DEC-Net: A Multi-Task Distillation Expert Collaboration Network for Glioma Grading and Molecular Subtyping Using Multimodal MRI

Fei Liu[1,†], Jianhong Cheng[3,†], Xu Tian[2], Jianxin Wang[1,2], Yue Hu[4], Jin Liu[1,2,*], and Hong-Dong Li[1,2,*]

[1] Xinjiang Key Laboratory of Intelligent Computing and Smart Applications, School of Software, Xinjiang University, Urumqi, Xinjiang, China
[2] Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, Hunan, China
[3] Institute of Guizhou Aerospace Measuring and Testing Technology, Guizhou Provincial Key Laboratory of Quality Inspection and Evaluation for Basic Components and Equipment, Guiyang, Guizhou, China
[4] School of Computer Science and Technology, Joint International Research Laboratory of Silk Road Multilingual Cognitive Computing, Xinjiang University, Urumqi, Xinjiang, China
{liujin06, hongdong}@csu.edu.cn

Abstract. Glioma grading and molecular subtyping are challenging due to significant intratumoral heterogeneity and inter-task interference in multi-task learning settings. Existing multi-task learning approaches often fail to effectively disentangle shared and task-specific representations, limiting their ability to capture complex biological patterns. To address these issues, we propose the Distillation Expert Collaboration Network (DEC-Net), which comprises two key modules. First, a Distillation ResModule utilizes a knowledge distillation mechanism to guide a ResNet backbone in learning discriminative multimodal representations from tumor-relevant regions. Second, a Progressive Mixture-of-Experts module hierarchically integrates shared and task-specific experts, enabling layer-wise adaptation of representation spaces to mitigate inter-task interference. Furthermore, a multi-task dependency loss is introduced to explicitly model the biological correlation between glioma grades and subtypes, thereby improving prediction consistency. Extensive experiments conducted on a large hybrid cohort comprising UCSF-PDGM and EGD datasets, along with external validation on the TCGA dataset, demonstrate that DEC-Net achieves superior performance and robust generalization compared with state-of-the-art methods for non-invasive glioma diagnosis.

Keywords: Glioma · Multi-task learning · Grading · Molecular subtyping.

[†] These authors contributed equally.
[*] Corresponding authors: Jin Liu and Hong-Dong Li.

1 Introduction

Gliomas are the most common and aggressive primary brain tumors [19]. The 2021 WHO classification emphasizes molecular biomarkers, such as IDH mutation and 1p/19q codeletion, alongside histological grade for precise diagnosis [11]. However, current methods for determining these biomarkers rely on invasive tissue sampling, which is costly and prone to sampling bias due to intratumoral heterogeneity [15]. In contrast, multimodal MRI offers a non-invasive and repeatable alternative for comprehensive tumor characterization [4, 6]. Therefore, developing non-invasive approaches to predict glioma grade and molecular subtypes is of great clinical significance.

Multimodal MRI is widely used for glioma grading and molecular subtyping, providing complementary information on tumor structure and biological heterogeneity [9, 14]. While grading typically relies on macroscopic characteristics linked to tumor aggressiveness [18], molecular subtyping identifies subtler imaging phenotypes correlated with IDH mutation and 1p/19q codeletion [13]. Multimodal MRI approaches face several limitations. Most methods rely on accurate tumor segmentation, increasing complexity, computational cost [27], and dependence on manual annotation, which suffers from inter-observer variability [28]. Despite the clinical and biological links between glioma grading and molecular subtyping, many studies model them independently [17], neglecting intrinsic correlations and shared representations, thus limiting multi-label prediction. Therefore, developing a unified framework that jointly models grading and subtyping remains a key challenge.

To address these challenges, we propose a distillation expert collaboration network for glioma grading and molecular subtyping. First, the Distillation ResModule (DRModule) leverages a knowledge distillation framework, where an anomaly-detection-guided [10] teacher directs the student ResNet to focus on discriminative tumor-related regions. Second, a Progressive Mixture of Experts (PMoE) employs a routing network to coordinate shared and task-specific experts, effectively balancing common and task-unique features. Additionally, a Dependency Loss (D-Loss) is introduced to model inherent biological correlations among tasks. Extensive experiments validate the effectiveness and clinical potential of our approach. Our main contributions are summarized as follows:

1. We propose a DRModule that employ an anomaly-detection-guided teacher-student mechanism, the module enables the network to adaptively focus on and refine discriminative tumor-relevant features.
2. We introduce a PMoE module to mitigate inter-task interference. This module dynamically balances shared and task-specific experts through a routing mechanism to enhance both grading and subtyping accuracy.
3. We develop a multi-task D-Loss to model the inherent biological correlations among grading and molecular subtyping. Extensive experiments on large-scale cohorts demonstrate our DEC-Net achieves SOTA performance.

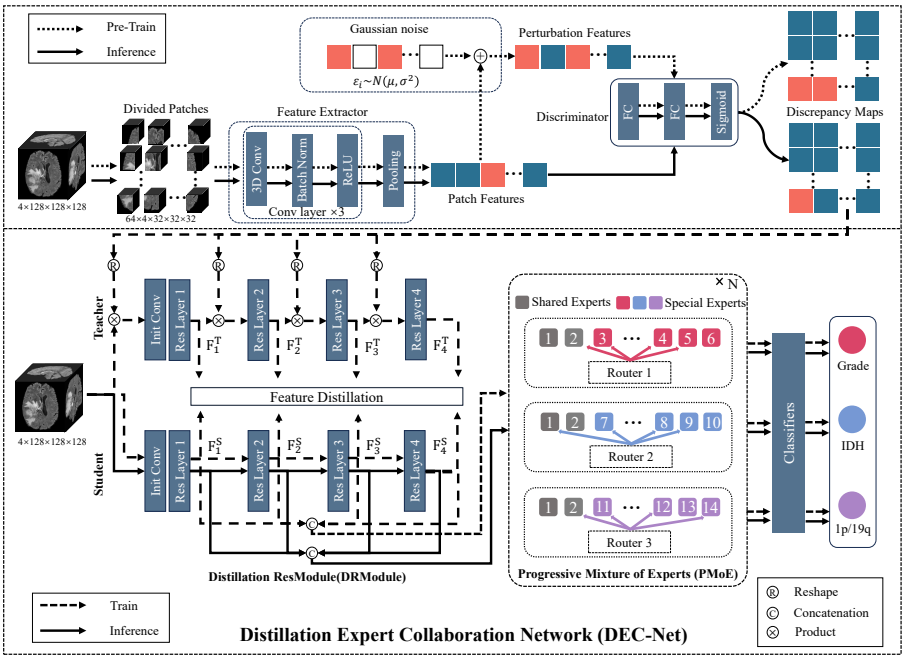

Fig. 1: Overview of the proposed DEC-Net.

2 Method

As illustrated in Fig. 1, the proposed DEC-Net is designed for glioma grading and molecular subtyping using multimodal MRI. Let $\mathbf{X}_i = [\mathbf{X}_i^{\mathrm{T1}}, \mathbf{X}_i^{\mathrm{T1ce}}, \mathbf{X}_i^{\mathrm{T2}}, \mathbf{X}_i^{\mathrm{FLAIR}}]$ denote the input modalities. DEC-Net integrates two key modules: the DRModule for adaptive tumor-relevant extraction and the PMoE for alleviating inter-task interference.

2.1 Distillation ResModule with a Pre-trained Network

The DRModule employs a Teacher-Student paradigm to address inter-modal discrepancies and enhance feature representation through knowledge distillation. To capture tumor-relevant regions, we first utilize an unsupervised pre-training framework. Specifically, patch features are extracted from multi-modal MRI. By applying Gaussian noise to create perturbation features, we train a discriminator to identify discrepancies between normal and anomalous patch features. This process yields a discrepancy map, which functions as anomaly-aware prior knowledge. This prior is then fed into the four stages of the teacher branch to enhance the accuracy of the learned features. Following this, we employ a ResNet backbone to extract multi-scale feature maps $\mathcal{F} = \{\mathbf{F}_i\}_{i=1}^{4}$, where i indexes the network stages. Each feature map $\mathbf{F}_i \in \mathbb{R}^{C_i \times D_i \times H_i \times W_i}$ denotes the feature map

with channel, depth, height, and width dimensions. $\mathbf{F}_0$ denote the init input four modalities. In DRModule, the student branch generates a set of multi-scale features: $\mathcal{F}_{\mathrm{S}} = \{\mathbf{F}_i^{\mathrm{S}}\}_{i=1}^{4}$. To improve the learning capability of student branch, a teacher branch is introduced. The teacher branch shares the same backbone architecture as the student branch but is additionally guided by the discrepancy map produced during the pre-training stage. The teacher network is not updated via gradient descent. Instead, its parameters are maintained as an exponential moving average (EMA) [22] of the student parameters throughout training. This EMA update ensures stable and consistent supervisory signals for the student. The discrepancy map is upsampled to match each backbone scale:

$$\mathbf{F}_i^{\mathrm{T}} = \mathrm{ResLayer}_i\big(\mathbf{F}_{i-1}^{\mathrm{T}} \odot \mathrm{UpSample}(\mathbf{M})\big)\,, \tag{1}$$

Specifically, $\mathbf{F}_{i-1}^{\mathrm{T}}$ denote the input activation of the i-th backbone stage in the teacher branch. The discrepancy map $\mathbf{M}$ is upsampled to match the spatial resolution of the stage input. $\mathrm{UpSample}(\cdot)$ denotes spatial upsampling. $\odot$ denotes element-wise multiplication. $\mathrm{ResLayer}_i(\cdot)$ representing the i-th backbone stage. Through this anomaly-guided fusion, the teacher branch produces a set of anomaly-aware feature representations: $\mathcal{F}_{\mathrm{T}} = \{\mathbf{F}_i^{\mathrm{T}}\}_{i=1}^{4}$, which emphasize modality discrepancies. This mechanism preserves deployment efficiency while transferring discrepancy aware knowledge to the student branch.

To enforce alignment between the student and teacher feature hierarchies, DRModule employs a multi-scale feature distillation objective. The distillation loss is formulated as:

$$\mathcal{L}_{\mathrm{distill}} = \lambda \sum_{i=1}^{4} \left\|\mathbf{F}_i^{\mathrm{S}} - \mathbf{F}_i^{\mathrm{T}}\right\|_2^2, \tag{2}$$

where $\mathbf{F}_i^{\mathrm{S}}$ and $\mathbf{F}_i^{\mathrm{T}}$ denote the student and teacher features at stage i, and λ controls the overall strength of distillation.

2.2 Progressive Mixture-of-Experts Module

To effectively balance shared representation learning and task-specific specialization, we introduce a PMoE module that adopts soft parameter sharing to dynamically allocate model capacity across tasks. PMoE performs hierarchical, layer-wise expert routing to progressively disentangle task-invariant features from task-specific patterns. This design alleviates negative transfer among tasks and enables stable and robust multi-task optimization.

The fused multi-scale feature $\mathcal{F}_{\mathrm{S}}$ extracted by the DRModule is first processed, yielding the initial representation $\mathbf{Z}^{(0)} = \mathcal{F}_{\mathrm{S}}$. The representation is subsequently fed into a cascade of L PMoE layers, where shared and task-special knowledge is progressively refined through expert routing.

Each PMoE layer consists of an expert pool and task-specific gating networks. The expert pool contains two types of experts. A set of shared experts $\{\mathrm{E}_{sh,i}^{(l)}\}_{i=1}^{n_{sh}}$ is accessible to all tasks and is responsible for modeling task-invariant representations. In parallel, for each task $t \in \{\mathrm{Grade}, \mathrm{IDH}, 1\mathrm{p}/19\mathrm{q}\}$, a set of task-specific

experts $\{\mathrm{E}_{sp,t,j}^{(l)}\}_{j=1}^{n_{sp}}$ is introduced to capture task-exclusive patterns. Dense routing is adopted throughout all experiments, allowing each task to attend to both shared and task-specific experts. The task-specific representation at layer l is computed as:

$$\mathbf{Z}_t^{(l)} = \sum_{k=1}^{n_{sh}+n_{sp}} \mathrm{Softmax}(\mathrm{G}_{t,k}^{(l)}(\mathbf{Z}^{(l-1)})) \cdot \mathrm{E}_{t,k}^{(l)}(\mathbf{Z}^{(l-1)}), \tag{3}$$

where $\mathrm{E}_{t,k}^{(l)}$ denotes the k-th expert available to task t. Each expert is implemented as a lightweight residual feed-forward network:

$$\mathrm{E}(\mathbf{Z}) = \mathbf{Z} + \mathrm{LayerNorm}\big(\mathbf{W}_2^{\mathrm{E}} \cdot \mathrm{ReLU}(\mathbf{W}_1^{\mathrm{E}}\mathbf{Z})\big), \tag{4}$$

where $\mathbf{W}_1^{\mathrm{E}}$ and $\mathbf{W}_2^{\mathrm{E}}$ are learnable projection matrices. For each task t at layer l, a task-specific gating network $\mathrm{G}_t^{(l)}$ produces routing scores over all accessible experts. The scores are normalized via a softmax function:

$$\mathrm{G}(\mathbf{Z}) = \mathbf{W}_1^{\mathrm{G}} \cdot \mathbf{Z} + \mathbf{b}, \tag{5}$$

After L PMoE layers, the final shared representation $\mathbf{Z}^{(L)}$ is fed into task-specific classification heads:

$$\mathbf{p}_t = \mathrm{Classifier}_t\big(\mathbf{Z}_t^{(L)}\big), \quad t \in \{\mathrm{Grade}, \mathrm{IDH}, \mathrm{1p/19q}\}. \tag{6}$$

Each head consists of a two-layer fully connected network with ReLU activation and dropout regularization.

2.3 Multi-Task Loss

We adopt focal loss to mitigate class imbalance and emphasize hard-to-classify samples for each task. The loss for task t is defined as:

$$\mathcal{L}_{\mathrm{focal}}^{(t)} = -\big(1 - \mathbf{p}_{t,y_t}\big)^{\gamma} \log \mathbf{p}_{t,y_t}, \tag{7}$$

where $\mathbf{p}_{t,y_t}$ is the predicted probability assigned to the ground-truth class y_t of task t, and γ is a focusing parameter that down-weights easy examples, thereby shifting the model's focus toward harder, more informative samples.

Given the strong statistical and biological dependencies inherent in predicting glioma grade, IDH mutation, and 1p/19q codeletion status, we introduce a dependency-aware regularization term. To this end, we leverage a data-driven conditional probability matrix to explicitly model these relationships as priors.

Each binary task contributes two class states, yielding six class-level variables: $\mathcal{C} = \{g_0, g_1, i_0, i_1, q_0, q_1\}$, corresponding to Grade, IDH, and 1p/19q, respectively. To capture inter-task dependencies, we construct a conditional dependency matrix $\mathbf{P} \in \mathbb{R}^{6\times 6}$, where each element $P_{i,j}$ represents the empirical conditional probability of class c_i given class c_j:

$$P_{i,j} = \frac{\mathrm{Count}(c_i \cap c_j)}{\mathrm{Count}(c_j)}, \tag{8}$$

Table 1: Clinical Characteristics of Internal and External Datasets.

	Internal Cohert		**External Cohert**
	UCSF-PDGM	EGD	TCGA
Total	495 (100%)	719 (100%)	167 (100%)
Grade			
LGG	99 (20%)	197 (27%)	65 (39%)
HGG	396 (80%)	474 (66%)	102 (61%)
Unknown	0 (0%)	48 (7%)	0 (0%)
IDH Status			
Mutated	103 (21%)	139 (20%)	58 (35%)
Wildtype	392 (79%)	276 (38%)	91 (54%)
Unknown	0 (0%)	304 (42%)	18 (11%)
1p/19q Status			
Co-deleted	15 (3%)	68 (10%)	13 (8%)
Intact	389 (79%)	160 (22%)	148 (89%)
Unknown	91 (18%)	491 (68%)	6 (3%)

This matrix is estimated from the joint label distribution of the training set to encode realistic statistical correlations among tasks. Accordingly, the dependency loss is defined as:

$$\mathcal{L}_{\text{dep}} = -\sum_{k=1}^{6} \mathrm{u}_k \sum_{z=1}^{6} P_{z,k} \log\left(\mathbf{p}_z + \epsilon\right), \tag{9}$$

where $\mathbf{p}_z$ is the predicted probability of class z, and ϵ is a small constant for numerical stability. $\mathrm{u}_k \in \{0, 1\}$ is a binary indicator that equals 1 if class k is a ground-truth label, and 0 otherwise. The final training objective is formulated as:

$$\mathcal{L}_{\text{total}} = \sum_{t} \mathcal{L}_{\text{focal}}^{(t)} + \alpha\, \mathcal{L}_{\text{dep}} \; + \beta\, \mathcal{L}_{\text{distill}}. \tag{10}$$

where α and β are balancing coefficients. In this work, α and β are set to 0.3 and 0.1, respectively, reflecting their relative contributions to the total loss.

3 Experiments

3.1 Datasets

The framework was evaluated on an internal cohort (UCSF-PDGM [1] and EGD [24]) and an external test cohort (TCGA-GBM/LGG [20,21]). Both include multimodal MRI (T1, T1ce, T2, FLAIR). Preprocessing involved co-registration to the SRI24 atlas, resampling to 1 mm isotropic resolution, bias field correction, skull-stripping, and z-score intensity normalization [7, 16]. Patient clinical characteristics for all cohorts are summarized in Table 1.

Table 2: Performance Comparison with SOTA Methods on Internal and External Test Datasets.

	Grade				IDH				1p/19q			
Methods	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)
Internal Test												
ASCP-Net [5]	81.24 ±3.86	82.80 ±3.47	77.07 ±6.99	72.19 ±7.86	78.28 ±4.41	74.21 ±2.80	73.22 ±7.58	78.39 ±6.30	78.85 ±2.80	70.68 ±1.63	61.10 ±8.16	76.84 ±3.47
MTTU-Net [3]	91.78 ±5.42	84.44 ±6.40	85.33 ±6.84	82.97 ±4.64	88.75 ±5.22	80.01 ±4.57	74.45 ±8.12	82.31 ±8.15	78.27 ±6.09	76.39 ±1.72	66.18 ±7.94	75.99 ±3.99
GliomaNet [23]	84.09 ±5.56	81.67 ±6.68	77.54 ±9.91	73.44 ±8.58	79.16 ±5.46	72.55 ±6.80	73.94 ±8.81	75.55 ±9.91	77.31 ±3.48	72.10 ±3.30	60.40 ±8.33	78.06 ±6.52
PSNet [25]	93.96 ±6.41	88.56 ±7.09	86.62 ±7.04	91.21 ±8.01	89.31 ±5.13	82.03 ±6.83	81.09 ±5.13	79.70 ±8.66	82.38 ±4.55	79.51 ±4.31	78.16 ±7.61	66.28 ±5.98
MFEFNet [29]	81.39 ±1.71	81.15 ±2.73	80.66 ±5.43	79.26 ±4.19	84.61 ±2.02	77.47 ±2.72	74.10 ±2.02	78.72 ±4.10	81.09 ±2.93	75.20 ±2.88	66.87 ±9.19	71.37 ±2.89
MDLNet [26]	92.83 ±4.62	87.76 ±5.30	86.88 ±4.12	84.64 ±11.15	89.34 ±1.16	84.65 ±2.15	83.29 ±6.70	83.91 ±3.52	84.05 ±2.58	74.49 ±6.38	72.12 ±7.22	72.10 ±8.76
CMTLNet [2]	91.38 ±4.50	86.93 ±4.02	83.11 ±5.34	79.37 ±4.38	89.28 ±4.68	82.46 ±3.18	78.71 ±4.96	83.81 ±3.29	76.69 ±4.28	70.38 ±3.33	62.96 ±4.25	76.08 ±3.28
MTAFNet [12]	85.26 ±1.64	84.37 ±1.92	84.29 ±4.21	73.87 ±8.36	84.93 ±1.54	80.48 ±2.24	74.06 ±5.77	82.64 ±3.80	77.81 ±4.38	76.23 ±4.32	68.62 ±5.76	72.66 ±5.17
MLNet [8]	83.73 ±1.26	81.38 ±1.88	87.69 ±2.84	79.64 ±3.81	80.52 ±1.43	82.93 ±1.82	76.58 ±4.89	82.92 ±2.71	78.27 ±4.09	75.98 ±2.88	62.06 ±10.87	78.28 ±2.50
Ours	**95.11** ±1.86	**89.14** ±1.34	**88.14** ±1.83	**91.55** ±2.47	**90.96** ±2.74	**85.45** ±1.14	**84.99** ±5.57	**85.47** ±2.79	**85.45** ±2.70	**80.73** ±2.98	**79.48** ±5.25	**79.89** ±4.28
External Test												
ASCP-Net [5]	77.83 ±7.62	74.66 ±8.43	73.53 ±12.77	64.15 ±16.22	72.77 ±8.53	70.60 ±5.20	61.66 ±14.68	64.84 ±15.45	64.62 ±7.78	80.99 ±8.11	33.08 ±6.88	87.84 ±7.81
MTTU-Net [3]	83.48 ±5.79	76.77 ±4.98	84.31 ±10.02	64.92 ±15.23	78.25 ±4.09	71.41 ±6.36	65.52 ±16.38	75.16 ±15.22	67.88 ±9.35	73.54 ±7.36	36.92 ±6.65	76.76 ±12.08
GliomaNet [23]	77.57 ±6.21	75.66 ±9.36	73.73 ±13.09	63.85 ±17.93	73.04 ±4.71	69.80 ±5.16	62.17 ±12.26	66.37 ±15.32	66.64 ±10.13	79.69 ±3.99	31.54 ±3.44	87.43 ±4.64
PSNet [25]	85.81 ±6.77	80.12 ±5.39	82.35 ±3.82	60.92 ±12.56	78.53 ±5.97	74.17 ±7.84	63.79 ±14.11	79.42 ±9.92	68.14 ±8.48	78.14 ±8.09	29.23 ±7.54	82.43 ±10.31
MFEFNet [29]	75.38 ±7.50	70.96 ±4.27	79.61 ±8.54	60.31 ±12.69	70.52 ±5.36	68.81 ±4.77	66.03 ±10.54	78.90 ±11.55	62.26 ±11.67	81.93 ±7.82	26.12 ±8.12	90.13 ±6.12
MDLNet [26]	82.32 ±7.23	77.96 ±6.90	86.43 ±4.35	61.54 ±16.06	75.65 ±5.32	74.77 ±6.45	62.07 ±10.42	78.86 ±15.58	66.74 ±11.17	83.11 ±5.39	36.15 ±6.44	89.86 ±6.34
CMTLNet [2]	82.95 ±6.02	76.89 ±7.33	81.18 ±9.94	70.15 ±15.69	78.65 ±6.58	72.46 ±4.89	71.03 ±12.82	66.81 ±13.28	68.97 ±8.59	68.20 ±8.62	35.38 ±4.08	71.08 ±8.66
MTAFNet [12]	70.38 ±6.78	72.93 ±6.27	75.10 ±13.84	65.85 ±16.34	70.35 ±4.91	71.69 ±9.47	64.31 ±18.15	62.86 ±18.34	67.05 ±8.91	77.33 ±6.52	28.16 ±8.45	85.40 ±7.10
MLNet [8]	79.44 ±4.34	79.76 ±4.47	83.53 ±11.48	62.46 ±18.96	74.19 ±7.07	67.07 ±2.73	61.03 ±16.92	76.95 ±12.90	63.20 ±2.26	81.30 ±10.88	27.42 ±3.82	89.32 ±4.96
Ours	**88.82** ±3.67	**81.32** ±3.34	**87.45** ±7.22	**71.69** ±11.65	**81.37** ±3.88	**74.90** ±3.96	**72.86** ±15.32	**80.06** ±10.15	**74.58** ±5.19	**85.22** ±4.88	**39.23** ±3.44	**91.89** ±5.41

3.2 Implementation Details

All experiments were implemented in PyTorch and conducted on NVIDIA RTX 4090 GPUs. The input size was fixed to $128 \times 128 \times 128$ after center cropping and random flipping for data augmentation. The DEC-Net was optimized using the AdamW optimizer with an initial learning rate of 1×10^{-4}. Batch size was set to 8, and the total number of epochs was 100. Samples with missing labels were excluded from model training, P-matrix estimation, and evaluation. For internal test, we conducted 5-fold cross-validation on the combined UCSF-PDGM and EGD cohort. The TCGA cohort was used exclusively as the external test cohort. Our five-fold cross-validation is site-stratified. To ensure a fair comparison, all experiments were evaluated under identical experimental settings.

3.3 Overall Performance on Internal and External Test Datasets

To comprehensively evaluate the effectiveness and generalization capability of the proposed method, we conducted experiments on both internal and external test datasets, comparing it with nine SOTA methods (ASCP-Net [5], MTTU-Net [3], GliomaNet [23], PSNet [25], MFEFNet [29], MDLNet [26], CMTLNet [2], MTAFNet [12], and MLNet [8]) on three key glioma classification tasks using Area Under the Curve (AUC) and Accuracy (ACC) with mean and standard deviation as evaluation metrics, with detailed results in Table 2. On the internal test dataset, DEC-Net consistently achieved the highest performance across all metrics. Notably, for Grade classification, our method reached an AUC of 95.11%, achieving a 1.22% improvement over the strongest baseline. In the more challenging IDH and 1p/19q subtyping tasks, DEC-Net also surpassed the second-best methods by 1.81% and 1.67% in AUC, respectively. The external validation on

Table 3: Module Ablation Study with Performance Metrics On Internal Dataset

Module			Grade				IDH				1p/19q			
DRModule	PMoE	D-Loss	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)
–	–	–	88.87 ±8.16	84.88 ±6.27	86.40 ±4.09	81.22 ±9.63	87.17 ±5.86	81.50 ±4.33	79.54 ±9.36	80.91 ±2.57	79.81 ±4.30	72.10 ±7.82	67.90 ±9.42	73.45 ±8.49
√	–	–	89.78 ±6.04	85.22 ±6.72	84.82 ±6.83	79.33 ±6.78	88.07 ±4.68	82.14 ±5.82	81.67 ±4.65	82.26 ±7.19	81.53 ±4.17	79.25 ±9.02	71.61 ±6.29	75.46 ±8.93
–	√	–	90.02 ±4.47	85.60 ±5.65	83.28 ±7.34	84.61 ±1.59	89.08 ±1.71	81.96 ±4.12	82.91 ±0.37	80.13 ±5.48	81.35 ±6.25	75.68 ±0.67	73.33 ±8.11	73.94 ±1.31
–	–	√	92.63 ±2.03	86.84 ±2.65	87.72 ±2.70	90.59 ±4.77	89.00 ±2.85	83.52 ±3.12	83.37 ±5.78	83.46 ±4.00	84.77 ±2.05	76.57 ±4.62	75.74 ±6.34	78.29 ±6.48
–	√	√	91.68 ±3.98	86.05 ±4.91	84.13 ±5.04	84.18 ±7.87	89.14 ±4.29	84.34 ±5.12	81.35 ±6.42	82.48 ±5.57	82.37 ±4.55	77.89 ±5.16	72.08 ±9.26	74.01 ±6.13
√	–	√	92.14 ±5.51	87.66 ±5.15	85.25 ±4.58	83.20 ±9.59	89.34 ±4.58	82.86 ±4.48	81.57 ±7.90	83.45 ±4.52	83.75 ±4.92	78.68 ±3.43	77.44 ±9.49	75.94 ±4.63
√	√	–	91.13 ±6.11	86.94 ±6.52	84.74 ±4.96	85.21 ±8.04	89.05 ±1.65	82.72 ±4.32	84.58 ±6.26	81.93 ±5.43	82.64 ±4.14	78.11 ±5.80	77.81 ±7.50	76.18 ±9.38
√	√	√	**95.11** ±1.86	**89.14** ±1.34	**88.14** ±1.83	**91.55** ±2.47	**90.96** ±2.74	**85.45** ±1.14	**84.99** ±5.57	**85.47** ±2.79	**85.45** ±2.70	**80.73** ±2.98	**79.48** ±5.25	**79.89** ±4.28

the TCGA dataset further underscores the robustness of our approach. While most SOTA methods experienced a sharp performance decline due to domain shift, DEC-Net retained leading results. Specifically, it achieved AUCs of 88.82% for Grade, 81.37% for IDH, and 74.58% for 1p/19q classification, outperforming the best competing model by 3.51%, 3.46%, and 8.13% in AUC, respectively.

3.4 Ablation Study

We conducted ablation studies on our internal dataset to validate the contribution of each proposed component (Table 3). The baseline model without any proposed modules achieved moderate performance. Introducing D-Loss improved prediction consistency across all tasks by modeling biological correlations between glioma grades and molecular subtypes, indicating the effectiveness of task-level relational regularization. Incorporating DRModule further guided the backbone to learn discriminative representations from tumor-relevant regions via knowledge distillation, while PMoE hierarchically integrated shared and task-specific experts to mitigate inter-task interference. In addition, the improvements of DRModule and PMoE are not strictly monotonic when combined with D-Loss separately, suggesting that these modules introduce complementary effects and may benefit different tasks or metrics to different extents. Ultimately, the full model integrating all three modules attained the best overall performance, outperforming all ablated variants on most metrics. These results confirm that DRModule, PMoE, and D-Loss are complementary and collectively enhance the model's generalization capability.

4 Conclusion

In this paper, we proposed the DEC-Net to address the challenges of task interference and representation learning in joint glioma grading and subtyping. By integrating the DRModule, PMoE module, and D-Loss, our network effectively captures discriminative multimodal representations, mitigates inter-task interference, and explicitly models the biological correlations between glioma grades and molecular subtypes. Extensive experiments on a large hybrid cohort of UCSF-PDGM and EGD, and external validation on the TCGA dataset, demonstrate that DEC-Net consistently outperforms SOTA methods. Our approach provides a powerful, non-invasive tool for comprehensive glioma profiling, holding significant potential to support precise clinical decision-making in neuro-oncology.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 62302119; the Scientific Research Fund of Hunan Provincial Education Department under Grant 23A0020; the Science and Technology Innovation Program of Hunan Province of China under Grant 2025RC3020; the Graduate Research and Innovation Project of Xinjiang Uygur Autonomous Region under Grants XJ2026G095 and XJ2026G047; the Central South University Innovation-Driven Research Program under Grant 2023CXQD018; the Guizhou Provincial Basic Research Program (Natural Science) under Grant QKH-ZK[2024]603; and the Guiyang City Science and Technology Plan Project under Grant [2024]2-18.

Disclosure of Interests. The authors declare no conflict of interest.

References

1. Calabrese, E., Villanueva-Meyer, J.E., Rudie, J.D., Rauschecker, A.M., Baid, U., Bakas, S., Cha, S., Mongan, J.T., Hess, C.P.: The university of california san francisco preoperative diffuse glioma mri dataset. Radiology: Artificial Intelligence **4**(6), e220058 (2022)
2. Chen, Q., Wang, L., Deng, Z., Wang, R., Wang, L., Jian, C., Zhu, Y.M.: Cooperative multi-task learning and interpretable image biomarkers for glioma grading and molecular subtyping. Medical Image Analysis **101**, 103435 (2025)
3. Cheng, J., Liu, J., Kuang, H., Wang, J.: A fully automated multimodal mri-based multi-task learning for glioma segmentation and idh genotyping. IEEE Transactions on Medical Imaging **41**(6), 1520–1532 (2022)
4. Cheng, J., Liu, J., Yue, H., Bai, H., Pan, Y., Wang, J.: 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 (2020)
5. Decuyper, M., Bonte, S., Deblaere, K., Van Holen, R.: Automated mri based pipeline for segmentation and prediction of grade, idh mutation and 1p19q codeletion in glioma. Computerized Medical Imaging and Graphics **88**, 101831 (2021)
6. Hu, Y., Cheng, J., Liu, F., Liu, J., Qian, Y.: Mvit-arnet: A lightweight mixed vision transformer with adaptive residual-enhanced network for 3d glioma mri segmentation. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2253–2258. IEEE (2025)
7. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)
8. Li, M., Jiang, D., Yun, J., Liu, R., Sun, Y., Li, G.: Cooperative multi-task learning and reliability assessment for glioma segmentation and idh genotyping. Pattern Recognition p. 112467 (2025)
9. Li, X., Liu, J., Kuang, H., Wang, Y., Wang, J.: Tetra-orientated mamba with t2-flair mismatch features for glioma segmentation, idh genotyping, and grading. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 554–564. Springer (2025)
10. Liu, Z., Zhou, Y., Xu, Y., Wang, Z.: Simplenet: A simple network for image anomaly detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20402–20411 (2023)

11. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H., Pfister, S.M., Reifenberger, G., et al.: The 2021 who classification of tumors of the central nervous system: a summary. Neuro-oncology **23**(8), 1231–1251 (2021)
12. Ma, M., Xi, Y., Li, R., Liu, X., Hu, Z., Li, Y., Wu, X., Hu, W., Kang, X., Yin, H., et al.: Mtaf-net: Multi-task learning for automated glioma segmentation and idh genotyping in multi-parametric mri. Biomedical Signal Processing and Control **112**, 108382 (2026)
13. Matsui, Y., Maruyama, T., Nitta, M., Saito, T., Tsuzuki, S., Tamura, M., Kusuda, K., Fukuya, Y., Asano, H., Kawamata, T., et al.: Prediction of lower-grade glioma molecular subtypes using deep learning. Journal of neuro-oncology **146**(2), 321–327 (2020)
14. Maynard, J., Okuchi, S., Wastling, S., Busaidi, A.A., Almossawi, O., Mbatha, W., Brandner, S., Jaunmuktane, Z., Koc, A.M., Mancini, L., et al.: World health organization grade ii/iii glioma molecular status: prediction by mri morphologic features and apparent diffusion coefficient. Radiology **296**(1), 111–121 (2020)
15. McKinnon, C., Nandhabalan, M., Murray, S.A., Plaha, P.: Glioblastoma: clinical presentation, diagnosis, and management. Bmj **374** (2021)
16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
17. Nalawade, S.S., Yu, F.F., Bangalore Yogananda, C.G., Murugesan, G.K., Shah, B.R., Pinho, M.C., Wagner, B.C., Xi, Y., Mickey, B., Patel, T.R., et al.: Brain tumor idh, 1p/19q, and mgmt molecular classification using mri-based deep learning: an initial study on the effect of motion and motion correction. Journal of Medical Imaging **9**(1), 016001–016001 (2022)
18. Naser, M.A., Deen, M.J.: Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images. Computers in biology and medicine **121**, 103758 (2020)
19. Ostrom, Q.T., Patil, N., Cioffi, G., Waite, K., Kruchko, C., Barnholtz-Sloan, J.S.: Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2013–2017. Neuro-oncology **22**(Supplement_1), iv1–iv96 (2020)
20. Pedano, N., Flanders, A.E., Scarpace, L., Mikkelsen, T., Eschbacher, J.M., Hermes, B., Sisneros, V., Barnholtz-Sloan, J., Ostrom, Q.: The cancer genome atlas low grade glioma collection (tcga-lgg). (No Title) (2016)
21. Scarpace, L., Mikkelsen, T., Cha, S., Rao, S., Tekchandani, S., Gutman, D., Saltz, J.H., Erickson, B.J., Pedano, N., Flanders, A.E., et al.: The cancer genome atlas glioblastoma multiforme collection (tcga-gbm). The Cancer Imaging Archive (2016)
22. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
23. Tripathi, P.C., Bag, S.: An attention-guided cnn framework for segmentation and grading of glioma using 3d mri scans. IEEE/ACM transactions on computational biology and bioinformatics **20**(3), 1890–1904 (2022)
24. van der Voort, S.R., Incekara, F., Wijnenga, M.M., Kapsas, G., Gahrmann, R., Schouten, J.W., Dubbink, H.J., Vincent, A.J., van den Bent, M.J., French, P.J., et al.: The erasmus glioma database (egd): Structural mri scans, who 2016 subtypes, and segmentations of 774 patients with glioma. Data in brief **37**, 107191 (2021)

25. 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.: Combined molecular subtyping, grading, and segmentation of glioma using multi-task deep learning. Neuro-oncology **25**(2), 279–289 (2023)
26. Wu, X., Zhang, S., Zhang, Z., He, Z., Xu, Z., Wang, W., Jin, Z., You, J., Guo, Y., Zhang, L., et al.: Biologically interpretable multi-task deep learning pipeline predicts molecular alterations, grade, and prognosis in glioma patients. npj Precision Oncology **8**(1), 181 (2024)
27. Yao, J., Liu, J., Cheng, J., Kuang, H., Wang, J.: M 3 ci-net: Multi-modal mri-based characteristics inspired network for idh genotyping. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1649–1654. IEEE (2023)
28. Yue, H., Kuang, H., Ma, L., Liu, J., Li, J., Cheng, J., Wang, J.: Visnet: A human visual system inspired lightweight dual-path network for medical images denoising. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 666–676. Springer (2025)
29. Zhang, J., Cao, J., Tang, F., Xie, T., Feng, Q., Huang, M.: Multi-level feature exploration and fusion network for prediction of idh status in gliomas from mri. IEEE Journal of Biomedical and Health Informatics **28**(1), 42–53 (2023)