An Intra- and Inter-Modality Fusion Model with Invariant- and Specific-Constraints Using MR Images for Prediction of Glioma Isocitrate Dehydrogenase Mutation Status

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, Email: is0490sr@ed.ritsumei.ac.jp (X.S.), yin-li@fc.ritsumei.ac.jp (Y.L.)
[2] The Affiliated Hospital of Zhengzhou University, Zhengzhou, China, Email: fccchengjl@zzu.edu.cn (J.C.), baijie113783501377@126.com (J.B.), ghzhao@ha.edu.cn (G.Z.)
*Correspondence: chen@is.ritsumei.ac.jp (Y.-W.C.)

***Abstract*—In the 2021 World Health Organization classification of gliomas, it is proposed that Isocitrate Dehydrogenase (IDH) plays a key role. The prognosis of glioma is largely affected by IDH mutation status. Therefore, IDH mutation status needs to be predicted in advance before surgery. In the past decade, with the development of machine learning, more and more machine learning methods, especially deep learning methods, have been applied to the development of computer-aided diagnosis systems. At present, in this field, many deep learning and radiomics based methods have been proposed for IDH prediction using multimodal Magnetic Resonance Imaging (MRI). In this study, we proposed an intra- and inter-modality fusion model with invariant- and specific- constraints to improve the performance of IDH status prediction. First, MRI-based radiomics features were fused with deep learning features in each modality (intra-modality fusion) and then the features extracted from each modality of brain MRI were fused by using an inter-modality fusion model with invariant and specific constraints. We experimented our proposed method on the dataset provided by the Affiliated Hospital of Zhengzhou University in Zhengzhou, China and demonstrated the effectiveness of the proposed method. In our study, we propose two inter-modality fusion models, and our experimental results show that our best proposed method outperformed state-of-the-art methods with an accuracy of 0.79, precision of 0.80, recall of 0.75, and F1 score of 0.78. Thus, we predicted the IDH mutation status for glioma treatment with a 2% increase in accuracy and 4% increase in precision to predict the IDH mutation status for glioma treatment.**

***Keywords*—glioma, isocitrate dehydrogenase, multi-modal learning, computer diagnosis**

I. Introduction

Brain tumors can be classified as primary and secondary tumors, with glioma being the most common primary brain tumor [1]. Glioblastoma (GBM) is the most aggressive type of glioma worldwide. Less than 5% of glioblastoma patients survive for five years after diagnosis [2]. According to the 2016 and 2021 World Health Organization (WHO) classification schemes for gliomas [3, 4], it is proposed that Isocitrate Dehydrogenase (IDH) plays a key role. 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 [5]. 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) and anaplastic astrocytoma (IDH wild-type, 20 months; IDH mutation, 65 months) [6]. Therefore, for the treatment of glioma, it is necessary for us to predict IDH mutation status in advance. Magnetic Resonance Imaging (MRI) is commonly used to diagnose gliomas. Generally, brain MRI produces four types of images: T1, T2, T1CE and FLAIR. Each modal has distinct features that are very useful for IDH state prediction. Although it is difficult for experienced doctors to predict IDH using MRI images, it can be predicted IDH using machine learning.

With the development of machine learning, particularly deep learning, many improvements have achieved recently in computer-aided diagnosis [7–15]. For the prediction of IDH status, Choi *et al.* [7] proposed a deep learning method using Convolutional Neural Networks (CNNs) and glioma MRI images. Zhang *et al.* [8] proposed a self-attention algorithm with a squeeze excitation network (SE-SA Net). 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.* [9] proposed

Manuscript received April 3, 2023; revised May 21, 2023; accepted July 5, 2023.

a machine learning method based on radiomics features and medical knowledge, which achieved more satisfactory results. Furthermore, Zhang *et al.* [10] improved their SE-SA net with Radiomics features.

Modality fusion has been widely used in previous studies. The radiomics-based machine learning method proposed by Yan *et al.* [9] adopted a regression structure to fuse the prediction results from four modalities as a fusion result, which improved the overall performance compared to a single modality. In the deep learning-based methods proposed by Zhang *et al.* [8], a squeeze-and-excitation network [16] was used for modality fusion by adding weights to each modality in the training stage, which also improved performance compared to the original deep CNN. Therefore, multimodal fusion of MRI images is an effective way to improve the performance of IDH mutation state prediction tasks. However, these fusion methods only consider multimodal fusion between each modality of MRI images, and the fusion methods used are relatively simple, which may ignore some modality invariant and specific information.

Our approach to improve the performance of IDH mutation status prediction is via an intra-modality fusion between deep features extracted from 2D tumor slices and radiomics features extracted from the 3D tumor area as well as an inter-modality fusion between each modality of MRI images.

In our previous work [17], an intra- and inter- modality fusion model was proposed to improve the performance of IDH mutation prediction. First, both radiomics features are fused with deep learning features in each modality (intra-modality fusion). Second, four Bayesian-Regularization Neural Networks (BRNN) [18] classifiers predict the probability of each modality using a learnable weight inter-modality regression fusion model to fuse the four modalities and predict the overall result. However, there are two main challenges in the inter-modality fusion stage:

(1) Owing to the different distributions between modalities, it is difficult to fuse them directly using a simple concatenation structure. Therefore, each modality must be aligned first for fusion.

(2) Multimodal MRI information contains redundant information for the IDH mutation status prediction task Therefore, it is necessary to make different modalities orthogonal to reduce redundancy and enhance complementarity, which could maximize the effective information of multi-modality data.

In addition, in this study, we proposed an invariant-and specific-constraint inter-modality fusion model to improve the performance of the inter-modality fusion stage. In this new inter-modality fusion model, we added invariant and specific constraints to extract features from multimodal MRI data, which improved the performance of the prediction model, especially the accuracy of positive cases (that is, precision).

- **Contributions**: Whereas most the state-of-the-art methods only focus on multimodal fusion within a single modality and simple fusion methods such as concatenation and regression, we propose an intra-modality fusion based on deep features and radiomics features and an inter-modality fusion using invariant- and specific-constraints inter-modality fusion. Our key contributions are as follows:
- **A novel intra- and inter-modality fusion multimodal fusion model**: We proposed an intra- and inter-modality fusion multimodal fusion model to improve the performance of IDH mutation status prediction. In each model, 3D MRI images and tumor area annotations were used to extract the radiomic features of glioma. In addition, we used a deep learning network to extract the hidden deep information in the image by selecting 2D slices with obvious tumor areas. Finally, the features extracted by radiomics and deep learning were combined, and a variety of statistical methods were used to screen the features to obtain the features useful for IDH status prediction. These features extracted from four modalities (T1, T1CE, T2, T2-FLAIR) of MRI images, were used for inter-modality fusion which significantly improved the overall prediction ability of the model.
- **An inter-modality fusion model with invariant- and -specific constraints**: In the inter-modality fusion stage, we proposed inter-modality fusion using invariant and specific constraints for multimodal MRI fusion. An invariant encoder was used to extract invariant features between different modalities, whereas several specific encoders were used to extract specific features that could reduce the redundancy and enhance the complementarity of each modality. Finally, we proposed similarity loss and difference loss as constraints to learn these features which improved the performance of the entire model.
- **Effective ablation study and contrast experiments**: To prove the effectiveness of each fusion part and the invariant- and specific-constraint inter-modality fusion, we conducted an effective ablation study on each part based on our data set from the First Affiliated Hospital of Zhengzhou University (FHZU) in Zhengzhou, China. We also conducted new experiments on our inter-modality fusion model and recent research on deep learning fused with radiomics, which was proposed by Zhang *et al.* [18]. When we compared our method with the current state-of-the-art methods, our proposed performed better.

A preliminary version of this work was presented as a four-page conference paper at the 2022 IEEE The Engineering in Medicine and Biology Conference (EMBC) [17]. The present draft involves both substantial conceptual and experimental extensions including the following:

(1) We evolved the inter-modality fusion model with an invariant and specific constraints inter-modality fusion.

(2) Through a series of input experiments, we discovered the optimal solution of the inter-modality fusion method.

(3) We integrated the new inter-modality fusion model with the previous model and performed ablation studies to demonstrate the effectiveness of our new model.

(4) we also added a new comparative experiment, adding the latest experimental results of related studies proposed by Zhang *et al.* [10] to our new method.

The remainder of this paper is organized as follows. The proposed method using intra- and inter-modality fusion models is described in Section II. Ablation studies and comparative experiments, as well as an exploration of inter-modality fusion model inputs, are presented in Section III. Finally, we summarize and conclude our work in Section IV.

II. Methods

A. Overview

In our research, we propose an intra- and inter-modal fusion model based on deep learning and machine learning methods. This model first fuses the deep learning features in each modal with the radiomics features (intra modality fusion), and then uses the inter-modal fusion model to fuse the features extracted from each modal of the MRI images. The overview of our proposed method is illustrated in Fig. 1. We extracted radiomics features from the 3D MRI images by four different modalities with their annotations. In addition, four deep learning classifiers were trained as feature extractors of the four models. At this stage, we adopted to select a self-attention convolutional network as the deep feature extractor, which achieves better performance for IDH status prediction using brain MRI.

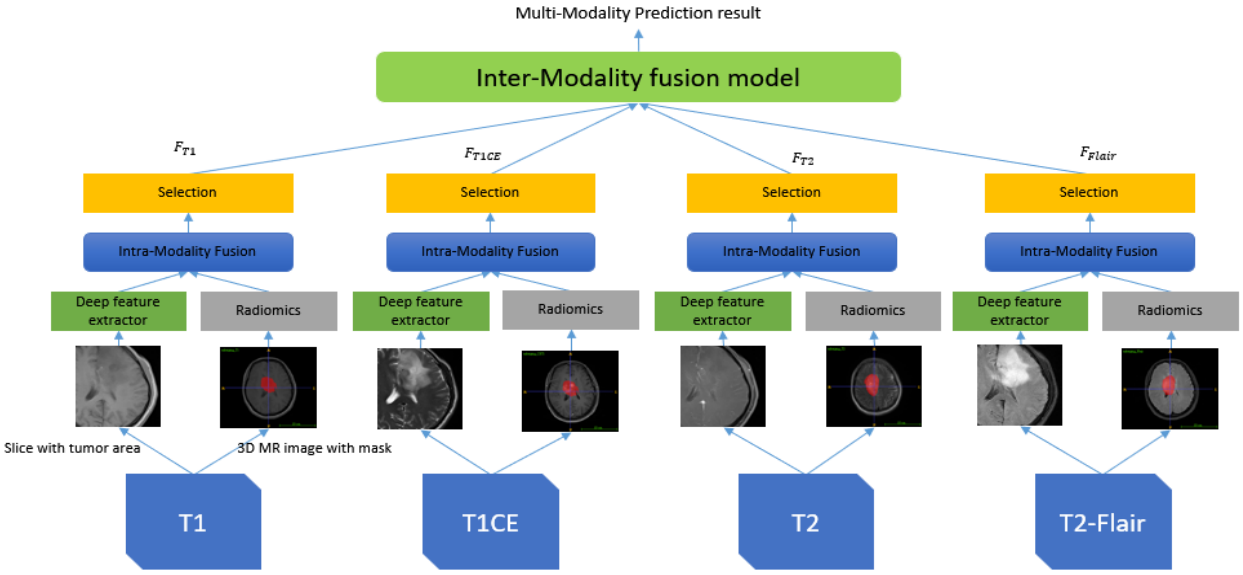

Figure 1. Overview of the intra- and inter-modality fusion model. For each modality of MRI images, the model first fuses both MRI-based radiomics features with deep learning features in each modality (intra-modality fusion) and then the features extracted from each modality of brain MRI are fused by using an inter-modality fusion model with invariant- and specific-constraints. The details of the whole model are introduced in Section II.

After extraction, the radiomics and deep learning features were concatenated into the fusion feature in the intra-modality fusion block. Due to the large number of fused features, and some not useful features we used several statistical methods to filter the features. After this stage, we proposed a deep learning method with invariant and specific constraints to extract invariant and specific features to improve the performance of the inter-modality fusion model. In this inter-modality fusion model, we use several encoders to train the fusion model by invariant and specific constraints, which could make modalities orthogonal to reduce the redundancy of each modality.

Details of the inter-modality fusion model with invariant- and specific-constraints are introduced in the following.

B. Radiomic Feature Extraction

Radiomics is a method of extracting a large number of features from medical images using data representation algorithms [19], which is widely used in computer-aided diagnosis systems. These radiomics features contain a large number of tumor features that are not recognizable by the naked eye. Traditional radiomics methods [9] to solve medical image problems include three parts, which are feature extraction, feature selection and classification or prediction.

The method uses the Python open-source package platform Pyradiomics 2.0.0 to extract glioma radiomics features from 3D MRI images with tumor region annotations (https://www.radiomics.io/pyradiomics.html). 873 tumor features were extracted from T1, T1CE, T2 and T2 FLAIR images, respectively. We referred to the method proposed by Yan *et al.* [9]. After manually selecting feature types, we can automatically extract the required features from MRI data by Pyradiomics. These features can be grouped as follows: histogram-based (n=18), shape and size-based (n=13), textural (n=68), wavelet-based (n=430), Laplacian of Gaussian (LoG) filter-based

(n=258), and those from the gradient magnitude of the given MRI volumes (n=86). Finally, we extracted 3492 (873×4) features from the four MRI modalities.

C. Deep Feature Extraction

In recent years, the use of machine learning has increased, particularly in medical diagnosis. Deep learning methods can automatically extract features from medical and radiology knowledge that are difficult for experienced doctors to observe. Although deep learning methods have made progress in the field of image processing, they are prone to overfitting problems on small medical datasets, especially brain MRI image datasets.

Therefore, a lightweight and efficient deep learning network is needed to solve the overfitting problem. Self-Attention Net (SA-Net) proposed in 2020 CVPR [20] is a light network with a pixel-based self-attention block. The following problems often arise with normal convolutional layers: as the receptive field increase, the number of parameters increases, and the convolution may lack rotation invariance. To overcome these limitations, self-attention networks have been proposed.

Compared with the ordinary stationary convolution kernel calculation method, SA-Net uses a 1×1 convolution layer to calculate the relationship between the target pixel and other pixels as a weighted average. Fig. 2 shows the pixel-wise self-attention network. The self-attention block is defined Eq. (1):

$$Y_i = \sum_{j \in R(i)} \gamma(\delta(x_i, x_j)) \odot \beta(x_j) , \quad (1)$$

where δ is the relation function between the target pixel and local neighbors and $\odot$ means a sum product calculation; subtraction is used for the relation function as noted in Eq. (2):

$$\delta(x_i, x_j) = \varphi(x_i) - \psi(x_j) \quad (2)$$

The output of the relation function is a single vector that shows the features x_i and x_j. γ, φ, ψ, β are 1×1 convolution layers. The output of γ is the weight of pixel x_j. In this case, the parameters do not increase when the receptive field becomes larger, and the rotation invariance is retained, which could alleviate the overfitting problem to some extent.

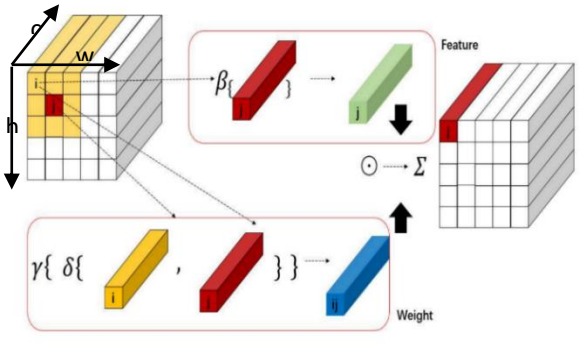

Figure 2. Pixel-based self-attention network.

After referring to the performance of various deep learning methods in IDH state prediction, the proposed method selected SA-Net introduced above as the deep feature extractor. As shown in Figs. 3 and 4, a common SA-Net was used as the feature extractor. We added a new full connection layer in front of the full connection layer (classifier) of the ordinary SA-Net as our feature extraction layer. In the training phase, we trained the deep network using IDH status labels using only the tumor slice images (366 cases) in the training dataset (Fig. 3). After that, we input the training and test cases for feature extraction (Fig. 4). The MRI image is fed into the trained network, which outputs 128 dimensional features (128→2) before the FC layer predicts the classification result. In order to use different representative features among the four modalities, four deep networks are trained using MRI images of the four modalities and the deep features based on the different modalities are extracted; this is to ensure that the information of each modality does not interfere with each other during intra-modality fusion. Finally, for each modality, this self-attention network as a deep feature extractor obtained 128-dimensional features.

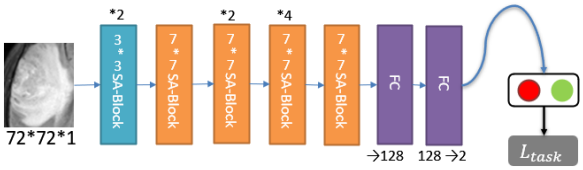

Figure 3. Deep feature extractor based on the self-attention network in training stage (The number after * represents the number of times this module is repeated in the network, for example, *2 means that this module is used twice in a row).

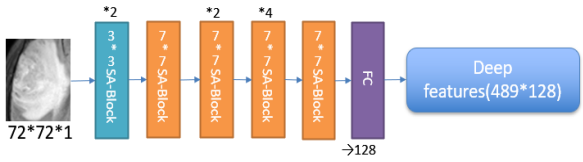

Figure 4. Self-attention network for feature extraction (The number after * represents the number of times this module is repeated in the network, for example, *2 means that this module is used twice in a row).

D. Intra-Modality Fusion and Selection

In other words, deep features share the same information with radiomics features to some extent, as both are related to IDH state prediction. Therefore, using the concatenate function to combine deep learning and radiomics features as global features can be considered as a kind of intra-modality fusion. However, it is not very intuitive to determine whether the selected features have a positive effect on IDH state prediction. Too many unfiltered features will interfere with the classification results and affect the classifier training effect. It is difficult for the classifier to train normally when the features with such a large dimension are directly input; Therefore, like commonly used methods in radiomics, we use statistical methods to perform feature selection first. Three-step feature selection was used to filter the features.

First, the Mann-Whitney U test [21] was used to eliminate the features that had no significant difference ($p \geq 0.05$) between the two groups. In the second step, to assess whether there is a correlation between feature pairs, we perform an analysis using Pearson's correction coefficient (PCC) [22] and randomly remove one feature from the feature pairs with a correlation coefficient $r > 0.9$. In the third step, we use the popular Least Absolute Shrinkage Selection Operator (LASSO) regression [23] to select informative features with non-zero coefficients by 10-fold cross validation. We input 873-dimensional radiomics features and 128-dimensional deep learning features. Four modalities with a total of 4004 dimensions (1001×4) are used for feature selection. Each modality performs the selection operation independently and saves the results separately. The features after selection were four T1 features (three radiomics and one deep feature), five T2 features (four radiomics and one deep feature), seven T1CE features (six radiomics and one deep feature) and seven FLAIR features (six radiomics and one deep feature).

E. Inter-Modality Fusion Model

We adopted to select Bayesian Regularized Neural Network (BRNN) as the classifier, in our previously proposed method. Four different classifiers are trained using the data of the four filtered modalities. The four probabilities (P_{T1}, P_{T1CE}, P_{T2}, P_{Flair}) of the IDH mutation were obtained from the four corresponding modality images. Finally, we use a linear regression model to learn the weights of each modality by using the predicted results of the four modalities and the actual labels to fuse the four modalities.

Usually, an average or learnable weighted-average model is considered in multimodal fusion. Owing to the importance of multimodal medical images in medical diagnosis, research on multimodal medical image processing based on deep learning has increased annually in recent years [24]. Therefore, an increasing number of multimodal fusion methods have been proposed such as the input-level fusion network proposed by Pereira *et al.* [25]. Similarly, the research on the fusion of multi-media modality fusion such as audio and video is also an important part of the multimodal learning. Hazarika *et al.* [26] proposed an invariant- and specific-constraints loss to improve the multimodal fusion performance of audio, video and text data. As shown in Fig. 5, for example, we have two modality features: u_1 and u_2. Because of the different distributions between modalities, it is difficult to fuse them directly using simple concatenation structure. Therefore, each modality must be aligned first for fusion. Multimodality MRI information contains a lot of repeated information, which is redundant for the prediction task. Therefore, it is necessary to make different modalities orthogonal to reduce redundancy and enhance complementarity, which could maximize the effective information of multi-modality data. We used an encoder with shared weights to process u_1 and u_2 to extract invariant features h_1^i and h_2^i between modalities, and additionally use two different encoders to process the u_1 and u_2 to extract specific features h_1^s and h_2^s.

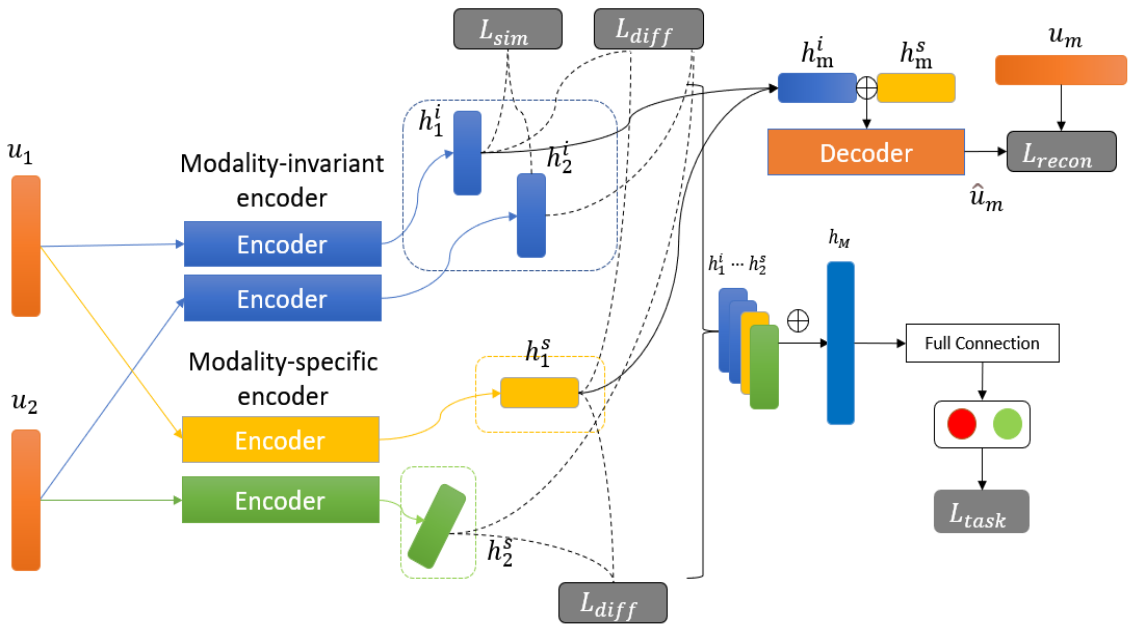

Figure 5. Invariant- and Specific-constraints inter-modality fusion model. u_1 and u_2 are two different modalities. A share-weight encoder (blue) is used for extracting invariant features from two modalities while two different encoders (yellow and green) are used for extracting specific features. Finally, invariant features and specific features are concatenated as one feature for the prediction. The model is trained with invariant- and specific-constraints, which the details are introduced in Section II. E. Inter-Modality Fusion Model.

To make h_1^i and h_2^i in an invariant subspace and h_1^s and h_2^s in two different specific subspaces, we used two loss functions L_{sim} and L_{diff} as constraints to learn this model. All features in each subspace were concatenated as one fusion feature, h_M, which was fed into a full connection layer to predict the IDH mutation status. Finally, the cross-entropy loss L_{task} estimated the quality of the prediction during the training stage. L_{sim} is an invariant constraint for aligning each modality with a common subspace to extract invariant features during the training stage. Minimizing this loss could reduce the discrepancy between the shared representations of each modality, which could better align the features extracted from one encoder into a common subspace. In our proposed method, we used KL-divergence as an invariant constraint to calculate the L_{sim} between each modality:

$$L_{sim}(h_1^i, h_2^i) = \sum p(h_1^i) log \frac{p(h_1^i)}{p(h_2^i)} \quad (3)$$

where $p(h_1^i)$ means the probability distribution of h_1^i.

L_{diff} is the difference loss to learn the specific features of each modality. This loss ensures that the modality-invariant and modality-specific representations capture different aspects of the input. When two vectors are orthogonal, they are completely unrelated. Therefore, we minimized the vector product of the features to compute L_{diff}:

$$L_{diff} = ||h_1^i - h_1^s||_2 + ||h_2^i - h_2^s||_2 + ||h_1^s - h_2^s||_2 \quad (4)$$

We also designed a reconstruction to assist in training the model. An obvious problem in training is that a share-weight encoder using only the KL-divergence invariant constraint can make both features become zero, which is the fastest way to make their distributions similar. Moreover, we expect that the invariant and specific features extracted from the same modality are complementary. Therefore, in the reconstruction part, we concatenated the invariant and specific features of the same modality and decoded the concatenated features to reconstruct them to the original. For example, for u_1, we concatenated the features h_1^i and h_1^s as h_1 and used a decoder to reconstruct this feature as $\hat{u}_1$. We used the Euclidean distance as the loss function of the reconstruction part to judge whether the reconstruction was successful and used it to prevent the features in the common subspace becoming zero, and to enhance the complementarity of the invariant- and specific-features extracted from the same modality. The reconstruction loss L_{recon} is expressed as:

$$L_{recon} = ||\hat{u}_m - u_m||_2 \quad (5)$$

The overall learning of the model was performed by minimizing:

$$L = L_{task} + \alpha L_{sim} + \beta L_{diff} + \gamma L_{recon} \quad (6)$$

In the proposed method, we set α = 0.7, β = 0.7, and γ = 0.7, which produced the best rate in our experiments.

All the encoders were full connection layers with an input size of 8 and an output size of 2. Before the invariant encoder and specifies encoders, the size of the selected features from intra-modality fusion stage of each modality was processed to eight with different full connection layers and layer normalization.

III. Experiments

A. Dataset

Through the First Affiliated Hospital of Zhengzhou University, China, we collected and processed a total of 489 multimodal MRI images as our dataset. The MRI images used in this experiment included two types: IDH wild-type and mutated-type. This dataset consists of 312 IDH wild-type patients and 197 IDH mutant patients. IDH status was obtained by Sanger sequencing, while we did not consider the effect of age. Each patient had four MR image modalities (T1, T2, T1ce, and FLAIR). The data for this study came from multiple impact centers, but all were from the same hospital. In the actual process of acquiring data, the instruments, parameters, and environment of MRI images are significantly different due to specific circumstances. This also directly led to the fact that a part of the MRI data used in the study also had significant differences.

To solve this problem, we used the simple ITK tool [27, 28] and registered all modes in T2 mode using the rigid registration method. All IDH mutation labels were obtained from the clinical diagnosis results of the patients. According to the image we obtained, the doctor also marked the tumor area to facilitate us to select the ROI of the lesion area.

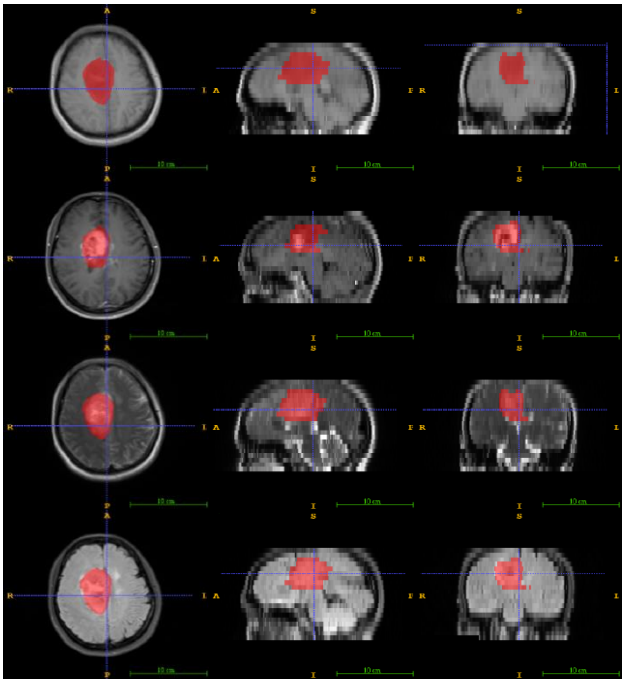

Figure 6. IDH mutation images Multi-modal MRI image (T1, T1CE, T2, Flair from top to bottom) in one case with Axial (Left), Sagittal (Middle), and Coronal (Right) planes with mask.

Extracting the tumor region by locating it as ROI and using this image for training is a widely used method. By focusing only on the tumor region, the network can better extract the features within the tumor region. Typically, we consider the tumor region as an ROI candidate and resize all ROIs to the same size. However, the aspect ratio of the outer rectangle in this method is large, which leads to the deformation of the image in the scaling process. To solve this problem, the length of the outer rectangle is used as the side length to extract the square ROI region. To solve this problem, the length of the outer rectangle is used as the side length to extract the square ROI region. The resolution of the MR images used in the deep learning stage was 72×72. Referring to other related work [7–9], we selected the ratio of the training set to the test set of 3:1. In our experiments, the data set were randomly divided using Python into training (366 cases) and test sets (122 cases). Brain MR images of the four modalities are shown in Fig. 6.

B. Experiments

To verify the impact of deep learning and radiomics features on the performance of prediction results, we conducted the following ablation studies. The first model (Model 1) was the CNN model. We used SA-Net with the same architecture as the deep feature extractor for training and verified the results. The second model (Model 2) was the radiomics model. We selected the same radiomics features and tested them using the same screening and classification manner. The third model (Model 3) was the proposed novel inter-modality fusion model, which fuses deep learning features with radiomics features but only use regression as an inter-modality fusion model. The last model (Model 4) is our proposed method based on the invariant- and specific-constraints fusion model as an inter-modality fusion model.

Then, the features were filtered and classified and fused to obtain the prediction results. We selected the area under the Curve (AUC), Accuracy (Acc), Precision (Pre), Recall (Rec), and F1 score (F1) as our evaluation measures.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 \times precision \times recall}{precison + recall}$$

Medical research has shown that gliomas patients with IDH mutation have a better survival rate. An investigation has shown that patients with IDH mutations have better prognostic performance and a significantly longer median survival for glioma (IDH wild type, 15 months; IDH mutant, 31 months) and anaplastic astrocytoma (IDH wild type: 20 months, IDH mutant, 65 months) [6]. Therefore, among the tasks of IDH mutation status prediction, unlike general tumor prediction tasks, we prefer to reduce mis discrimination of the negative case, that is, a higher precision is more favorable.

We used the Adam optimizer with a batch size of 16 and a learning rate of 0.0001, without data shuffling in our deep learning model. All the training was conducted on one Nvidia GeForce RTX 3090 24GB GPU. The proposed method achieves the best performance.

C. Experimental Results

The results of the ablation studies are presented in Table I. As shown in Table I, Models 1 and 2 show the results of using only deep learning features or radiomics features. Model 3 shows that better results are achieved when integrating the two features whereas Model 4 shows that the invariant- and specific-constraint fusion model performs better than regression fusion model in the IDH prediction task.

TABLE I. ABLATION EXPERIMENTS RESULTS

	CNN	Radiomics	Constraints fusion	AUC	Acc	Pre	Rec	F1
Model 1	√			0.77	0.70	0.69	0.70	0.69
Model 2		√		0.77	0.71	0.74	0.62	0.67
Model 3	√	√		0.82	0.77	0.77	0.77	0.77
Model 4 (Proposed method	√	√	√	0.81	0.79	0.80	0.75	0.77

We also compared the different inter-modality fusion methods. Table II shows the prediction results using only one a single modality MLP model. Obviously, T1CE and FLAIR performed better in the IDH status prediction task. As shown in Table III, the inter-modality fusion regression model (our preliminary work) was compared with the conventional average model (simply considering the average value of the four modality results as the final result), which uses a learnable weight for each modality in the inter-modality fusion. As shown in Table IV, the weight of each modality in the regression model shows the same result as the independent result, which means that T1CE, and T2-FLAIR are more important in this task. Because the IDH mutation status prediction performance of each modality is different, input methods will also have different effects on the results. The input methods are illustrated in Fig. 7. Therefore, when we compared several input methods of invariant- and specific-constraints inter-modality fusion models, we first tested the better modalities in terms of independent results which is T1CE and T2-FLAIR (Input Method 1). Secondly, we tested the results of the four modalities input in fusion Method 2. As shown in Table III, Input Method 1 shows improved but lower results in other evaluation measures, whereas Input Method 2 shows a more balanced result. In Input Method 2, the difficulty in the training stage was also significantly increased because the four modalities performed orthogonal operations on each other. Furthermore, this method increases the parameters of the model. Therefore, we aim to find an improved way to simplify the model and use useful information from the four modalities

simultaneously. We noted that in radiology, T1CE acts as an enhanced modality of T1 after contrast injection, while T2-FLAIR is an enhanced modality of T2. Therefore, we first concatenated T1 with T1CE and T2 with T2-FLAIR as our Input Method 3. The results of Input Method 3 showed that the accuracy, F1 score and precision outperformed the other models. Finally, we conducted a comparative experiment with the current state-of-the-art methods, proposed by Zhang *et al.* [8, 10], Yan *et al.* [9] and our preliminary work [17]. The results are presented in Table V. Our proposed method achieved better accuracy and precision than the state-of-the-art methods whereas the other indices did not decrease, which means that our model performs better in the IDH mutation status prediction task.

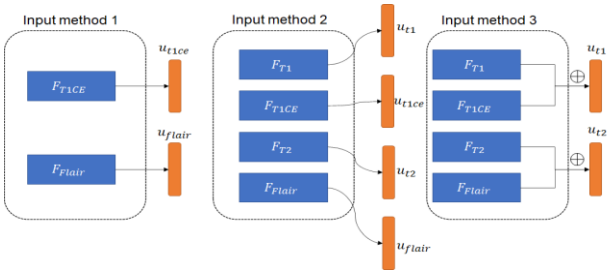

Figure 7. Input methods of the invariant- and specific-constraints inter-modality fusion model.

TABLE II. COMPARISON OF INDEPENDENT MODALITIES

	AUC	Acc	PRE	Rec	F1
T1	0.53	0.54	0.53	0.58	0.57
T1CE	0.76	0.74	0.73	0.73	0.73
T2	0.56	0.50	0.49	0.42	0.45
FLAIR	0.68	0.66	0.66	0.63	0.64

TABLE III. COMPARISON OF INTER-MODALITY FUSION METHODS

	AUC	Acc	Pre	Rec	F1
Average	0.79	0.74	0.71	0.78	0.75
Regression	0.82	0.77	0.77	0.77	0.77
Method 1	0.79	0.71	0.65	0.87	0.74
Method 2	0.77	0.76	0.78	0.70	0.74
Method 3	**0.81**	**0.79**	**0.80**	**0.75**	**0.77**

TABLE IV. WEIGHTS OF EACH MODALITY IN REGRESSION MODEL

T1	T1CE	T2	T2-FLAIR	Bias
0.40	0.49	0.30	0.47	−0.28

TABLE V. COMPARISON WITH STATE-OF-THE-ART METHODS

	AUC	Acc	Pre	Rec	F1
Zhang *et al.* [8]	0.77	0.72	0.72	0.70	0.71
Yan *et al.* [9]	0.77	0.71	0.74	0.62	0.67
Zhang *et al.* [10]	0.78	0.76	0.72	0.83	0.77
Shi *et al.* [17]	0.82	0.77	0.77	0.77	0.77
Proposed method	**0.81**	**0.79**	**0.80**	**0.75**	**0.77**

IV. DISCUSSION AND CONCLUSION

Based on the multi-modal fusion method, this paper proposes an IDH state prediction model combining deep learning and radiomics features, which has an intra-modal and inter-modal fusion structure. For the inter-modality fusion model, we proposed an invariant- and specific-constraint fusion method to fuse the MRI data of patients with glioma. Through the experimental results, we achieved an AUC of 0.81, accuracy of 0.79, precision of 0.80, recall of 0.75 and F1 score of 0.77, which achieved better performance than state-of-the-art methods. The most important measurement in IDH status mutation prediction and precision has also exhibited improved performance. In the field of computer diagnosis system, radiomics is a very important component of machine learning. Owing to dataset limitations, we should use more than just deep learning methods to process data in medical image processing. Especially in the glioma IDH status prediction task, when facing the medical image processing task, it is not enough to solve the problem using only the deep learning method. The multimodal fusion of medical data is also very important in computer diagnosis research. By proposing an invariant- and specific-constraints inter-modality fusion model with multiple loss- function-assisted learning, our model can learn the similar and complementary parts between different modalities. Therefore, we chose to effectively select the fused features by combining radiomics features and deep learning features and using statistical methods.

In the future, we intend to improve our approach to improve the performance of IDH status prediction by enhancing deep learning feature extraction methods and research on radiomics and statistics. Improving the efficiency of the multimodal fusion model is also a direction that we need to consider in the future.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Xaoyu Shi, Yinhao Li and Yen-Wei Chen conducted the research; Jingliang Cheng, Jie Bai and Guohua Zhao collected and processed the data. Xiaoyu Shi, Yinhao Li and Yen-Wei Chen wrote the paper; all authors had approved the final version.

FUNDING

This work was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20KK0234, No. 20K21821, No. 21H03470 and No. 21K17774.

REFERENCES

[1] Q. T. Ostrom, H. Gittleman, J. Xu *et al.*, "Primary brain and other central nervous system tumors diagnosed in the United States in 2009-2013," CBTRUS Statistical Report, Neuro Oncol, 2016.

[2] A. F. Tamimi and M. Juweid, "Epidemiology and outcome of glioblastoma," in *Glioblastoma [Internet]*, S. De Vleeschouwer, Eds. Brisbane (AU): Codon Publications; 2017, ch. 8, PMID: 29251870. doi: 10.15586/codon.glioblastoma

[3] D. N. Louis, A. Perry, G. Reifenberger *et al.*, "The 2016 World Health Organization classification of tumors of the central nervous system: a summary," *Acta Neuropathol*, vol. 131, no. 6, pp. 803–820, 2016.

[4] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat *et al.*, "The 2021 WHO classification of tumors of the central nervous system: A summary," *Neuro-Oncology*, vol. 23, issue 8, pp. 1231–1251, 2021,

[5] H. Yan, D. W. Parsons, G. Jin *et al.*, "IDH1 and IDH2 mutations in gliomas," *N Engl J Med*, 2009.

[6] S. Han, Y. Liu, S. J. Cai, M. Qian *et al.* "IDH mutation in glioma: Molecular mechanisms and potential therapeutic targets," *Br J Cancer*, 2020. doi: 10.1038/s41416-020-0814-x,
[7] Y. S. Choi *et al.*, "Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics," *Neuro-Oncology*, 2020.
[8] X. Zhang, I. Yutaro *et al.*, "IDH mutation status prediction by modality self attention network," in *Innovation in Medicine and Healthcare. Smart Innovation, Systems and Technologies*, Y. W. Chen, S. Tanaka, R. J. Howlett, and L. C. Jain, Eds. vol. 242, Springer, Singapore, 2021, pp. 51–57.
[9] J. Yan, B. Zhang, S. Zhang *et al.*, "Quantitative MRI-based radiomics for noninvasively predicting molecular subtypes and survival in glioma patients," *NPJ Precis*, 2021. doi: 10.1038/s41698- 021-00205-z
[10] X. Zhang, X. Shi, Y. Iwamoto, *et al.*, "IDH mutation status prediction by a radiomics associated modality attention network," *Visual Comput*., 2022, doi: 10.1007/s00371-022-02452-y
[11] Y.-W. Chen and L.C. Jain, *Deep Learning in Healthcare*, Springer, 2020.
[12] D. Liang, L. Lin, H. Hu, *et al.*, "Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, MICCAI 2018, Lecture Notes in Computer Science*, A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger, Eds. LNCS7951, Springer, 2018, pp. 666–675.
[13] T. Kitrungrotsakul, Q. Chen, H. Wu *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.
[14] H. Huang, H. Zheng, L. Lin *et al.*, "Medical image segmentation with deep atlas prior," *IEEE Trans. Medical Imaging*, vol. 40, no. 12, pp. 3519–3530, 2021.
[15] L. Peng, L. Lin, H. Hu *et al.*, "Classification 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.
[16] J. Hu, L. Shen, G. Sun *et al.*, "Squeeze-and-excitation networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[17] X. Shi, X. Zhang, Y. Iwamoto *et al.*, "An Intra- and Inter-Modality Fusion Model Using MR Images for Prediction of Glioma Isocitrate Dehydrogenase (IDH) Mutation," in *Proc. The IEEE Engineering in Medicine and Biology Society*, 2022.
[18] F. Burden, D. Winkler *et al.*, "Bayesian regularization of neural networks," in *Artificial Neural Networks. Methods in Molecular Biology™*, D. J. Livingstone, Eds. vol 458, Humana Press., 2018. doi: 10.1007/978-1-60327-101-1_3
[19] R. J. Gillies, E. K. Paul, H. Hedvig *et al.*, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
[20] H. Zhao, J. Jia, K. Vladlen *et al.*, "Exploring self-attention for image recognition," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.
[21] P. E. McKnight and J. Najab, "Mann-Whitney U test," *American Cancer Society, the Corsini Encyclopedia of Psychology*, 2010. doi: 10.1002/9780470479216.corpsy0524
[22] K. Wilhelm *et al.*, *Pearson's Correlation Coefficient*, Encyclopedia of Public Health, Springer Netherlands, January 2008, ch. 2569. doi:10.1007/978-1-4020-5614-7_
[23] R. Tibshirani *et al.* "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
[24] T. Zhou, R. Su, C. Stéphane *et al.*, "A review: Deep learning for medical image segmentation using multi-modal fusion," arXiv preprint, arXiv:2004.10664 [eess.IV], 2019. doi: 10.1016/j.array.2019.100004
[25] S. Pereira, A. Pinto, V. Alves, C. A. Silva *et al.*, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
[26] H. Devamanyu, Z. Roger, P. Soujanya *et al.* "MISA: Modality-invariant and -specific representations for multimodal sentiment analysis," arXiv preprint, arXiv:2005.03545 [cs.CL], 2020.
[27] B. C. Lowekamp, D. T. Chen, L. Ibáñez *et al.*, "The design of simpleITK," *Frontiers in Neuroinformatics*, vol. 7, no. 45, p. 45, 2013. doi: 10.3389/fninf.2013.00045
[28] Z. Yaniv, B. C. Lowekamp, H. J. Johnson *et al.*, "SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research," *J. Digit Imaging*, vol. 31, pp. 290–303, 2018. doi: 10.1007/s10278-017-0037-8