Deep wavelet scattering orthogonal fusion network for glioma IDH mutation status prediction

Qijian Chen, Lihui Wang, Zhiyang Xing, Li Wang, Xubin Hu, Rongpin Wang, Yue-Min Zhu

Deep Wavelet Scattering Orthogonal Fusion Network for Glioma IDH Mutation Status Prediction

Qijian CHEN[a], Lihui WANG[a,*], Zhiyang Xing[b], Li WANG[a], Xubin HU[a], Rongpin WANG[b] and Yue-Min ZHU[c]

[a]*Engineering Research Center of Text Computing & Cognitive Intelligence, Ministry of Education, Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China*

[b]*Department of Radiology, International Exemplary Cooperation Base of Precision Imaging for Diagnosis and Treatment, NHC Key Laboratory of Pulmonary Immune-related Diseases, Guizhou Provincial People's Hospital, Guiyang, 550002, China*

[c]*University Lyon, INSA Lyon, CNRS, Inserm, IRP Metislab CREATIS UMR5220, U1206, Lyon 69621, France*

ARTICLE INFO

Keywords:
Gliomas
IDH mutation status
Wavelet scattering
Multimodal information fusion
Orthogonal projection
Multi-center

ABSTRACT

Accurately predicting the isocitrate dehydrogenase (IDH) mutation status of gliomas is greatly significant for formulating appropriate treatment plans and evaluating the prognoses of gliomas. Although existing studies can accurately predict the IDH mutation status of gliomas based on multimodal magnetic resonance (MR) images and machine learning methods, most of these methods cannot fully explore multimodal information and effectively predict IDH status for datasets acquired from multiple centers. To address this issue, a novel wavelet scattering (WS)-based orthogonal fusion network (WSOFNet) was proposed in this work to predict the IDH mutation status of gliomas from multiple centers. First, transformation-invariant features were extracted from multimodal MR images with a WS network, and then the multimodal WS features were used instead of the original images as the inputs of WSOFNet and were fully fused through an adaptive multimodal feature fusion module (AMF2M) and an orthogonal projection module (OPM). Finally, the fused features were input into a fully connected classifier to predict IDH mutation status. In addition, to achieve improved prediction accuracy, four auxiliary losses were also used in the feature extraction modules. The comparison results showed that the prediction area under the curve (AUC) of WSOFNet on a single-center dataset was 0.9966 and that on a multicenter dataset was approximately 0.9655, which was at least 3.9% higher than that of state-of-the-art methods. Moreover, the ablation experimental results also proved that the adaptive multimodal feature fusion strategy based on orthogonal projection could effectively improve the prediction performance of the model, especially for an external validation dataset.

1. Inroduction

Isocitrate dehydrogenase (IDH) mutation status are significant prognostic markers in gliomas (Wesseling and Capper, 2018), and studies have shown that low-grade glioma (LGG) patients with IDH-mutant have better prognoses than those with IDH-wildtype (Sun et al., 2013; Wang et al., 2014). Therefore, accurately detecting IDH mutation status allows personalized treatment plans to be made and accordingly improves the prognoses of gliomas. Currently, IDH mutation status are mainly determined by performing immunohistochemistry or gene sequencing on tissue specimens (Bangalore Yogananda et al., 2020; Tietze et al., 2017). However, due to the particularity of the locations of gliomas, sampling a tissue specimen with biopsy or surgical resection is difficult and risky. In addition, considering the heterogeneity of gliomas, it is not guaranteed that the sampled tissue will contain sufficient IDH mutation information (Zhao et al., 2020).

Due to the non-invasive nature and the ability of imaging whole tumor, magnetic resonance imaging (MRI) has been considered the most promising candidate to replace the biopsy for determining IDH mutation status accurately prior to surgery (Kim et al., 2020). Numerous studies have shown that MRI-based radiomics can effectively predict molecular glioma subtypes (Yu et al., 2017; Han et al., 2018; Lu et al., 2018; Zhou et al., 2017).However, the prediction performance of radiomics is severely dependent on complex processing pipelines, such as data processing, glioma segmentation, handcrafted feature extraction, and feature selection. Such dependencies make it difficult to further improve the prediction accuracy of this method, particularly for multicenter datasets. Recently, with the successful applications of convolutional neural networks (CNNs) in image classification tasks, the use of CNNs to predict IDH mutation status has attracted the attention of many researchers. For instance, Li et al. (Li et al., 2017) developed a deep learning-based radiomics method for predicting the IDH mutation status of gliomas with T2-weighted (T2w) and FLAIR images, in which the handcrafted features of traditional radiomics were replaced with the semantic features learned from a CNN, and the highest area under the curve (AUC) reached 92%. Liang et al. (Liang et al., 2018) proposed a 3D DenseNet model to predict IDH status with T2w images and obtained an AUC of 81.6%. Choi et al. (Choi et al., 2019) designed an explainable recurrent neural network (RNN) to predict IDH status with perfusion MR images, achieving an accuracy of 92.8% on their validation set. Nalawade et al. (Nalawade et al., 2019) presented an automated pipeline for noninvasively predicting IDH status using deep learning and T2w images, and this approach achieved approximately 83.8%

*Corresponding author: e-mail: lhwang2@gzu.edu.cn;

accuracy on the testing set. Ai et al.(Ai et al., 2022) proposed a three-directional attention block network to predict IDH status with FLAIR images and achieved an AUC of 96.44%.

Considering that multi-modal MRI can provide tumor information from different perspectives, combining multi-modal information is therefore beneficial for promoting further the prediction accuracy. Li et al (Li et al., 2017) and Liang et al (Liang et al., 2018) accordingly extended their models to multimodal images and improved their AUCs to 95% and 85.7%, respectively. Subsequently, Chang et al. (Chang et al., 2018) used FLAIR, T2w, T1 pre-contrast and T1 post-contrast images to predict IDH status in gliomas based on a residual CNN, resulting in an AUC of 0.93 on the validation set. Van der Voort et al. (van der Voort et al., 2022) developed a multitask CNN that uses multimodal structural MR images to segment gliomas and simultaneously predict glioma grades, IDH mutations, and 1p/19q codeletion status. The AUC of IDH status prediction reached 90% on the independent testing set. Cheng et al.(Cheng et al., 2022) proposed a semi-supervised multitask learning framework for both glioma segmentation and IDH genotyping, and with the help of pseudo labels, the IDH status prediction process achieved an AUC of 90.37%. However, most existing multimodal image-based IDH prediction methods do not fully and effectively explore multimodal information; they simply combine multimodal images or semantic multimodal features through concatenation or addition. The reasonable fusion of multimodal information with adaptive weights determines the prediction performance of the employed model. In addition, most IDH prediction studies focus on single-center datasets, and the prediction performance achieved on multicenter datasets is not satisfactory. The difference between the image contrast levels and intensity distributions of datasets acquired from multiple centers significantly influences the generalization ability of deep learning models. To cope with this issue, several histogram specification methods have been used (Zhao et al., 2019; Chen et al., 2019), although these methods can force a multicenter dataset to have almost the same image intensity distribution, the histogram specification process easily introduces extra noise, which influences the resulting prediction performance.

In this work, to address the above-mentioned problems of the existing IDH status prediction models, we proposed a deep wavelet scattering orthogonal fusion network (WSOFNet), in which the transformation invariant wavelet scattering features for multicenter dataset were used as the input of the network to deal with the influences of image intensity distribution, and a two-stage fusion strategy based adaptive attention and orthogonal projections were designed to effectively fuse the multimodal feature information.

2. Materials And Methods

2.1. Data Description

Two kinds of datasets were used in this work. One dataset was downloaded from the MICCAI Challenge 2020 website for glioma segmentation (Menze et al., 2014; Bakas et al., 2017; Clark et al., 2013), including T1w, T1-Gd-enhanced, T2w, and FLAIR images. The corresponding clinical information of some patients could be found in the cancer imaging archive (TCIA) by matching the patient IDs and names. In total, 194 patients with effectively labeled IDH status (96 positive cases (IDH-Mutant) and 108 negative cases (IDH-WT)) were included in this dataset. To overcome the influence of patient motions, skull stripping was first performed, followed by image registration to ensure that the multimodal images were strictly aligned for the same patient. After registration, the spatial resolution was 1 mm × 1 mm × 1 mm. The regions of interest (ROIs) were effectively labeled by experienced radiologists, including edema, non-enhancing solid core, necrotic core, and enhancing core. In addition, the images of each patient were standardized using Z score normalization. For this dataset, 145 cases were randomly selected as the training set, and the remaining 49 cases were selected as the testing set.

Table 1
Clinical features of the patients

	MICCAI challenge dataset		GZPH dataset
	Train	Test	Valid
Patients			
WT	81(56%)	27(55%)	40(52%)
Mutant	64(44%)	22(45%)	37(48%)
Slices			
WT	3181(52%)	1093(49%)	1396(53%)
Mutant	2948(48%)	1140(51%)	1243(47%)
Gender			
Female	75(51%)	20(29%)	35(45%)
Male	70(49%)	28(67%)	42(55%)
NA	0	1(4%)	0
Age			
<=40	38(26%)	11(23%)	40(52%)
<=50	27(19%)	7(14%)	15(19%)
>50	80(55%)	31(63%)	22(29%)

Another dataset was acquired from Guizhou Provincial People's Hospital (GZPH), including T1-Gd-enhanced and T2w images of 77 patients. To complete the multimodal image information, we used the method presented in (Zhou et al., 2020) to generate T1w and FLAIR images. Subsequently, all the preprocessing steps mentioned above were also implemented for these data. To test the generalization ability of our proposed method, this dataset was used as the external validation set. The detailed clinical information of the datasets used in this work is given in Table 1.

In this work, to alleviate the error induced by tumor delineation and save time and labor, a minimum rectangular box completely bounding the largest tumor area (including the tumor core and necrotic area) was taken as the ROI. Furthermore, to fully consider the local spatial information around the tumor, the adjacent slices were also considered in the prediction process and used to form a 3D ROI, as illustrated in Fig. 1.

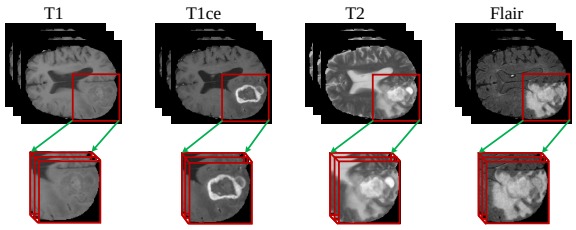

Figure 1: Illustrations of tumor ROIs and the corresponding zoomed-in versions. The red boxes indicate the bounding boxes containing the largest tumor region of the slice under different modalities.

2.2. Wavelet Scattering

Considering the advantages of invariant features (Bruna and Mallat, 2013; Mallat, 2012) in image classification, this work uses wavelet scattering (WS) features instead of the original images as the input of the network. Letting the input image be x, its WS features at different levels S can be expressed by

$$S = \begin{bmatrix} S_0 = U_0 * \phi_J \\ S_1 = U_1 * \phi_J \\ \vdots \\ S_m = U_m * \phi_J \end{bmatrix} \tag{1}$$

where $\phi_J(u) = 2^{-2J}\phi(2^{-J}u)$ represents the scaling function of wavelet transform at the maximum scale J. In this work, $\phi(u)$ is a Gaussian function expressed as $\phi(u) = e^{-u^2/2\sigma^2}$, "$*$" indicates the convolution operation, U_m represents the scattering propagator for the m^{th} order wavelet scattering, formulated as

$$U = \begin{bmatrix} U_0 = x \\ U_1 = \left|U_0 * \psi_{j_0,r}\right| \\ \vdots \\ U_m = \left|U_{m-1} * \psi_{j_{m-1},r}\right| \end{bmatrix} \tag{2}$$

where $j = 0, 2, ..., J$ and $r = 1, 2, ..., L$, $\psi_{j_m,r}(u) = 2^{-2j}\psi\left(2^{-j}r^{-1}u\right)$ represents the directional wavelet function along the direction r at the scale j used in the m^{th} level WS network, J is the maximum wavelet decomposition scale and L is the maximum direction number. The scattering order m is determined by decomposition times of high-frequency component.

The second-order ($M = 2$) WS decomposition for an image block is shown in Fig. 2. The blue, red, and green solid lines represent the 0th, 1st and 2ed order WS propagation paths, respectively, resulting in the scattering propagators U_1 and U_2. Based on the (2), if $M = 2$ and $L = 4$, we will get 1 U_0 propagator, 8 U_1 propagators and 16 U_2 propagators. Using the (1), we can get accordingly a total of 25 scattering feature maps, in which the numbers of S_0, S_1 and S_2 are 1, 8 and 16, respectively. Generally, if the original image size is $H \times W$, for any given L and J, the second-order WS decomposition will result in $(1 + LJ + L^2J(J-1)/2)$ scattering feature maps with a size of $\frac{H}{2^J} \times \frac{W}{2^J}$.

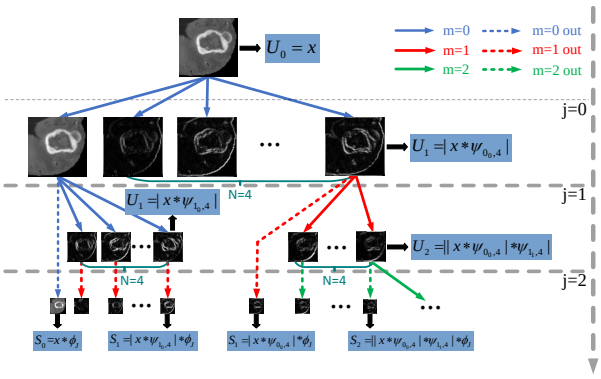

Figure 2: Schematic diagram of the second-order WS network when the wavelet decomposition scale is set as $J = 2$,with U_i representing the wavelet scattering propagator and S_i the wavelet scattering coefficient ($i = 0, 1, 2$), the maximum scattering order is $M = 2$, and the maximum scattering direction is $L = 4$.

2.3. IDH Mutation Status Prediction based on Deep WSOFNet

After extracting the WS features of multimodal MR image blocks, we designed a deep WSOFNet to predict IHD mutation status in gliomas. The overall framework of WSOFNet is shown in Fig. 3. First, the WS feature maps of multimodal images (T1w, T2w, T1-CE, and FLAIR images) were input respectively to their own pre-feature extraction modules (PEM) to extract the modality-dependent semantic features. As shown in Figure 3, the PEMs were mainly composed of convolutional, batch normalization, and GeLU activation layers. Subsequently, these features were fused with an adaptive multimodal feature fusion module (AMF2M). Considering that the fused features may lose some beneficial information for prediction, to fully exploit the useful multimodal information, the features lost by the fusion process were further recovered by an orthogonal projection module (OPM). The recovered features of each modality were further fused with the AMF2M. Next, the two fused feature maps were further mined with the feature extraction module (FEM), which combines the advantages of a transformer and a CNN to more comprehensively characterize multimodal fusion features. The detailed structure of the FEM is demonstrated at the bottom of Fig. 3. The MBConv module in the FEM has the same structure as that used in CoAtNet (Dai et al., 2021), and the parameter settings are also the same. Finally, the outputs of two FEMs were concatenated and fed into a FC layer to predict IDH mutation status. To achieve improved prediction accuracy, in addition to the cross-entropy loss used in the last classifier layer, four auxiliary cross-entropy losses were also used in the first four PEMs to force them to extract the useful single-modality image features that were highly related to IDH mutation status, as illustrated in Fig. 3. Since the AMF2M and OPM are the kernels of the proposed method, the detailed structures of the AMF2M and the principles of the OPM will be specified in the following subsections.

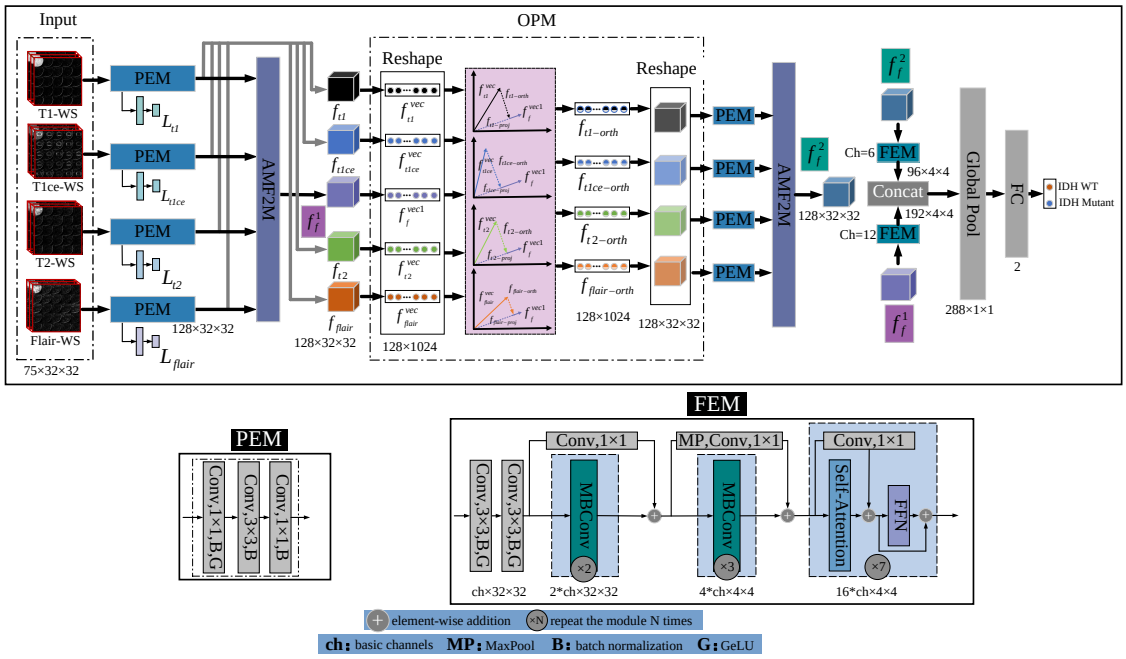

Figure 3: The overall architecture of the proposed WSOFNet. The wavelet scattering maps of multimodal images pass through the PEM blocks to extract modality-dependent semantic features, these features are then sequentially fused using AMF2M, OPM and AMF2M blocks. The features outputted by the first AFM2M block are noted as f_f^1, and those outputted by the second AFM2M block are noted as f_f^2. f_f^1 and f_f^2 are further mined with FEM block and finally concatenated to predict the IDH status. The detailed structure of PEM and FEM are demonstrated at the bottom of the Figure 3, while the structure of AMF2M is given in Figure 4.

2.3.1. Adaptive Multi-modal Feature Fusion Module

Multimodal images can provide complementary information, and how to fuse this information determines the achieved IDH mutation status prediction accuracy. Different from the method of directly splicing multimodal image features, the AMF2M used in this work fuses multimodal features by adaptively weighting the spatial attention and channel attention, and the detailed structure of this module is shown in Fig. 4.

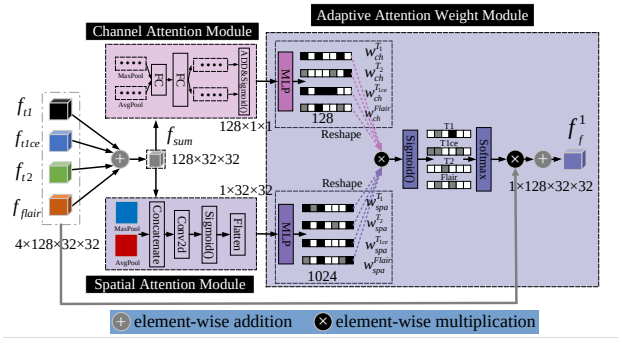

Figure 4: The detailed structure of adaptive multimodal feature fusion module (AMF2M), where the multimodal features are adaptively fused using the multiplications of channel attentions and spatial attentions.

The WS features of each modality were input into PEMs, and the corresponding outputs were noted as f_{t1}, f_{t1ce}, f_{t2} and f_{flair} respectively. These outputs were then summed in an elementwise manner and used to form the feature map f_{sum},

$$f_{sum} = f_{t1} \oplus f_{t1ce} \oplus f_{t2} \oplus f_{flair} \tag{3}$$

f_{sum} was then input into channel attention module (CAM) and spatial attention module (SAM) respectively. The CAM was used to calculate the channel attention w_{ch} based on maximum pooling and average pooling operations performed on f_{sum}, as illustrated in Fig. 4, w_{ch} can be formulated as,

$$w_{ch} = \sigma(MLP(MaxPool(f_{sum})) + MLP(AvgPool(f_{sum}))) \tag{4}$$

where MLP represents the two shared fully connection (FC) layers and σ indicates the sigmoid activation function. The SAM was responsible for calculating the spatial attention w_{spa}, which was derived from the maximum and average pooling operations performed on f_{sum} along the channel dimension. Specifically, denoting the channel-wise maximum and average pooling results as $M_{f_{sum}}$ and $A_{f_{sum}}$ respectively, $M_{f_{sum}}$ and $A_{f_{sum}}$ were concatenated and then passed through a 2D convolutional layer and a sigmoid activation layer. The output was finally flattened to yield w_{spa},

$$w_{spa} = flatten(\sigma(\underbrace{Cov2d_{5\times5}}_{ch=(2,1)}(Concat(M_{f_{sum}}, A_{f_{sum}}))) \quad (5)$$

To adaptively weight the features of different modalities, the channel attention w_{ch} and spatial attention w_{spa} were fed into different FC layers to generate the channel and spatial weights of different image modalities, which are denoted as $w_{ch}^{T_1}$, $w_{ch}^{T_2}$, $w_{ch}^{T_{1ce}}$, w_{ch}^{Flair} and $w_{spa}^{T_1}$, $w_{spa}^{T_2}$, $w_{spa}^{T_{1ce}}$, w_{spa}^{Flair}, respectively. Subsequently, element-wise multiplications were performed for the weights of the same image modality, and this was followed by sigmoid activation to generate the adaptive weight for each modality:

$$\begin{bmatrix} w_{adp}^{T_1} \\ w_{adp}^{T_2} \\ w_{adp}^{T_{1ce}} \\ w_{adp}^{Flair} \end{bmatrix} = \sigma \begin{bmatrix} w_{ch}^{T_1} \otimes w_{spa}^{T_1} \\ w_{ch}^{T_2} \otimes w_{spa}^{T_2} \\ w_{ch}^{T_{1ce}} \otimes w_{spa}^{T_{1ce}} \\ w_{ch}^{Flair} \otimes w_{spa}^{Flair} \end{bmatrix} \quad (6)$$

These weights were finally normalized with a SoftMax activation to guarantee that the weight sum of different modalities is 1 at a given position. The final fused features of AMF2M can be obtained by,

$$f_f^1 = f_{t1} \otimes w_{adpn}^{T_1} \oplus f_{t2} \otimes w_{adpn}^{T_2} \oplus f_{t1ce} \otimes w_{adpn}^{T_{1ce}} \oplus f_{flair} \otimes w_{adpn}^{flair} \quad (7)$$

where $w_{adpn}^{T_1}$, $w_{adpn}^{T_{1ce}}$, $w_{adpn}^{T_2}$, w_{adpn}^{Flair} represent the normalized adaptive weights of different modalities.

2.3.2. Principles of Orthogonal Projection Module

The fused features f_f^1 output by the AMF2M may loss some important information that related with the IDH mutation status. To recover this missed information, we proposed to extract the features that orthogonal to f_f^1 form the original features f_{t1}, f_{t1ce}, f_{t2} and f_{flair} respectively with OPM, as illustrated in Fig. 3. The single modality features and the fused features were flattened firstly as one-dimensional vector, formulated as,

$$\begin{bmatrix} f_{t1}^{vec} \\ f_{t2}^{vec} \\ f_{t1ce}^{vec} \\ f_{flair}^{vec} \\ f_f^{vec1} \end{bmatrix} = flatten(\begin{bmatrix} f_{t1} \\ f_{t2} \\ f_{t1ce} \\ f_{flair} \\ f_f^1 \end{bmatrix}) \quad (8)$$

Then we projected the feature vectors of each modality onto the fused feature vector through

$$f_{model-proj} = \frac{f_{model}^{vec} \cdot f_f^{vec1}}{sum(f_f^{vec1} \otimes f_f^{vec1})} f_f^{vec1} \quad (9)$$

The orthogonal features were accordingly derived by subtracting the projected features from the original features (where $model \in [t1, t2, t1ce, flair]$)

$$f_{model-orth} = f_{model}^{vec} - f_{model-proj}. \quad (10)$$

Since the resulted orthogonal features were not contained in the fused feature f_f^1, therefore, it can be guaranteed that these orthogonal features can compensate the information lost in the adaptive fusion process. However, not all the lost information can continue to improve the prediction ability of the model. Therefore, these orthogonal feature vectors were reshaped as the feature maps and then passed through PEM and AMF2M to obtain the second fused features f_f^2.

2.3.3. Evaluation metrics

We used accuracy (ACC), sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), and area under the curve (AUC) to evaluate the predictive performance of different models for IDH mutation status. AUC needs to be calculated based on the receiver operating characteristic curve (ROC) and is a commonly used metric for classification models. It mainly measures whether the model can effectively distinguish positive and negative examples under different thresholds. A larger AUC indicates that the model is less affected by the threshold, which reflects that the model is more stable under the condition of high precision. These evaluation metrics can be calculated with, ACC = (TP+TN)/(TP+FN+FP+TN), specificity = TN/(FP+TN), sensitivity = TP/(TP+FN),NPV = TN/(FN+TN), PPV = TP/(TP+FP). where TP (true positive), FN (false negative), FP (false positive) TN (true negative) can be derived from the confusion matrix by counting the number of samples according to the predicted and ground-truth labels. The higher the above six quantitative evaluation metrics, the better the prediction performance of the model.

3. Experiments and Results

3.1. Experimental Setup

All experiments in this work used the same training, test, and validation datasets, and the sizes of the image patches that were input into the network were adjusted to 128×128. To prevent overfitting, data augmentation strategies including random rotation, translation, shearing, scaling, flipping, and Gaussian noise were incorporated into the training process, and the sharpness-aware minimization-based optimization strategy (Foret et al., 2020) was also used. In WOSFNet, the FC layers were initialized with a normal distribution possessing a variance of 0.001, and the convolutional layers were initialized with a Kaiming normal distribution. The stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and an initial learning rate of 1e-4 was used. The learning rate was gradually decreased with the ReduceLROnPlateau strategy; that is, if

an evaluation metric of the network did not improve for a patient number of epochs, the learning rate of the network was adaptively reduced. In this work, the Patience = 8, BatchSize = 64, and Epoch = 150 for all experiments.

3.2. Comparison of different IDH mutation status prediction models on test dataset

To verify the glioma IDH mutation status prediction performance of the proposed model, we compared it with the mainstream deep learning models, including VGG (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), InceptionV4 (Szegedy et al., 2017), DenseNet (Huang et al., 2017), ShuffleNet (Zhang et al., 2018), MobileNetV3 (Howard et al., 2019), Res2Net (Gao et al., 2019), EfficientNet (Tan and Le, 2019), MLP-Mixer (Tolstikhin et al., 2021), RepVGG (Ding et al., 2021), Swin-Transformer (Liu et al., 2021), and ViTAEv2 (Zhang et al., 2022), as well as some state-of-the-art glioma prediction methods (Naser and Deen, 2020; Ge et al., 2018; Lu et al., 2020; Li et al., 2022; Jiang et al., 2022; Wu et al., 2023).

Fig. 5 shows the ROC curves of different models for predicting IDH mutation status on the test set (originate from the same center as training set) and Table 2 shows the corresponding quantitative metrics. It can be found that the AUCs of all models are larger than 0.90, and our proposed model WSOFNet achieves the highest ACC (0.9796), AUC (0.9966), sensitivity (1.0) and NPV (1.0), which are improved respectively by 2.1%, 1.0%, 4.8% and 3.8% at least comparing against the best existing models. Although the specificity and PPV of ResNet, InceptionV4 and the method of Naser et al are comparable to those of our model, their sensitivity and NPV are a little lower than ours. As to the other methods, they have the worst specificity and PPV. In this work, IDH-mutant is considered as the positive example while IDH-WT is taken as the negative, low PPV and specificity indicate that these models prefer to classify the IDH-WT as IDH-mutant. However, our WSOFNet can recognize the IDH-mutant glioma precisely, simultaneously, fewer IDH-WT gliomas are wrongly classified as IDH-mutant.

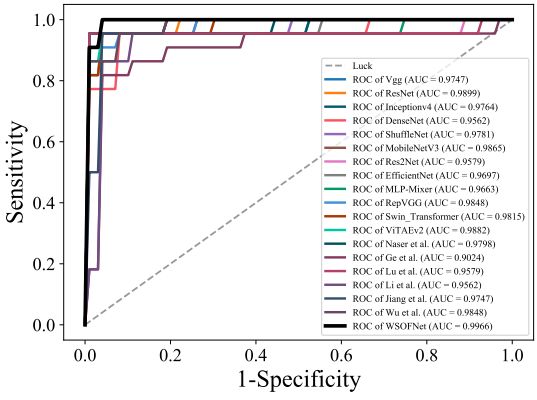

Figure 5: ROCs of different models for predicting IDH mutation status on the single-center test set.

To further compare the performance of different models more intuitively, box plots of the predicted probabilities corresponding to the ground-truth class of each patient in the testing set are given in Fig. 6. For a given IDH status label, if the prediction is correct, the corresponding prediction probability should be larger than 0.5. We found that in VGG, InceptionV4, ShuffleNet, MobileNetV3, Res2Net, RepVGG, VitAEv2 and the method of Ge et al., some patients were not correctly classified (outliers with probabilities much smaller than 0.5 were produced). Among the remaining models, although there were no outliers, the minimum prediction probability values were lower than 0.5, especially for ResNet, DenseNet, EfficientNet, MLP-Mixer, VGG and the work of Naser et al. The minimum value of our proposed WSOFNet was also smaller than 0.5 but not as small as the others. In addition, among the models without outliers, WSOFNet had the smallest interquartile range, and such a stable prediction range indicates that the generalization performance of WSOFNet was better than that of the other models.

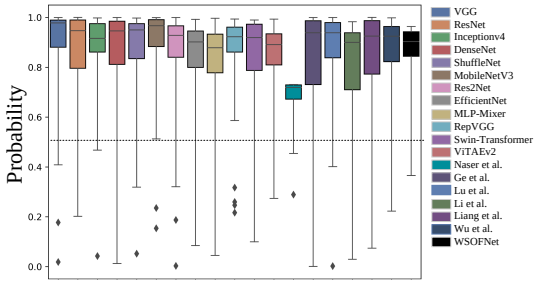

Figure 6: Box plots of the predicted probabilities of different methods for all the samples in single-center test set. If the ground-truth label of a sample is 1, the predicted probability equals to the model output, otherwise, using 1 minus the model output as the predicted probability.

3.3. Comparison of different IDH mutation status prediction models on external validation dataset

To verify the generalization ability of the different models for predicting IDH mutation status on multi-center data, we quantitatively compared different methods on an external validation set.

As shown in Fig. 7 and Table 3, among all the comparison methods, the method proposed by Wu et al. achieved the highest ACU of 0.9291 when tested on a dataset obtained from a different site with varying imaging parameters. Nevertheless, this was still somewhat inferior when compared to our proposed WSOFNet, which managed an AUC of 0.9655, representing an increase of around 3.9% over the best comparison model. From Table 3, we notice that the sensitivity and NPV of WSOFNet were slightly smaller than those of ResNet, ShuffleNet, Res2Net, MLP-Mixer, and ViTAEv2, whose sensitivity and NPV metrics were equal to 1. However, the corresponding specificity and PPV of these models are too low, for instance, the specificity of ResNet,

Table 2
Quantitative evaluation metrics of different models for predicting IDH mutation status in gliomas on single-center test set

Model	ACC	Sensitivity	Specificity	PPV	NPV	AUC	AUC 95%CI
VGG	0.9388	0.9545	0.9259	0.9130	0.9615	0.9749	[0.96-0.98]
ResNet	0.9592	0.9545	**0.9630**	**0.9545**	0.9630	0.9899	[0.98-0.99]
Inceptionv4	0.9592	0.9545	**0.9630**	**0.9545**	0.9630	0.9764	[0.97-0.98]
DenseNet	0.8776	0.9545	0.8148	0.8077	0.9565	0.9562	[0.95-0.97]
ShuffleNet	0.9184	0.9545	0.8889	0.8750	0.9600	0.9781	[0.98-0.99]
MobileNetV3	0.9592	0.9545	0.9630	**0.9545**	0.9630	0.9865	[0.98-0.99]
Res2Net	0.9184	0.9545	0.8889	0.8750	0.9600	0.9579	[0.93-0.96]
EfficientNet	0.9184	0.9545	0.8889	0.8750	0.9600	0.9697	[0.96-0.98]
MLP-Mixer	0.9592	0.9545	**0.9630**	0.9545	0.9630	0.9663	[0.96-0.98]
RepVGG	0.9184	0.9545	0.8889	0.8750	0.9600	0.9848	[0.97-0.99]
Swin-Transformer	0.9184	0.9545	0.8889	0.8750	0.9600	0.9815	[0.97-0.99]
ViTAEv2	0.9184	0.9545	0.8889	0.8750	0.9600	0.9882	[0.97-0.99]
Naser et al. (Naser and Deen, 2020)	0.9592	0.9545	**0.9630**	**0.9545**	0.9630	0.9798	[0.97-0.98]
Ge et al. (Ge et al., 2018)	0.8776	0.8636	0.8889	0.8636	0.8889	0.9024	[0.89-0.93]
Lu et al. (Lu et al., 2020)	0.9388	0.9545	0.9259	0.9130	0.9615	0.9579	[0.94-0.97]
Li et al. (Li et al., 2022)	0.8980	0.9091	0.8889	0.8696	0.9231	0.9562	[0.94-0.96]
Jiang et al. (Jiang et al., 2022)	0.9388	0.9545	0.9259	0.9130	0.9615	0.9747	[0.97-0.98]
Wu et al. (Wu et al., 2023)	0.9388	0.9545	0.9259	0.9130	0.9615	0.9848	[0.98-0.99]
WSOFNet	**0.9796**	**1**	**0.9630**	**0.9545**	**1**	**0.9966**	**[0.99-0.99]**

Table 3
Quantitative evaluation metrics of different models for predicting IDH mutation status of glioma on external validation dataset

Model	ACC	Sensitivity	Specificity	PPV	NPV	AUC	AUC 95%CI
VGG	0.8442	0.8649	0.8250	0.8205	0.8684	0.9128	[0.91-0.93]
ResNet	0.5974	**1**	0.2250	0.5441	**1**	0.8142	[0.79-0.83]
Inceptionv4	0.7662	0.8649	0.6750	0.7111	0.8438	0.9061	[0.89-0.92]
DenseNet	0.6753	0.8649	0.5000	0.6154	0.8000	0.7824	[0.76-0.80]
ShuffleNet	0.6623	**1**	0.3500	0.5873	**1**	0.8689	[0.86-0.89]
MobileNetV3	0.6494	0.9459	0.3750	0.5833	0.8824	0.8696	[0.85-0.88]
Res2Net	0.6364	**1**	0.3000	0.5692	**1**	0.8534	[0.83-0.86]
EfficientNet	0.7532	0.9189	0.6000	0.6800	0.8889	0.8270	[0.87-0.89]
MLP-Mixer	0.7662	**1**	0.5500	0.6727	**1**	0.8797	[0.83-0.86]
RepVGG	0.6883	0.9730	0.4250	0.6102	0.9444	0.8581	[0.83-0.86]
Swin-Transformer	0.7792	0.8919	0.6750	0.7174	0.8710	0.8541	[0.83-0.86]
ViTAEv2	0.7403	**1**	0.5000	0.6491	**1**	0.9176	[0.91-0.93]
Naser et al. (Naser and Deen, 2020)	0.8312	0.9730	0.7000	0.7500	0.9655	0.9128	[0.90-0.92]
Ge et al. (Ge et al., 2018)	0.6623	0.7027	0.6250	0.6341	0.6944	0.7054	[0.68-0.71]
Lu et al. (Lu et al., 2020)	0.7403	0.9730	0.5250	0.6545	0.9545	0.8682	[0.85-0.88]
Li et al. (Li et al., 2022)	0.6104	1.0000	0.2500	0.5522	1.0000	0.9182	[0.91-0.93]
Jiang et al. (Jiang et al., 2022)	0.7662	0.9189	0.6250	0.6939	0.8929	0.8838	[0.87-0.89]
Wu et al. (Wu et al., 2023)	0.8312	0.8919	0.7750	0.7857	0.8857	0.9291	[0.92-0.94]
WSOFNet	**0.9610**	0.9730	**0.9500**	**0.9474**	0.9744	**0.9655**	**[0.95-0.97]**

ShuffleNet, Res2Net, MLP-Mixer and ViTAEv2 are 0.225, 0.35, 0.3, 0.55, and 0.5 respectively, which means that in these models, most of the IDH-WT gliomas are classified as the IDH-mutant gliomas. This result is consistent with the finding from the test dataset, but the performance of these models on the external dataset was even worse. In contrast, both specificity and sensitivity of WSOFNet are higher than 0.95, which means that our model can accurately recognize IDH-WT and IDH-mutant gliomas, even on the external validation dataset.

Similar to the testing set, we also present the box plots of the prediction probabilities produced for the external validation set in Fig. 8. We see that although DenseNet, ShuffleNet, MobileNetV3, Res2Net, RepVGG, and the method of Naser et al. did not have extreme outliers, their interquartile zones passed through the probability threshold line (0.5), which means that most examples were incorrectly classified.

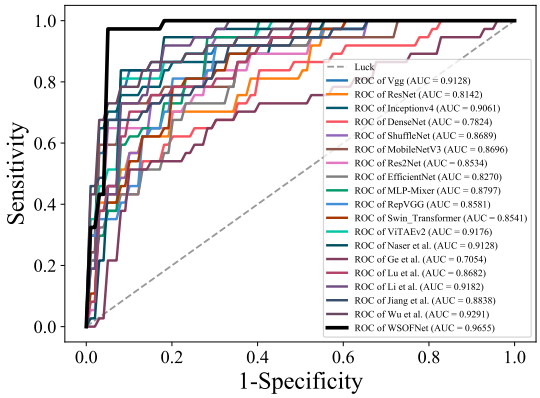

Figure 7: ROCs of different models for predicting IDH mutation status on external validation set.

Among the remaining models, our WSOFNet had the narrowest range (between its minimum and maximum values), with all the probabilities being larger than 0.5 except for the outliers. Even though the minimum value of VGG was also above 0.5, it produced many more outliers than our method. The distributions of the prediction probabilities on the external validation dataset further validate the robustness and generalization ability of our proposed model.

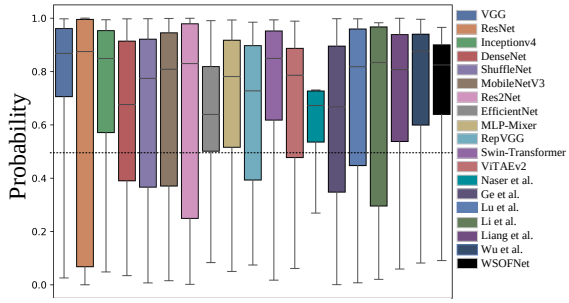

Figure 8: Box plots of the predicted probabilities of different methods for all the samples.

3.4. Visualizing the learned representations

To further demonstrate the effectiveness of the features extracted by WSOFNet for separating IDH-WT and IDH-mutant gliomas on both the test and validation sets, we performed unsupervised clustering of the features extracted from the FC layers of different models using t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton, 2008). As shown in Fig. 9, when testing on the single-center dataset, the features extracted with almost all the models can separate well the IDH-WT and IDH-mutant gliomas, and WSOFNet performs the best. However, when testing on dataset acquired at the center different from train set, the differentiating ability of other models for IDH-WT and IDH mutant gliomas decreases dramatically (Fig. 10), especially for ResNet, ShuffleNet, MobileNetV3, and RepVGG, they cannot distinguish these two gliomas. Comparing with other models, the proposed WSOFNet produce the best separation of IDH-WT and IDH-mutant features, with the fewest samples being wrongly clustered, which means that WSOFNet can learn more discriminative features.

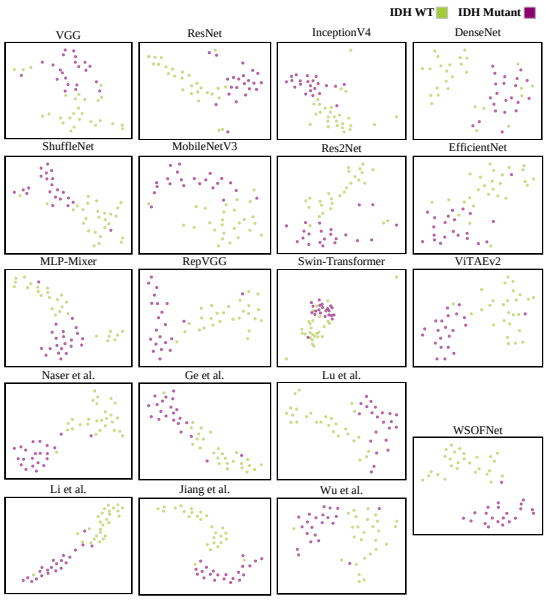

Figure 9: t-SNE visualization of features extracted with different models on single-center test set. Purple indicates the features of gliomas with IDH-mutant, and green means the features of gliomas with IDH-WT. The more separated the two kinds of features are, the better the prediction performance of the corresponding model is.

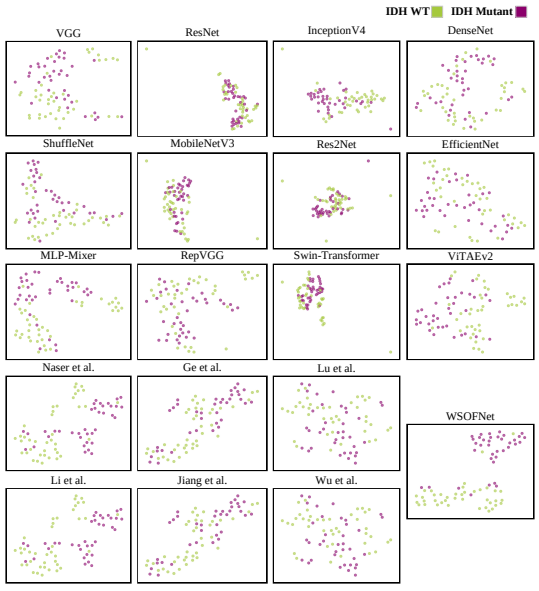

Figure 10: t-SNE visualization of features extracted with different models on external validation set.

Table 4
Ablation experiment setting and corresponding prediction results.

	WS	Multi-losses	AMF2M	OPM	Test		Valid	
					ACC	AUC	ACC	AUC
WSOFNet-v1					0.9184	0.9916	0.7143	0.8108
WSOFNet-v2	√				0.9388	0.9714	0.8052	0.8905
WSOFNet-v3	√	√			0.9388	0.9815	0.8312	0.8912
WSOFNet-v4	√	√	√		**0.9796**	0.9949	0.8571	0.9372
Proposed	√	√	√	√	**0.9796**	**0.9966**	**0.9610**	**0.9655**

3.5. Ablation experimental results

To verify the effectiveness of the different modules and strategies proposed in this work, several ablation experiments were implemented, as detailed in Table 4, in which WSOFNet-v1 represents the baseline model that used the original image as input for prediction; WSOFNet-v2 also used the baseline model but the wavelet scattering features as input; WSOFNet-v3 had the same structure as WSOFNet-V2 but was trained with multi-losses; in WSOFNet-v4, the AMF2M block was embedded based on WSOFNet-V3; the proposed model was constructed by introducing OPM strategy into the WSOFNet-V4. Comparing WSOFNet-v1 and WSOFNet-v2, it can be found that using WS features instead of the original image as input, the ACC on the test set and external validation set were increased by 2.2% and 12.7 % respectively. Although the AUC of WSOFNet-v2 was slightly lower than that of WSOFNet-v1 on single-center test set, it was much higher than that of WSOFNet-V1 (about 10.4%) on the external validation set. As to the effectiveness of the auxiliary losses (comparing WSOFNet-v3 and WSOFNet-v2), they can also improve the prediction performance on multi-center datasets. From the comparison between WSOFNet-v3 and WSOFNet-v4, we notice that with the adaptive multi-modal feature fusion module (AMF2M), both ACC and AUC can be further significantly improved, which validates the effectiveness of the AMF2M. With the help of OPM (comparing WSOFNet-v4 and proposed WSOFNet), we interestingly find that the improvement in ACC and AUC on the single-center test set was not significant, however, on the external validation set, the ACC and AUC were increased by 12.1% and 3.0% respectively, indicating that OPM is useful for fusing the features from the different sources.

4. Disscussion

We proposed and validated a novel deep-learning model called WSOFNet for predicting the IDH status of gliomas. We used transformation-invariant WS features instead of the original images as the WSOFNet inputs and effectively fused the multimodal image features with adaptive fusion and orthogonal projections. In addition, to effectively extract the single-modality features related to IDH status, four auxiliary losses were also used. The IDH status predictions obtained by WSOFNet were more accurate than those of the state-of-the-art methods, especially on the external validation dataset acquired from another center, proving the effectiveness and robustness of the proposed method.

A fundamental problem faced by deep learning-based IDH-status prediction models is how to stabilize their classification abilities for data acquired from different sites. Even though the current models can achieve high IDH status prediction accuracies for datasets derived from the same center (Fig. 5 and Table 2), testing on the data from other centers is usually not as well as expected (Fig.7 and Table 3). This may be caused by the different image intensity distributions of multi-center dataset. Using the images with a fixed intensity distribution to train the model and then using the images with another intensity distribution to test will loss the coherence and influence the generalization ability of the model. To deal with this issue, we proposed to use WS features instead of the original images as the input of the network. As illustrated in Fig. 11, we compared the kernel density estimations (KDE) of original T1ce images and those of WS features for train, test, and external validation sets, respectively. We notice that the KDEs of T1ce images in train set and test set are also the same, but those between test set and validation set are significantly different (Fig. 11(a) and Fig. 11(b)), however, regarding to the KDEs of WS features, there is no significant difference between train and test sets, nor between test and validation sets. This indicates that WS features can reduce the discrepancies in multicenter datasets and increase the generalization capability of the developed model, this also explains why using the WS features rather than original images as input in our model yields better prediction performance (comparing WSOFNet-v1 and WSOFNet-v2 in Table 4).

Considering that different modalities may provide different compensatory information for characterizing gliomas, fusing this multimodal information is more helpful for improving the IDH status prediction accuracy of the model. The commonly used fusion strategies are splicing together the multimodal images and then using convolutional layers to extract the fused features or using the different convolutions to extract singlemodal features and finally splicing them together.The former fusion strategy fully considers the image-level interactions among different modalities, which can be regarded as equally-weighted summations of different modalities. However, the contributions of different modalities to the predictive ability of the model are inconsistent; therefore, this fusion strategy is less effective for predicting

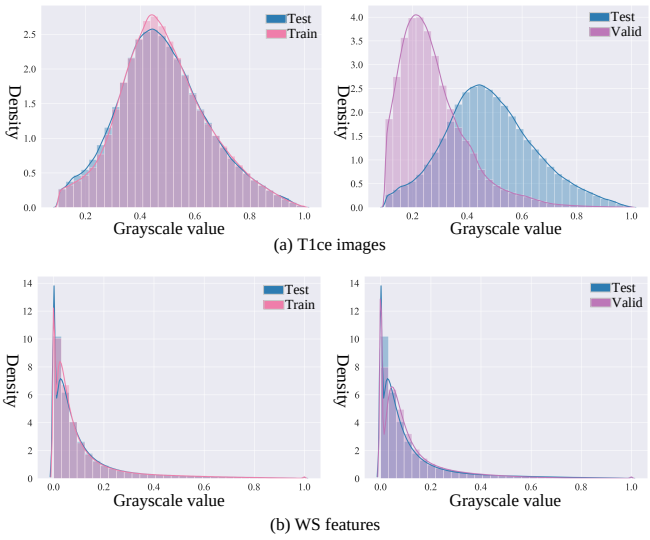

Figure 11: Kernel density estimations (KDE) of T1ce images and those of the corresponding WS features. Red, purple, and blue curves represent the intensity distributions on the training set, external validation set and test set, respectively.

multicenter datasets, in which the contributions of different modalities in the training set may be totally different from those in the testing set. The latter fusion strategy can address this problem, but it does not consider the interactions between feature maps with different modes and may influence the resulting prediction performance. Considering the issues of the current multimodal fusion strategies, we proposed an AMF2M module to fuse the different modal features in an adaptive manner, combining spatial attention and channel attention modules to adaptively update the weights of multimodal features. These adaptive weights allow the network to emphasize the features that greatly contribute to IDH prediction and to overcome the inconsistency among the image intensities in multicenter datasets. This explains why in the ablation experiments, WSOFNet-v4 performed much better than WSOFNet-v3 on the external validation set. After presenting the AMF2M module, we also introduced the OPM module to fuse the useful features that may be lost by the AMF2M. Exploiting the properties of orthogonal projection, we used the OPM to extract the lost information, which could guarantee that the features output by the OPM were not contained in the fused features of the AMF2M. These lost features were fused again to compensate for the previously fused information. Table 4 shows that the OPM could further improve the prediction ability of the full model, especially on the external validation set, because combining two orthogonal fused features can provide compensation information for fully characterizing glioma properties and therefore increase the generalization ability of the prediction model.

All the results validate that using transformation-invariant WS features as input and exploring the orthogonal projection-based adaptive rules to fuse multimodal features are more advantageous in predicting IDH status in multicenter gliomas. However, there are still several limitations to be addressed in the future. First, even though the WS features can reduce the intensity difference of multi-center images, we must admit that they cannot solve the problem of the texture difference in the multicenter dataset, if the images from multi-centers have significantly different texture, the prediction performance of WSOFNet will be influenced. Secondly, in this work, since in the external validation dataset, the T1w and Flair images were not acquired while synthesized with GAN model, the performance of the generation model may have a certain impact on the prediction performance for external validation set, acquiring the multi-center dataset with multiple modalities to further validate our model will be preferred in the future. Finally, although the orthogonal projection used in this work can effectively decouple the features from the fused information at a numerical level, how to combine the novel decoupling methods to disentangle the fused features at semantic level may be possible to further improve the prediction performance.

5. Conclusion

We proposed a WSOFNet for noninvasively and accurately predicting IDH mutation status in multicenter gliomas with multimodal MR images. Using the transformation invariant WS features as the input of the network can overcome the influences of different intensity distributions of the multicenter datasets. In addition, to fully exploit the multimodal MR image information for promoting IDH prediction accuracy, a two-stage fusion strategy based on adaptive spatial and channel attention, as well as orthogonal projections were used. The comparison and ablation results illustrated that, with the help of the WS features, fusion strategies, and auxiliary losses, the proposed method can greatly improve the prediction accuracy of IDH mutation status in multicenter gliomas, suggesting the potential utility of the WSOFNet in computer-aided glioma diagnosis.

Acknowledgments

This work was partially funded by the National Nature Science Foundations of China (Grant No.62161004), Guizhou Provincial Science and Technology Projects (QianKeHe ZK [2021] Key 002), Nature Science Foundations of Guizhou Province (QianKeHe [2020]1Y255).

References

Ai, L., Bai, W., Li, M., 2022. Tdabnet: Three-directional attention block network for the determination of idh status in low-and high-grade gliomas from mri. Biomedical Signal Processing and Control 75, 103574. Doi:10.1016/j.bspc.2022.103574.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C., 2017. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data 4, 1–13. Doi:10.1038/sdata.2017.117.

Bangalore Yogananda, C.G., Shah, B.R., Vejdani-Jahromi, M., Nalawade, S.S., Murugesan, G.K., Yu, F.F., Pinho, M.C., Wagner, B.C., Mickey,

B., Patel, T.R., et al., 2020. A novel fully automated mri-based deep-learning method for classification of idh mutation status in brain gliomas. Neuro-oncology 22, 402–411. Doi:10.1093/neuonc/noz199.

Bruna, J., Mallat, S., 2013. Invariant scattering convolution networks. IEEE transactions on pattern analysis and machine intelligence 35, 1872–1886. Doi:10.1109/TPAMI.2012.230.

Chang, K., Bai, H.X., Zhou, H., Su, C., Bi, W.L., Agbodza, E., Kavouridis, V.K., Senders, J.T., Boaro, A., Beers, A., et al., 2018. Residual convolutional neural network for the determination of idh status in low-and high-grade gliomas from mr imagingneural network for determination of idh status in gliomas. Clinical Cancer Research 24, 1073–1081. Doi:10.1158/1078-0432.CCR-17-2236.

Chen, X., Wu, Y., Zhao, G., Wang, M., Gao, W., Zhang, Q., Lin, Y., 2019. Automatic histogram specification for glioma grading using multicenter data. Journal of healthcare engineering 2019. Doi:10.1155/2019/9414937.

Cheng, J., Liu, J., Kuang, H., Wang, J., 2022. A fully automated multimodal mri-based multi-task learning for glioma segmentation and idh genotyping. IEEE Transactions on Medical Imaging Doi:10.1109/TMI.2022.3142321.

Choi, K.S., Choi, S.H., Jeong, B., 2019. Prediction of idh genotype in gliomas with dynamic susceptibility contrast perfusion mr imaging using an explainable recurrent neural network. Neuro-oncology 21, 1197–1209. Doi:10.1093/neuonc/noz095.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al., 2013. The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of digital imaging 26, 1045–1057. Doi:10.1007/s10278-013-9622-7.

Dai, Z., Liu, H., Le, Q.V., Tan, M., 2021. Coatnet: Marrying convolution and attention for all data sizes, in: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.. pp. 3965–3977. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/20568692db622456cc42a2e853ca21f8-Paper.pdf.

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J., 2021. Repvgg: Making vgg-style convnets great again, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13733–13742. Doi:10.1109/cvpr46437.2021.01352.

Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B., 2020. Sharpness-aware minimization for efficiently improving generalization. arXiv preprint arXiv:2010.01412 Doi:10.48550/arXiv.2010.01412.

Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P., 2019. Res2net: A new multi-scale backbone architecture. IEEE transactions on pattern analysis and machine intelligence 43, 652–662. Doi:10.1109/tpami.2019.2938758.

Ge, C., Gu, I.Y.H., Jakola, A.S., Yang, J., 2018. Deep learning and multi-sensor fusion for glioma classification using multistream 2d convolutional networks, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE. pp. 5894–5897. Doi:10.1109/EMBC.2018.8513556.

Han, Y., Xie, Z., Zang, Y., Zhang, S., Gu, D., Zhou, M., Gevaert, O., Wei, J., Li, C., Chen, H., et al., 2018. Non-invasive genotype prediction of chromosome 1p/19q co-deletion by development and validation of an mri-based radiomics signature in lower-grade gliomas. Journal of neuro-oncology 140, 297–306. Doi:10.1007/s11060-018-2953-y.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. Doi:10.1109/cvpr.2016.90.

Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al., 2019. Searching for mobilenetv3, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 1314–1324. Doi:10.48550/arXiv.1905.02244.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708. Doi:10.1109/cvpr.2017.243.

Jiang, L., Ning, C., Li, J., 2022. Glioma classification framework based on se-resnext network and its optimization. IET Image Processing 16, 596–605. Doi:10.1049/ipr2.12374.

Kim, M., Jung, S.Y., Park, J.E., Jo, Y., Park, S.Y., Nam, S.J., Kim, J.H., Kim, H.S., 2020. Diffusion-and perfusion-weighted mri radiomics model may predict isocitrate dehydrogenase (idh) mutation and tumor aggressiveness in diffuse lower grade glioma. European radiology 30, 2142–2151. Doi:10.1007/s00330-019-06548-3.

Li, Z., Wang, Y., Yu, J., Guo, Y., Cao, W., 2017. Deep learning based radiomics (dlr) and its usage in noninvasive idh1 prediction for low grade glioma. Scientific reports 7, 1–11. Doi:10.1038/s41598-017-05848-2.

Li, Z.C., Yan, J., Zhang, S., Liang, C., Lv, X., Zou, Y., Zhang, H., Liang, D., Zhang, Z., Chen, Y., 2022. Glioma survival prediction from whole-brain mri without tumor segmentation using deep attention network: a multi-center study. European Radiology 32, 5719–5729. Doi:10.1007/s00330-022-08640-7.

Liang, S., Zhang, R., Liang, D., Song, T., Ai, T., Xia, C., Xia, L., Wang, Y., 2018. Multimodal 3d densenet for idh genotype prediction in gliomas. Genes 9, 382. Doi:10.3390/genes9080382.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022. Doi:10.1109/iccv48922.2021.00986.

Lu, C.F., Hsu, F.T., Hsieh, K.L.C., Kao, Y.C.J., Cheng, S.J., Hsu, J.B.K., Tsai, P.H., Chen, R.J., Huang, C.C., Yen, Y., et al., 2018. Machine learning–based radiomics for molecular subtyping of gliomasmachine learning for molecular subtyping of gliomas. Clinical Cancer Research 24, 4429–4436. Doi:10.1158/1078-0432.CCR-17-3445.

Lu, Z., Bai, Y., Chen, Y., Su, C., Lu, S., Zhan, T., Hong, X., Wang, S., 2020. The classification of gliomas based on a pyramid dilated convolution resnet model. Pattern Recognition Letters 133, 173–179. Doi:10.1016/j.patrec.2020.03.007.

van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of Machine Learning Research 9.

Mallat, S., 2012. Group invariant scattering. Communications on Pure and Applied Mathematics 65, 1331–1398. Doi:10.1002/cpa.21413.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2014. The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging 34, 1993–2024. Doi:10.1109/TMI.2014.2377694.

Nalawade, S., Murugesan, G.K., Vejdani-Jahromi, M., Fisicaro, R.A., Yogananda, C.G.B., Wagner, B., Mickey, B., Maher, E., Pinho, M.C., Fei, B., et al., 2019. Classification of brain tumor isocitrate dehydrogenase status using mri and deep learning. Journal of Medical Imaging 6, 046003. Doi:10.1117/1.JMI.6.4.046003.

Naser, M.A., Deen, M.J., 2020. Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images. Computers in biology and medicine 121, 103758. Doi:10.1016/j.compbiomed.2020.103758.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 Doi:10.48550/arXiv.1409.1556.

Sun, H., Yin, L., Li, S., Han, S., Song, G., Liu, N., Yan, C., 2013. Prognostic significance of idh mutation in adult low-grade gliomas: a meta-analysis. Journal of Neuro-oncology 113, 277–284. Doi:10.1007/s11060-013-1107-5.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence. Doi:10.1609/aaai.v31i1.11231.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR. pp. 6105–6114. Doi:10.48550/arXiv.1905.11946.

Tietze, A., Choi, C., Mickey, B., Maher, E.A., Ulhøi, B.P., Sangill, R., Lassen-Ramshad, Y., Lukacova, S., Østergaard, L., Von Oettingen, G., 2017. Noninvasive assessment of isocitrate dehydrogenase mutation status in cerebral gliomas by magnetic resonance spectroscopy in a clinical setting. Journal of neurosurgery 128, 391–398. Doi:10.3171/2016.10.JNS161793.

Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A., 2021. Mlp-mixer: An all-mlp architecture for vision 34, 24261–24272. URL: https://proceedings.neurips.cc/paper/2021/file/cba0a4ee5ccd02fda0fe3f9a3e7b89fe-Paper.pdf.

van der Voort, S.R., Incekara, F., Wijnenga, M.M., Kapsas, G., Gahrmann, R., Schouten, J.W., Nandoe Tewarie, R., Lycklama, G.J., De Witt Hamer, P.C., Eijgelaar, R.S., et al., 2022. Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning. Neuro-oncology Doi:10.1093/neuonc/noac166.

Wang, X.W., Ciccarino, P., Rossetto, M., Boisselier, B., Marie, Y., Desestret, V., Gleize, V., Mokhtari, K., Sanson, M., Labussière, M., 2014. Idh mutations: genotype-phenotype correlation and prognostic impact. BioMed research international 2014. Doi:10.1155/2014/540236.

Wesseling, P., Capper, D., 2018. Who 2016 classification of gliomas. Neuropathology and applied neurobiology 44, 139–150. Doi:10.1111/nan.12432.

Wu, P., Wang, Z., Zheng, B., Li, H., Alsaadi, F.E., Zeng, N., 2023. Aggn: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion. Computers in Biology and Medicine 152, 106457. Doi:10.1016/j.compbiomed.2022.106457.

Yu, J., Shi, Z., Lian, Y., Li, Z., Liu, T., Gao, Y., Wang, Y., Chen, L., Mao, Y., 2017. Noninvasive idh1 mutation estimation based on a quantitative radiomics approach for grade ii glioma. European radiology 27, 3509–3522. Doi:10.1007/s00330-016-4653-3.

Zhang, Q., Xu, Y., Zhang, J., Tao, D., 2022. Vitaev2: Vision transformer advanced by exploring inductive bias for image recognition and beyond. arXiv preprint arXiv:2202.10108 Doi:10.48550/arXiv.2202.10108.

Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6848–6856. Doi:10.1109/cvpr.2018.00716.

Zhao, G., Wu, Y., He, M., Bai, J., Cheng, J., Lin, Y., 2019. Preprocessing and grading of glioma data acquired from multicenter. Journal of Medical Imaging and Health Informatics 9, 1236–1245. Doi:10.1166/jmihi.2019.2724.

Zhao, J., Huang, Y., Song, Y., Xie, D., Hu, M., Qiu, H., Chu, J., 2020. Diagnostic accuracy and potential covariates for machine learning to identify idh mutations in glioma patients: evidence from a meta-analysis. European Radiology 30, 4664–4674. Doi:10.1007/s00330-020-06717-9.

Zhou, H., Vallières, M., Bai, H.X., Su, C., Tang, H., Oldridge, D., Zhang, Z., Xiao, B., Liao, W., Tao, Y., et al., 2017. Mri features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro-oncology 19, 862–870. Doi:10.1093/neuonc/now256.

Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L., 2020. Hi-net: hybrid-fusion network for multi-modal mr image synthesis. IEEE transactions on medical imaging 39, 2772–2781. Doi:10.1109/tmi.2020.2975344.