioma, and in low-grade gliomas, IDH mutant gliomas have a similar prognosis to IDH wild-type gliomas; however, IDH mutant glioblastomas have a better prognosis than IDH wild-type glioblastomas [8]. A follow-up survey showed that the presence of an IDH mutation predicted a good disease outcome and extended the median survival period of glioblastoma (IDH wild-type, 15 months; IDH mutant, 31 months) [9]. Therefore, it is necessary to predict the IDH mutation status for the treatment of glioma. 3D Magnetic resonance imaging (MRI) is commonly used to diagnose gliomas; see Fig. 1. Generally, 3D brain MRI produces four types of images: T1, T2, T1CE and FLAIR. Each modal has unique characteristics that are very useful for IDH status prediction. Moreover, the spatial information of 3D images is also very effective for IDH status prediction. Although it is difficult for experienced doctors to predict IDH using only MRI images, it can be predicted IDH using machine learning. For high-cost surgical diagnosis, the use of brain MRI to diagnose IDH status using a computer-aided diagnosis system is less costly and easier to perform. Therefore, the development of computer-aided diagnosis system based on machine learning is hot research.

With the development of machine learning, particularly deep learning, many improvements have achieved recently in computer-aided diagnosis. For the prediction of IDH status, Choi et al. [4] proposed a deep learning method using convolutional neural networks (CNNs) and glioma MRI images. Zhang et al. proposed a self-attention algorithm with a squeeze excitation network (SE-SA Net) [33]. However, owing to the limited amount of medical image data, the deep learning method often ignores the importance of medical guidance of doctors and faces the problem of overfitting. Therefore, Yan et al. [31] proposed a machine learning method based on radiomics features and medical knowledge, which achieved more satisfactory results. Furthermore, Zhang et al. [32] improved their SE-SA net with Radiomics features.

However, a general problem with these studies is the need to use glioma regions for IDH status prediction. Deep learning-based tumor IDH status prediction and gene prediction tasks usually use tumor regions as region of interest (ROI) to cut images for IDH status prediction, whereas radiomics based methods directly use tumor boundary masks to obtain internal tumor features. In practical application scenarios, it is high cost to obtain tumor region markers in advance. In addition, the segmentation and labeling of tumor regions can also assist doctors in the diagnosis of glioma. Fur

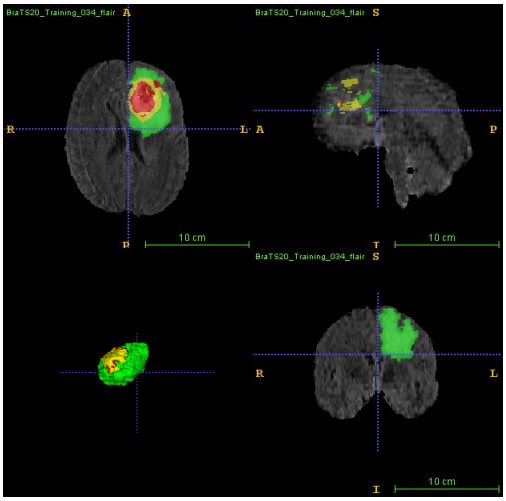

Fig. 1. Example of 3D brain MRI with glioma

-thermore, the IDH status also contains information about the glioma region owing to the existing relationship between the glioma region and the IDH status. There were several related works about multi-task learning for IDH status prediction and glioma segmentation [3,28,29]. Nevertheless, these works used entire glioma representation features for both segmentation and IDH status prediction, which ignored the importance of glioma ROI for IDH status prediction.

1.2 Contributions

To solve the above problems, we proposed a multi-task model based on deep learning, which used 3D multimodal brain MRI images to segment tumor regions simultaneously and use the segmentation results to predict the IDH genotypes. This model contains a segmentation part as well as a IDH prediction part for IDH prediction. Our key contributions are as follows:

A multi-task model for both segmentation and IDH status prediction tasks:

In this research, an integrated model is proposed to simultaneously segment glioma regions and predict IDH status. First, we use a segmentation model to segment the tumor region of the 3D MRI image. Second, we apply the obtained segmentation results to the original input to obtain the image of glioma region. Finally, we used a IDH status prediction model to extract high-level features of the tumor region and use them for IDH status prediction. Compared with the existing methods, we effectively improve the accuracy of segmentation task and prediction task using a multi-task joint model.

Effective ablation study and contrast experiments:

To prove the effectiveness of multi-task model, we conducted an effective ablation study on each part based on public dataset from the multimodal brain tumor segmentation (BraTS2020) [24] challenge. We also conducted contrast experiments with several state-of-the-art methods for brain tumor segmentation and IDH status predic

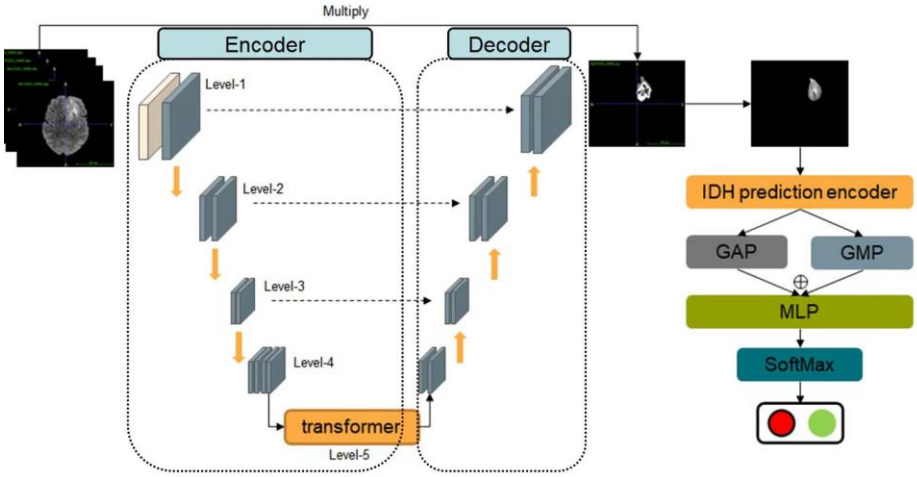

Fig. 2. Overview of multi-task model (In the multi-task network, input of 3D multimodal brain MRI generates the glioma segmentation results firstly. Then the segmentation prediction multiplied with input to get tumor area image, which is processed by IDH prediction network to predict the IDH status.)

-tion. When we compared our method with the current state-of-the-art methods, our proposed performed better both segmentation and prediction results.

The remainder of this paper is organized as follows. The proposed method using multi-task model is described in Section II. Ablation studies and comparative experiments are presented in Section III. Finally, we summarize and conclude our work in Section IV.

2 Methods

2.1 Overview

The overview of the proposed method is illustrated in the Fig. 2. In this research, we proposed a multi-task model, which contained a segmentation part for glioma region segmentation and a IDH status prediction part for IDH status prediction. This model first segmented glioma boundary using a TransBTS [27] as a segmentation network with 3D U-Net [11] encoder, decoder, skip-connection structure and a high-level vision transformer [6] layer. Then, the masks generated by the prediction results were multiplied with the original input to obtain the part containing only the tumor region. Finally, we selected to adopt the same encoder structure and transformer layer for high-level features as IDH status prediction network to extract tumor region features and used them for IDH status prediction.

In the training stage, the two models are trained with the labels of segmentation and IDH status respectively, and the loss function of the entire model is obtained by the weighted average of the loss functions of the two parts. The way the weighted average

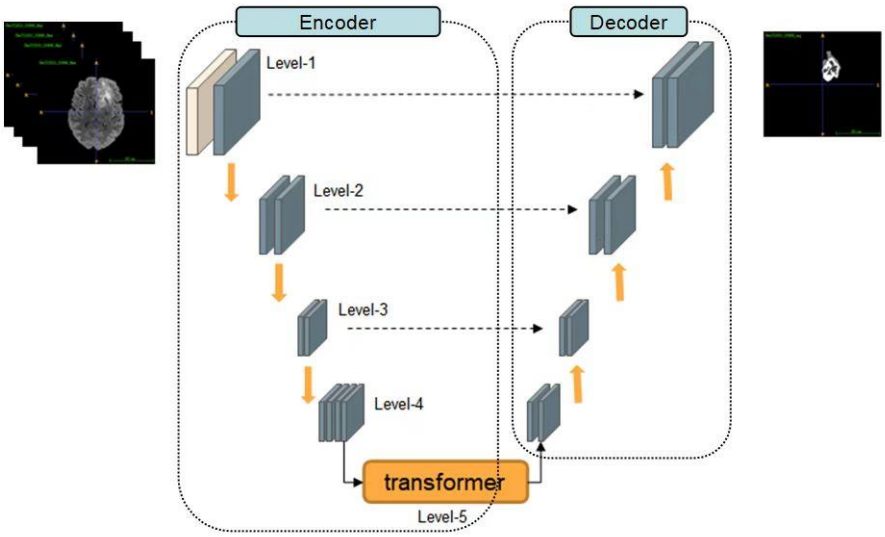

Fig. 3. Segmentation network

calculated will be introduced in the following sections. In the testing stage, only 3D multimodal brain MRI data is input to obtain the predicted tumor boundary mask and IDH status information. In the proposed method, the multi-task model using only one model to complete the requirements of two tasks at the same time and obtain better results. Details of the segmentation model, IDH status prediction model and multi-task learning loss function are introduced in the following.

2.2 Glioma segmentation

For the study of glioma, the automatic segmentation of tumor regions in multimodal brain MRI images by computer can effectively help doctors to make auxiliary diagnosis. In recent years, with the development of deep learning in computer vision, especially in the field of image processing, more and more tumor segmentation methods based on deep learning have appeared. Among them, various studies [15,30] are mainly based on U-Net [23], which proposed skip-connection structure. Compared with the past methods, these deep learning-based studies have greatly improved the accuracy of segmentation. However, limited by the size of the data set and network, most of these methods intercept MRI slices as data for training and testing, which makes the model lack of spatial relevant information about the tumor region. With the progress of equipment, many studies have recently used 3D convolutional neural networks to directly process 3D images, which retained the spatial information of the tumor as much as possible, and further improved the segmentation results, such as the use of 3D U-Net for glioma segmentation [11]. Furthermore, with the development of Transformers using self-attention mechanism in the field of computer vision in the past two years, transformers [6] are applied to more and more medical image segmentation research [19,22,23]. For the segmentation of 3D multimodal brain MRI images,

Wang et al. proposed a TransBTS [27] method based on 3D U-Net and transformer encoder to further improve the segmentation results of Glioma. In our proposed multi-task model, we selected this method as our segmentation network due to the high accuracy of glioma segmentation.

The structure of segmentation network is shown in Fig. 3.

3D CNN-transformer encoders:

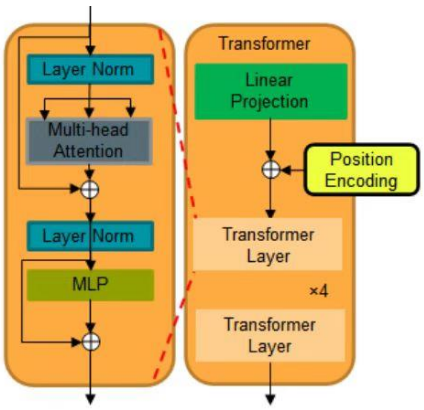

Fig. 4. Transformer unit

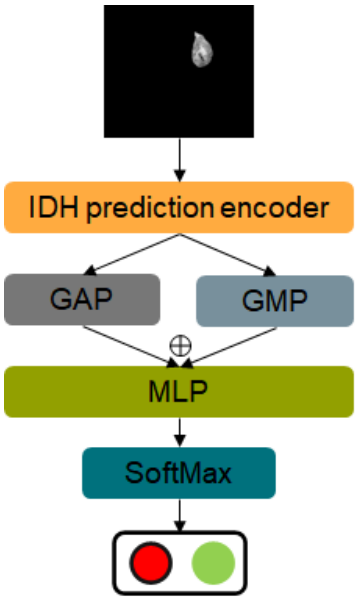

Fig. 5. IDH status prediction network

The dimension of the input $x \in R^{C\times H\times W\times D}$ can be expressed as a 3D resolution of $H \times W \times D$ and C channels. The input was first operated by an initial convolution

with a kernel size of 3 × 3 × 3 to generate feature maps with 16 channels. In order to capture the higher-level features and the spatial relationship of the images, we down sampled the feature maps T times by using residual convolutions. For each down-sampling operation, we used a 3 × 3 × 3 convolution with a stride of 2 the output channel was set to twice the input channel. Each residual convolution block (Resnet Block) consists of two convolution blocks and a residual connection was applied to between the input and output of the residual block. Each convolution block was composed of a batch normalization, a Leaky ReLU function, and a 3 × 3 × 3 convolution layer. The output of encoder E, feature maps $\mathrm{F}_{\lambda}^{\mathrm{C}^{'}\times\mathrm{H}^{'}\times\mathrm{W}^{'}\times\mathrm{D}^{'}}$ in level λ can be denoted by:

$$\mathrm{F}_{\lambda} = E(\boldsymbol{\theta}; x; \lambda) \in \mathrm{R}^{2^{\lambda+3}\times\frac{\mathrm{H}}{2^{\lambda-1}}\times\frac{\mathrm{W}}{2^{\lambda-1}}\times\frac{\mathrm{D}}{2^{\lambda-1}}}, \tag{1}$$

where $\boldsymbol{\theta}$ is the parameters of encoder, x is the input.

The architecture of transformer unit is shown in Fig. 4. Inspired by the vision transformer, TransBTS enhances the long-term connections of images by adding a transformer layer based on self-attention mechanism at the highest-level features (level-4 →level-5). One problem that needs to be solved is that for 3D images, the method of directly cutting the image into several patches and generating tokens is difficult to implement because of the large computational cost. Therefore, in this transformer, we focus more on high-level features. First, we select the features for the last layer, λ=4 and $F_4 \in \mathrm{R}^{128\times\frac{\mathrm{H}}{8}\times\frac{\mathrm{W}}{8}\times\frac{\mathrm{D}}{8}}$. We then embedded these features to 512 channels by 3 × 3 × 3 convolutions, keeping the size constant and reshaping them into $N \times d$, where the $N = \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}$ and is the dimension of feature maps. To learn the positional information, a learnable position embedding $E_{pos} \in \mathrm{R}^{\mathrm{N}\times\mathrm{d}}$ was fused into the patch embedding $E_{pat} \in \mathrm{R}^{\mathrm{N}\times\mathrm{d}}$ by an addition operation, which is denoted as follows:

$$Z_0 = E_{pos} + E_{pat}, \tag{2}$$

where Z_0 denotes the feature of transformer encoder input. Like the regular transformer encoder, the self-attention mechanism triplet (Q, K, V) at l-layer transformer encoder can be denoted as:

$$Q = z^{L-1}w_Q, K = z^{L-1}w_K, V = z^{L-1}w_V, \tag{3}$$

where the $w_Q, w_K, w_V \in \mathrm{R}^{\mathrm{d}\times\mathrm{d}}$ are learnable parameters of the three linear projection layers. Then the self-attention SA can be calculated as:

$$SA(z^{L-1}) = z^{L-1}\left(\frac{Qk^T}{\sqrt{d}}\right)V. \tag{4}$$

Multi-head self-attention (MSA) is a very important part in transformer. It can complete the learning of representations within different subspaces by multiple heads.

Specifically, it divides the input of transformer layer into an independent parts, processes each part in parallel using SA operation, and then projects these concatenated results using a linear projection layer. Therefore, MSA can be denoted as:

$$MSA(z^{L-1}) = Concat(\, SA_1(z^{L-1}), \ldots, SA_n(z^{L-1}))w_O, \tag{5}$$

where $w_O \in \mathrm{R}^{\mathrm{d}\times\mathrm{d}}$ are learnable parameters of the linear projection layer. Then, the output of MSA was converted by an MLP block with a residual skip as the layer output z^L. The transformer encoder can be denoted as follow:

$$z_k^l = z^{L-1} + MSA(z^{L-1}), \tag{6}$$
$$z^L = z_k^l + MLP(z_k^l). \tag{7}$$

Segmentation decoder:

In order to obtain voxel-level segmentation results, we need to pass high-level features through decoder to generate regions for glioma segmentation. In this process, we select 3 × 3 × 3 3D convolutions with stride 2 to generate segmentation results as $H \times W \times D$ from feature maps $d \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}$. Skip-connection was used to further improve the segmentation effect. In this process, we used the features of the corresponding level in the encoder to perform skip-connection. Finally, a 1 × 1 × 1 convolution followed by a SoftMax function as the segmentation layer of the decoder was applied to generate the segmentation result.

2.3 IDH status prediction

It has been confirmed in past studies that the information extracted from the glioma region image has the information of IDH status. Some of these studies used deep learning methods, and some used radiomics to extract the features of the tumor region to predict the IDH status of patients. In this study, we selected to use the 3D CNN-transformer encoder with the same structure as the segmentation network to extract high-level features from tumor regions and use MLP as a classifier to predict IDH categories.

As illustrated in Fig. 5, the IDH status prediction network used the same structure as the segmentation encoder. We selected to adopt level 4 and level 5 feature maps as IDH prediction features to perform the prediction. The sizes of feature maps are $128 \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}$ and $512 \times \frac{H}{8} \times \frac{W}{8} \times \frac{D}{8}$. First, we reshaped feature maps into $128 \times N$ and $512 \times N$. Second, in order to highlight important features, we computed a global Max pooling on the same feature map in addition to the usual global average pooling. Using pooling also helps to save computation because pooling does not involve the computation of parameters. After performing two kinds of pooling on the two feature maps, we obtained the following feature maps: $\mathrm{F}_{\mathrm{GAP}}^{4} \in \mathrm{R}^{128}$, $\mathrm{F}_{\mathrm{GMP}}^{4} \in \mathrm{R}^{128}$, $\mathrm{F}_{\mathrm{GAP}}^{5} \in \mathrm{R}^{512}$, $\mathrm{F}_{\mathrm{GMP}}^{5} \in \mathrm{R}^{512}$. To predict the IDH status using these feature maps, an MLP consisting of multiple linear projection layers was used as a classifier. Finally, we combined these feature maps by concatenation, and used MLP to reduce the

features. A SoftMax converts the features into the probability value of IDH mutation and non-mutation prediction and calculated the loss function using the prediction probability for training. This process can be denoted as:

$$P = SoftMax(MLP(Concat(\mathrm{F}^4_{\mathrm{GAP}}, \mathrm{F}^4_{\mathrm{GMP}}, \mathrm{F}^5_{\mathrm{GAP}}, \mathrm{F}^5_{\mathrm{GMP}}))) \quad (8)$$

2.4 Multi-task model learning

Fig .2 shows the overview of the multi-task model. We can think of the network as the following parts: (1) Segmentation encoder, E, (2) Segmentation decoder, D, (3) IDH status prediction network, P. For a multimodal 3D brain MRI as input $x \in \mathrm{R}^{\mathrm{C\times H\times W\times D}}$, we first input it into the segmentation encoder E to obtain the high-level features of the entire MRI image, and then segmented the tumor boundaries using the decoder D and a SoftMax function to obtain the segmentation probability result $\mathrm{g_p}$. Second, we multiplied the predicted tumor segmentation region by the original input x to generate the tumor region part, $\tilde{\mathrm{x}}\in\mathrm{R}^{\mathrm{C\times H\times W\times D}}$, and input it into the IDH state prediction network to obtain the predicted IDH result $\mathrm{y}^{'}$. The overall computational process of the model can be denoted as:

$$\mathrm{g_p} = \mathrm{SoftMax(D(E(x)))}, \quad (9)$$

$$\mathrm{y}^{'}=\mathrm{SoftMax(P(g_p \times x))}. \quad (10)$$

We got the output of the segmentation model, g_p and the prediction model, $\mathrm{y}^{'}$. The model used dice loss function to train the model for the segmentation task and cross-entropy function to train the model for the prediction task. The loss function can be denoted as:

$$\mathrm{L_{seg}}=1-\frac{\mathrm{g_p g_t}}{\mathrm{g_p+g_t}}, \quad (11)$$

$$\mathrm{L_{idh}}=-\sum \mathrm{ylogy}^{'}, \quad (12)$$

where the g_t is the ground truth of tumor mask, and y is the label of IDH status.

In order to jointly train the multi-task model, it is necessary to select appropriate weights for each task to train. However, manually selected weights take a lot of time to adjust and are also not practical. Therefore, we refer to the training method based on uncertainty model [22], use maximum likelihood estimation to calculate the weight of each model, and define two learnable parameters to dynamically adjust the weight of the loss function, so as to obtain better training results. The total loss function for the multi-task model is denoted as:

$$\mathrm{L_{joint}} = \frac{\mathrm{L_{seg}}}{2\sigma^2_{\mathrm{seg}}} + \frac{\mathrm{L_{idh}}}{2\sigma^2_{\mathrm{idh}}} + \log\sigma_{seg}\sigma_{idh}, \quad (13)$$

where the σ_{seg} and σ_{idh} are the learnable parameters for network training. These two values are initialized to 1 to ensure that the training starts with the same weight for both tasks.

3 Experiments

3.1 Dataset

We obtained 148 multimodal MRI images from the public dataset from the multimodal brain tumor segmentation (BraTS2020) [24] challenge training dataset as our training dataset, whereas 70 multimodal brain MRI images from the BraTS2020 validation set as test set. Part of BraTS2020 data belongs to The Cancer Imaging Archive (TCIA) [14], which can be available from our public repository. Genomic information is provided from The Cancer Genome Atlas (TCGA) [19]. The MRI images used in this experiment included two types: IDH wild-type and mutated-type. This training dataset contained 57 patients with an IDH mutation and 91 patients with wild type IDH. The test dataset contained 32 mutant cases and 38 wild cases.

Each patient data used in the experiments had four MR image modalities (T1, T2, T1ce, and FLAIR) with IDH status information. Registration of multimodal brain MRI images is very important due to the differences between different modalities. In BraTS2020, both training set and validation set data are labeled by experts and registered on T1 modal. In our experiments, we directly use the registered data. In this dataset, there are three parts of annotations for the tumor region, which are respectively whole tumor (WT), tumor core (TC), and enhancing tumor (ET).

3.2 Experiments

To confirm the influence of multi-task learning between segmentation task and IDH status prediction task, we performed the following ablation studies. The first model (Model 1) was the IDH status prediction model. We trained the IDH network same as the multi-task learning in the same manner and validated the IDH status prediction results. The second model (Model 2) was the segmentation model. We used the same method as the proposed method to segment the tumor area. The third model (Model 3) was the proposed novel multi-task model, which used the segmentation prediction as the tumor area to predict the IDH status using IDH network. We also conduct comparative experiments with the state-of-the-art methods for single task.

We selected the area under the curve (AUC), accuracy (Acc), sensitivity (Sens), and specificity (Spec) as our IDH status prediction evaluation measures. We selected the Dice coefficient (Dice) as the segmentation evaluation measure. For the three parts of tumor area, we computed dice for the three parts independently and finally compute the three-part average.

We used the Adam optimizer with a batch size of 1 and a learning rate of 0.001, without data shuffling in our deep learning model. All the training was conducted on two Nvidia GeForce RTX 3090 24GB GPU. The proposed method achieves the best performance. In order to ensure the effectiveness of the segmentation results, we first pre-train the segmentation network and then add the IDH status prediction network for fine-tuning.

3.3 Experimental results

The results of the ablation studies are presented in Table 1. As shown in Table 1, Models 1 and 2 show the results of using only IDH status prediction or segmentation network. Model 3 shows that better results are achieved when integrating the two tasks. We thought that because the tumor region was used as the input of the IDH state prediction network, the prediction ability of the IDH state was improved compared with the method of inputting the whole image, and the information of the IDH state further improved the ability of the segmentation network to locate the tumor region.

Finally, we conducted a comparative experiment with the current state-of-the-art methods [7,12,13,23,28,29]. The results are presented in Table 2. Our proposed method achieved better IDH status prediction and segmentation performance than the state-of-the-art methods which means that our multi-task model performs better. The segmentation performance of our proposed multi-task method in TC and ET is significantly higher than other models, which is because IDH is more sensitive to the information within the internal boundaries of the tumor rather than glioma edema region.

4 Discussion and Conclusion

In this study, we proposed a multi-task model that combines both IDH status prediction task and brain tumor segmentation. From the experimental results of our proposed method, we achieved AUC of 88.2%, accuracy of 84.2%, sensitivity of 71.8%, specificity of 94.7% for IDH status prediction task and Dice of whole tumor of 86.84, tumor core of 76.33, enhancing tumor of 74.66, which achieved better performance than state-of-the-art methods. Multi-task learning is a field worth exploring in the field of medical image processing. Multi-task learning not only improves the performance of two tasks using the complementarity of multiple tasks to a certain extent, but also needs computer-aided diagnosis systems that can complete multiple tasks to assist doctors in diagnosis in the medical field. Since the IDH status prediction task needs to predict using the tumor target region, we used the tumor region generated by the segmentation task as the input of the prediction network to complete the need to achieve two tasks at the same time. Meanwhile, the IDH network further improves the performance of the segmentation network in the joint learning process.

In the future, we would like to improve our multi-task network by more efficient segmentation networks as well as IDH status prediction networks. At present, the problem we face is that for 3D medical images, the 3D network model we now design is large and has some shortcomings in the training process and practicability. We hope to explore more connections between multiple tasks and use smaller networks to achieve better results with less cost.

Table 1. Ablation study of single task and multi-task model

	AUC(%)	Acc(%)	Sens(%)	Spec(%)	WT (mean%)	TC (mean%)	ET (mean%)	Avg (mean%)
Model1	88.2	81.4	62.5	**97.3**	-	-	-	-
Model2	-	-	-	-	**88.61**	74.46	68.65	77.24
Model3	**88.2**	**84.2**	**71.8**	94.7	86.84	**76.33**	**74.66**	**79.27**

Table 2. Comparison with the state-of-the-art methods

	AUC(%)	Acc(%)	Sens(%)	Spec(%)	WT (mean%)	TC (mean%)	ET (mean%)
SENet101[12]	79.6	77.1	65.6	86.8	-	-	-
DenseNet121[25]	78.2	77.1	65.6	86.8	-	-	-
ResNet50[13]	77.5	74.2	62.5	84.2	-	-	-
U-Net[23]	-	-	-	-	88.79	71.35	71.84
V-Net[7]	-	-	-	-	**88.84**	70.44	69.67
UNETR[1]	-	-	-	-	87.42	68.68	70.39
SGPNet[28]	81.3	80.0	62.5	94.7	87.04	69.33	71.43
CMSVNet[29]	75.9	72.9	56.3	86.8	78.29	58.43	67.50
Proposed method	**88.2**	**84.2**	**71.8**	**94.7**	86.84	**76.33**	**74.66**

needs to predict using the tumor target region, we used the tumor region generated by the segmentation task as the input of the prediction network to complete the need to achieve two tasks at the same time. Meanwhile, the IDH network further improves the performance of the segmentation network in the joint learning process.

In the future, we would like to improve our multi-task network by more efficient segmentation networks as well as IDH status prediction networks. At present, the problem we face is that for 3D medical images, the 3D network model we now design is large and has some shortcomings in the training process and practicability. We hope to explore more connections between multiple tasks and use smaller networks to achieve better results with less cost.

References

1. A. Hatamizadeh et al., “UNETR: Transformers for 3D medical image segmentation,” arXiv:2103.10504, 2021.
2. Chen, Y.-W., and Jain, L.C., (Eds): *Deep Learning in Healthcare*, Springer, 2020.
3. Cheng J, Liu J, Kuang H, Wang J. A Fully Automated Multimodal MRI-Based Multi-Task Learning for Glioma Segmentation and IDH Genotyping. IEEE Trans Med Imaging. 2022 Jun;41(6):1520-1532. doi: 10.1109/TMI.2022.3142321. Epub 2022 Jun 1. PMID: 35020590.
4. Choi, Yoon Seong, et al. "Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics." Neuro-oncology,2020.
5. David N Louis, Arie Perry, Pieter Wesseling, Daniel J Brat, et al. “The 2021 WHO IDH status prediction of Tumors of the Central Nervous System: a summary”, *Neuro-Oncology*, Volume 23, Issue 8, Pages 1231–1251, 2021,
6. Dosovitskiy, Alexey, et al. “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.

7. F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully convolutional neural networks for volumetric medical image segmentation,” *in Proc. 4th Int. Conf. 3D Vis. (3DV)*, pp. 565–571, Oct. 2016.
8. Hai Yan, D.Williams Parsons, Genglin Jin, et al. “IDH1 and IDH2 mutations in gliomas.” *N Engl J Med*, 2009.
9. Han S, Liu Y, Cai SJ, Qian M, et al. “IDH mutation in glioma: molecular mechanisms and potential therapeutic targets”. *Br J Cancer*. PMID: 32291392; PMCID: PMC7250901. Doi: 10.1038/s41416-020-0814-x, 2020.
10. Huang, H., Zheng, H., Lin, L., et. al. “Medical Image Segmentation with Deep Atlas Prior.” *IEEE Trans. Medical Imaging*, Vol.40, No.12, pp.3519-3530, 2021.
11. J. Cheng, J. Liu, L. Liu, Y. Pan, J. Wang, et al. “Multi-level glioma segmentation using 3D U-Net combined attention mechanism with atrous convolution,” *in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 1031–1036
12. J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit*., pp. 7132–7141, Jun. 2018,
13. K. Chang, H. X. Bai, H. Zhou, and C. Su, “Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR imaging,” *Clin. Cancer Res*., vol. 24, no. 5, pp. 1073–1081, 2018.
14. K. Clark et al., “The cancer imaging archive (TCIA): Maintaining and operating a public information repository,” *J. Digit. Imag*., vol. 26, no. 6, pp. 1045–1057, 2013.
15. K. Qi et al. “X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies,” in Proc. Int. Conf. *Med. Image Comput. Comput.-Assist.* Intervent. Cham, Switzerland: Springer, pp. 247–255, 2019.
16. Kitrungrotsakul, T., Chen, Q., Wu, H., et. al. “Attention-RefNet: Interactive Attention Refinement Network for Infected Area Segmentation of COVID-19,” *IEEE Journal of Biomedical and Health Informatics*, Vol.25, No.7, pp. 2363-2373, 2021.
17. Liang, D., Lin, L., Hu, H., et. al. “Combining Convolutional and Recurrent Neural Networks for IDH status prediction of Focal Liver Lesions in Multi-Phase CT Images,” *In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018, Lecture Notes in Computer Science*, LNCS7951, Springer, pp.666-675, 2018.
18. Louis DN, Perry A, Reifenberger G, et al. "The 2016 World Health Organization IDH status prediction of tumors of the central nervous system: a summary". Acta Neuropathol; 131(6):803–820, 2016.
19. M. Ceccarelli et al., “Molecular proling reveals biologically discrete subsets and pathways of progression in diffuse glioma,” *Cell*, vol. 164, no. 3, pp. 550–563, 2016.
20. Ostrom QT, Gittleman H, Xu J, et al. “Primary brain and other central nervous system tumors diagnosed in the United States in 2009-2013”. *CBTRUS statistical report,* Neuro Oncol, ;18(suppl_5): v1–v75, 2016.
21. Peng, L., Lin, L., Hu, H., et. al. “IDH status prediction and Quantification of Emphysema Using a Multi-Scale Residual Network,” *IEEE Journal of Biomedical and Health Informatics*, Vol.23, No.6, pp.2526-2536, 2019.
22. R. Cipolla, Y. Gal, and A. Kendall, et al. “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics,” *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 7482–7491, Jun. 2018.
23. Ronneberger, Olaf, Philipp Fischer, Thomas Brox, et al. “U-Net: Convolutional networks for biomedical image segmentation,” *International Conference on Medical image computing and computer-assisted intervention*, Springer, Cham, 2015.

24. S. Bakas et al., “Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Scientific data*, vol. 4, Art. no. 170117, Sep. 2017.
25. S. Liang et al., “Multimodal 3D DenseNet for IDH genotype prediction in gliomas,” *Genes*, vol. 9, no. 8, p. 382, Jul. 2018.
26. Steven De Vleeschouwer S. “Glioblastoma [Internet].” *Brisbane (AU): Codon Publications*; PMID: 29251853, 2017. Doi: 10.15586/codon.glioblastoma, 2017.
27. Wang, Wenxuan, et al. “Transbts: Multimodal brain tumor segmentation using transformer,” I*nternational Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2021.
28. Y. Wang, Y. Wang, C. Guo, S. Zhang, and L. Yang, “SGPNet: A threedimensional multitask residual framework for segmentation and IDH genotype prediction of gliomas,” Comput. Intell. Neurosci., vol. 2021, pp. 1–9, Apr. 2021.
29. Y. Zhou et al., “Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images,” Med. Image Anal., vol. 70, May 2021, Art. no. 101918.
30. Y. Zhou, W. Huang, P. Dong, Y. Xia, S. Wang, et al. “D-UNet: A dimension-fusion U shape network for chronic stroke lesion segmentation,” *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 940–950, May 2021
31. Yan, J., Zhang, B., Zhang, S. et al. “Quantitative MRI-based radiomics for noninvasively predicting molecular subtypes and survival in glioma patients.” *npj Precis*. Doi: 10.1038/s41698- 021-00205-z, 2021.
32. Zhang, X., Shi, X., Iwamoto, Y., et al. “IDH mutation status prediction by a radiomics associated modality attention network”. *Visual Comput*. Doi: 10.1007/s00371-022-02452-y, 2022.
33. Zhang, Xinran, Iwamoto Yutaro, et al. “IDH Mutation Status Prediction by Modality Self Attention Network.” *Innovation in Medicine and Healthcare*. Smart Innovation, Systems and Technologies, vol 242. Springer, Singapore, pp.51-57, 2021.