Proceedings of Machine Learning Research vol 278:1–6, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Multi-instance Causal Representation Learning-based Network for Glioma Grading

Hanbing Zhang
School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China

Guohua Zhao
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China
Department of Magnetic Resonance Imaging, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China

Yusong Lin* YSLIN@HA.EDU.CN
School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450002, China
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Gliomas are the most common primary intracranial malignant tumors, characterized by high heterogeneity, recurrence, and mortality. Accurate grading is essential for treatment planning and prognosis assessment. MRI, as a non-invasive modality, is widely used, but traditional diagnosis depends on expert experience, leading to subjectivity and inefficiency. AI-based automatic grading has made progress, yet challenges persist due to tumor boundary ambiguity, structural heterogeneity, and the "black box" nature of AI models, limiting robustness, generalization, and interpretability. To address these issues, this study proposes a multi-instance causal representation learning-based network for glioma grading (MCRNet). MCRNet employs multi-instance learning to aggregate MRI slice features, effectively handling tumor heterogeneity. The causal-aware attention mechanism (CAAM) and causal-aware dynamic aggregation mechanism (CDAM) enhance feature selection and aggregation efficiency. Evaluated on BraTS2020 and a private clinical dataset, MCRNet improves robustness, generalization, and interpretability. It minimizes the performance gap between validation and test sets, reducing the AUC difference by up to 3.21% compared to existing methods, demonstrating its potential for reliable clinical application.

Keywords: Medical Image, Glioma Grading, Deep Learning, Attention Mechanism, Causal Inference.

1. Introduction

Gliomas, arising from glial cells, are among the most common primary malignant brain tumors, with high incidence, recurrence, and mortality. They account for 40-60% of primary central nervous system (CNS) tumors, with glioblastomas being the most aggressive subtype. According to the 2021 World Health Organization classification, gliomas are graded 1-4, where low-grade gliomas (LGG, grades 1-2) progress slowly, while high-grade gliomas (HGG, grades 3-4) are highly aggressive, with survival often under two years despite treatment. Accurate grading is crucial for clinical decision-making. MRI-based (magnetic resonance imaging, MRI) automatic glioma grading, driven by deep learning, has improved classification, with convolutional neural networks (CNN) effectively extracting local tumor features. However, gliomas' high heterogeneity leads to challenges: CNN focus on local features, neglecting global spatial relationships; attention mechanisms enhance grading

but struggle with nonlinear feature interactions and reliance on precise region weighting; and the lack of explicit causal relationships limits interpretability, hindering clinical adoption. Addressing these limitations is key to advancing AI-assisted glioma diagnosis.

To overcome these issues, this study proposes a multi-instance causal representation learning-based network for glioma grading (MCRNet). By integrating attention mechanisms, causal inference, and graph neural networks within a multi-instance learning framework, MCRNet enhances robustness, generalization, and interpretability. Specifically, MCRNet incorporates a causality-aware attention mechanism and a causality-aware dynamic aggregation mechanism, making AI decisions more clinically reliable and transparent.

2. Methods

In this section, the proposed MCRNet is described in detail (see Figure 1), which mainly consists of four phases: feature extraction, feature coding, feature aggregation, and classification.

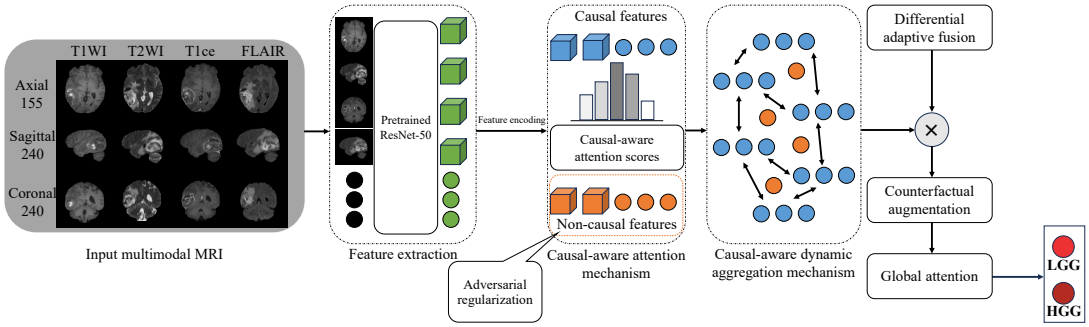

Figure 1: Schematic diagram of the overall architecture of the proposed MCRNet.

2.1. Data preprocessing

To address the heterogeneity and complexity of glioma MRI, this study adopts multi-instance learning as the foundational framework, necessitating targeted data preprocessing (see Figure 2) to ensure an effective grading model. First, the four-modal MRIs of each patient are normalized to eliminate intensity variations across modalities. A min-max normalization maps intensity values to [0,1], reducing biases from scanning equipment, patient positioning, and other factors, ensuring stable model training. Next, whole-brain slices are extracted along axial, sagittal, and coronal planes for all four modalities (T1WI, T2WI, T1ce, and FLAIR) to comprehensively capture tumor features. To standardize imaging resolution, bilinear interpolation ensures uniform pixel size across slices. Finally, following the multi-instance learning framework, all slices of a patient are grouped into a package, treating each slice as an independent instance. The tumor grade (HGG or LGG) serves as the package label, inheriting the patient's pathological diagnosis. Through weakly supervised learning, the model infers the overall tumor grade from multiple slices, enhancing grading robustness and generalization.

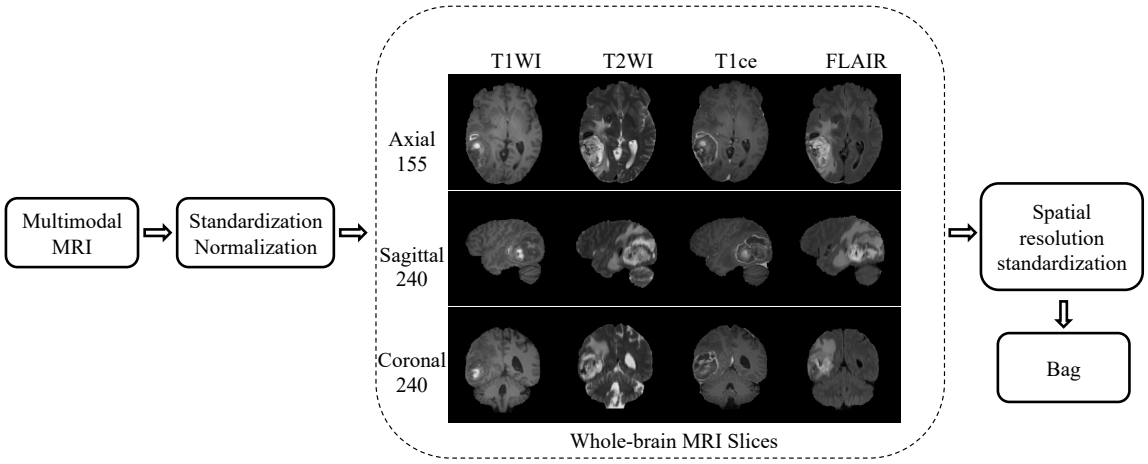

Figure 2: Flowchart of data preprocessing.

2.2. Causal-aware attention mechanism

The causal-aware attention mechanism (CAAM), a core module of MCRNet, decouples each instance's features into causal and non-causal components and computes a causal-attention matrix to highlight key tumor grading features in glioma MRI. By focusing on causal features, it effectively reduces interference from irrelevant regions. Specifically, instance features are first projected into causal and non-causal spaces for further processing. With this transformation, the network is able to capture feature changes in the data caused by causal and other potential factors, respectively.

To minimize interference from non-causal features, an adversarial network with two fully connected layers (ReLU) extracts noisy signals. The mean output defines the adversarial regularization loss L_{adv}, integrated with negative weighting into the total loss, guiding the network to prioritize causal features.

To refine tumor focus, the network computes a causal attention matrix. Before this, to prevent excessive sparsity, causal and non-causal features are weighted and fused appropriately. Then, Q and K are obtained by performing two linear transformations on the weighted fused features h_{causal} to compute the attention matrix:

$$Q, K = W_q h_{\text{causal}}\,, W_k h_{\text{causal}} \tag{1}$$

$$\text{attn} = \frac{Q \cdot K^T}{\sqrt{d'}} \tag{2}$$

Where W_q and W_k are the linear transformation weights used to generate Q and K, d' is the dimension of h_{causal}, and $1/\sqrt{d'}$ is used as a scaling factor to prevent the value from being too large in the computation of the attention matrix, which can lead to the disappearance of the gradient or instability. Ultimately, the CAAM assigns weights to the features of each instance in order to better weigh the neighbors when performing feature aggregation, which in turn focuses the model on the key features of the tumor region, thus improving the robustness and generalization of the grading.

2.3. Causal-aware dynamic aggregation mechanism

To fully utilize contextual information between instances, this study designs a causal-aware dynamic aggregation mechanism (CDAM), consisting of three steps: adjacency matrix construction, Top-k neighbor selection, and graph neural network aggregation.

Adjacency matrix construction. Using causal-aware features h_{causal} and attention weights $attn$, an adjacency matrix A is built. Specifically, a Gaussian kernel computes the similarity between instances i and j, integrating causal-aware attention scores. Where α_i is the causal-aware attention score of i, computed by $\alpha_i = \sigma(W h^i_{\text{causal}} + b)$.

$$A_{ij} = \exp\left(-\left\| h^i_{\text{causal}} - h^j_{\text{causal}} \right\|^2\right) \cdot \frac{\alpha_i + \alpha_j}{2} \tag{3}$$

Topk neighbor selection. Since fully connected graphs lead to increased computational overhead, a Topk neighbor selection strategy (Li et al., 2024) is used for each i, retaining only the k neighbors that are most relevant to it. Specifically, denote the set of Topk neighbors of i by $\mathcal{N}(i)$ and construct the sparse adjacency matrix $A^{(k)}$. This strategy both reduces redundant information and ensures that the most representative neighbor information is fully utilized.

$$A^{(k)}_{ij} = \begin{cases} A_{ij}, & j \in \mathcal{N}(i) \\ 0, & \text{Others} \end{cases} \tag{4}$$

Graph neural network aggregation. After obtaining the sparse neighbor matrix $A^{(k)}$, it is necessary to use a graph neural network for information aggregation. For the feature $h^{(l)}_i$ (initial value $h^{(0)}_i = h^i_{\text{causal}}$) at layer l for each i, the update process is as follows. Where $W^{(l)}$ and $b^{(l)}$ are the weights and biases of the lth layer graph neural network, respectively, and $\sigma(\cdot)$ represents the ReLU.

$$h^{(l+1)}_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} A^{(k)}_{ij} W^{(l)} h^{(l)}_j + b^{(l)}\right) \tag{5}$$

After l-layer aggregation, the aggregated feature $h_{\text{agg}} \in \mathbb{R}^{B \times N \times d''}$ (d'' is the dimension of the aggregated feature) is obtained, which allows each instance of the feature to contain both its own information and incorporate the contextual information of its most relevant neighbors to better capture the spatial heterogeneity of the gliomas on different slices.

3. Experiences and results

3.1. Experimental setup

In this study, the BraTS2020 dataset (Menze et al., 2015; Bakas et al., 2017a, 2018, 2017b) and the private dataset from the First Affiliated Hospital of Zhengzhou University were used to carry out the related experiments. To evaluate the proposed MCRNet, the BraTS2020 dataset was divided into training and validation sets in a ratio of 4:1, and the private dataset was used as the test set.

The evaluation metrics used in the above experiments include accuracy, F1 score, and AUC. The experiments are completed on a Linux server equipped with an NVIDIA GeForce RTX 3090 graph-

ics card, and the deep learning frameworks used are PyTorch 1.13.1, CUDA 11.6.0, and Python 3.8. The experimental hyper-parameters are set as shown in Table 1.

Table 1: Hyperparameter settings.

Hyperparameter	Value	Description
λ_{cf}	0.1	L_{cf} weight.
Topk	6	Number of neighbor nodes.
dropout	0.3	Prevention of overfitting.
λ_{adv}	0.01	L_{adv} coefficient.

3.2. Experimental results

As shown in Table 2, MCRNet achieves superior performance on both validation and test sets. Especially in the AUC metric, it reaches 99.60% (validation) and 99.55% (test), with the smallest difference between the two, indicating excellent generalization capability. MCRNet is more robust in handling complex, heterogeneous glioma MRIs and achieves accurate grading, comparable to ResMT, a hybrid 3D MRI network. Other models such as IPTV, U-Net, CSF-Glioma, and TransMIL generally perform better on validation, especially in AUC, but lack robustness. Moreover, MCRNet's results show minimal fluctuation: only 0.16% drop in accuracy, 0.17% in F1, and 0.05% in AUC, fully validating its robustness and generalization—attributable to the integration of multi-instance learning, CAAM, and CDAM.

Table 2: Comparison of MCRNet's performance with five state-of-the-art models (maximum values of this metric in bold)

Models	Validation set			Test set		
	Accuracy (%)	F1 score (%)	AUC (%)	Accuracy (%)	F1 score (%)	AUC (%)
IPTV (Cheng et al., 2022)	94.00	95.70	97.50	90.37	90.40	95.23
ResMT (Cui et al., 2024)	97.01	97.57	99.53	96.84	96.83	99.48
U-Net (Yu et al., 2022)	89.29	89.53	89.00	84.02	83.93	90.78
CSF-Glioma (Zheng et al., 2023)	91.04	93.74	96.14	87.21	87.27	92.83
TransMIL (Shao et al., 2021)	95.38	95.36	99.22	94.97	94.98	98.69
MCRNet	**97.08**	**97.08**	**99.65**	**96.92**	**96.91**	**99.55**

4. Conclusion

This study proposes MCRNet, integrating multi-instance learning, attention mechanisms, causal inference, and graph neural networks. It introduces CAAM and CDAM to enhance grading robustness, generalization, and interpretability. CAAM dynamically adjusts instance weights, while CDAM utilizes graph neural networks with Top-k neighbor selection for improved feature aggregation. Experimental results demonstrate significant performance improvements. Future work will

validate MCRNet on multi-center datasets and refine causal inference methods to further enhance interpretability and clinical applicability.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China under Grant No. 82441022, and Collaborative Innovation Major Project of Zhengzhou under Grant No.20XTZX05015.

References

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, and et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, 4 (1):170117, 2017a. doi: 10.1038/sdata.2017.117.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, and et al. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection, 07 2017b.

Spyridon Bakas, Mauricio Reyes, András Jakab, and et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *CoRR*, abs/1811.02629, 2018.

Jianhong Cheng, Jin Liu, Hailin Yue, Harrison Bai, Yi Pan, and Jianxin Wang. Prediction of glioma grade using intratumoral and peritumoral radiomic features from multiparametric mri images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2):1084–1095, 2022. doi: 10.1109/TCBB.2020.3033538.

Honghao Cui, Zhuoying Ruan, Zhijian Xu, and et al. Resmt: A hybrid cnn-transformer framework for glioma grading with 3d mri. *Computers and Electrical Engineering*, 120:109745, 2024. ISSN 0045-7906. doi: https://doi.org/10.1016/j.compeleceng.2024.109745.

Jiawen Li, Yuxuan Chen, Hongbo Chu, and et al. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis, 2024.

B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, and et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans Med Imaging*, 34(10):1993–2024, 2015. doi: 10.1109/tmi.2014.2377694.

Zhuchen Shao, Hao Bian, Yang Chen, and et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 2136–2147. Curran Associates, Inc., 2021.

X. Yu, Y. Wu, Y. Bai, and et al. A lightweight 3d unet model for glioma grading. *Phys Med Biol*, 67(15), 2022. doi: 10.1088/1361-6560/ac7d33.

Y. Zheng, D. Huang, Y. Feng, and et al. Csf-glioma: A causal segmentation framework for accurate grading and subregion identification of gliomas. *Bioengineering (Basel)*, 10(8), 2023. doi: 10.3390/bioengineering10080887.