Joint modeling histology and molecular markers for cancer classification

Xiaofei Wang [a,1], Hanyu Liu [b,1], Yupei Zhang [a], Boyang Zhao [b], Hao Duan [c],
Wanming Hu [d], Yonggao Mou [c], Stephen Price [a], Chao Li [a,b,e,f,*]

[a] Department of Clinical Neurosciences, University of Cambridge, UK
[b] School of Science and Engineering, University of Dundee, UK
[c] Department of Neurosurgery, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, China
[d] Department of Pathology, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, China
[e] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK
[f] School of Medicine, University of Dundee, UK

ARTICLE INFO

ABSTRACT

Cancers are characterized by remarkable heterogeneity and diverse prognosis. Accurate cancer classification is essential for patient stratification and clinical decision-making. Although digital pathology has been advancing cancer diagnosis and prognosis, the paradigm in cancer pathology has shifted from purely relying on histology features to incorporating molecular markers. There is an urgent need for digital pathology methods to meet the needs of the new paradigm. We introduce a novel digital pathology approach to jointly predict molecular markers and histology features and model their interactions for cancer classification. Firstly, to mitigate the challenge of cross-magnification information propagation, we propose a multi-scale disentangling module, enabling the extraction of multi-scale features from high-magnification (cellular-level) to low-magnification (tissue-level) whole slide images. Further, based on the multi-scale features, we propose an attention-based hierarchical multi-task multi-instance learning framework to simultaneously predict histology and molecular markers. Moreover, we propose a co-occurrence probability-based label correlation graph network to model the co-occurrence of molecular markers. Lastly, we design a cross-modal interaction module with the dynamic confidence constrain loss and a cross-modal gradient modulation strategy, to model the interactions of histology and molecular markers. Our experiments demonstrate that our method outperforms other state-of-the-art methods in classifying glioma, histology features and molecular markers. Our method promises to promote precise oncology with the potential to advance biomedical research and clinical applications. The code is available at github.

1. Introduction

Cancers are malignant tumors that are lethal to humans. Early and accurate diagnosis is crucial for managing cancers. Currently, pathology remains the gold standard for diagnosing cancers. However, traditional pathology is labour-intensive, time-consuming, and heavily dependent on the expertise of neuropathologists. Digital pathology, which utilizes automated algorithms to analyze tissue whole slide images (WSIs) (Lu et al., 2021), holds the promise of rapid diagnosis for timely and precise treatment.

Recently, deep learning (DL)-based digital pathology approaches have been successfully applied in diagnosing various cancers (Jose et al., 2023). These methods primarily focus on cancer diagnosis using histology features according to pathology diagnostic criteria. Alongside these efforts, the past decade has witnessed a surge in the discovery of molecular markers for cancer diagnosis, catalyzing a paradigm shift in the cancer diagnostic criteria from traditional histopathology towards molecular pathology. For instance, glioma is one of the most prevalent malignant primary tumors in adults, with a median overall survival of less than 14 months for high-grade glioma (Molinaro et al., 2019; Zhang et al., 2024b). The 2021 WHO Classification of Brain Tumours (Louis et al., 2021) has established several key molecular markers, such as isocitrate dehydrogenase (IDH) mutations, co-deletion of chromosome 1p/19q, and homozygous deletion (HOMDEL) of cyclin-dependent kinase inhibitor 2A/B (CDKN). According to the updated

criteria, glioblastoma is primarily diagnosed based on IDH mutations, whereas it was previously based purely on histology features such as necrosis and microvascular proliferation (NMP). In addition, a new class of high grade astrocytoma has been defined by IDH and CDKN. This prominent paradigm shift has revolutionized the clinical pathway for cancers and offers new opportunities for developing novel digital pathology approaches.

Integrating molecular markers and histology into clinical diagnosis still faces several practical challenges. Firstly, the methods of assessing molecular markers, such as gene sequencing and immunostaining, are often time-consuming and costly. Secondly, histology assessment and molecular markers follow separate diagnostic workflows, hindering timely and effective integration for clinical diagnosis. DL approaches promise to tackle the above challenges with unique advantages: (1) Predicting molecular markers from WSIs: Mounting research indicates that histology features are associated with molecular alterations, making it feasible to predict molecular markers directly from WSIs (Wang et al., 2024a). (2) Integrating diagnostic flows: DL-based approaches can effectively integrate the diagnostic flows of molecular markers and histology features. Additionally, they can model the interactions between data types, thereby advancing cancer diagnosis under the latest diagnostic paradigm. There is a pressing need for developing DL approaches to jointly predict molecular markers and histology while capturing their interactions. This integrated approach holds the potential to streamline and enhance the accuracy of cancer diagnosis.

In this paper, we introduce a novel digital pathology approach, namely multi-scale multi-task modeling for cancer classification (M3C2), aligning with the emerging paradigm of molecular pathology for cancer diagnosis. Previous approaches have attempted to integrate histology and genomics for tumour diagnosis (Xing et al., 2022). For example, Ding et al. (2023) developed a multi-modal transformer with unsupervised pretraining to integrate pathology and genomics for predicting colon cancer survival. Despite success, most existing methods only use molecular markers as supplementary input. They are incapable of simultaneously predicting histology and molecular markers and further modeling their interaction, which limits their clinical relevance under current diagnostic scheme. To address this limitation and align with the updated clinical diagnostic pathway, we leverage a novel hierarchical multi-task framework based on vision transformer (Dosovitskiy et al., 2020). Our approach features two partially weight-sharing components that jointly predict histology and molecular markers, effectively modeling their interactions and enhancing clinical relevance.

First, we design our model to extract informative features from multiple WSI magnifications. In clinical practice, pathology diagnosis involves inspecting tissue sections under different magnification, e.g., from 20X ($0.25 \ \mu m \ px^{-1}$ at cellular-level) to 10X ($1 \ \mu m \ px^{-1}$ at tissue-level) (Schmitz et al., 2021). Hence, we design a multi-scale disentangling module to capture the crucial WSI features required for cancer diagnosis. Specifically, we employ a novel disentanglement loss to efficiently extract features for both histology and molecular markers.

Further, we focus on modeling relationship among different molecular markers. It is known that molecular markers are inherently associated due to underlying cancer evolution and biology. Consequently, multiple molecular markers are often required for accurate cancer classification according to guidelines. To mirror real-world scenarios, we formulate the prediction of multiple molecular markers as a multi-label classification (MLC) task. Although previous MLC methods have effectively captured label correlations (Li et al., 2022b), existing approaches may overlook the co-occurrence and intrinsic associations of molecular markers during prediction (Zhang et al., 2023). To address this gap, we propose a co-occurrence probability-based label-correlation graph (CPLC-Graph) network to model the co-occurrence and relationships of molecular markers.

Lastly, we model the interaction between molecular markers and histology predictions. Specifically, we introduce a cross-modal interaction module to capture the interplay between molecular markers and

histology features, such as IDH mutation and NMP, both crucial for diagnosing glioblastoma. In this module, we firstly design a dynamic confidence constrain (DCC) loss, guiding the model to focus on similar WSI regions for both tasks. Beyond loss-level interaction, we also develop the cross-modal gradient modulation (CMG-Modu) learning strategy to coordinate the training process of histology and molecular marker predictions. To our best knowledge, this is the first attempt to classify cancer via modeling the interaction of histology and molecular markers predictions.

Our main contributions are summarized as follows. (1) We propose a multi-task multi-instance learning framework to jointly predict histology and molecular markers and classify glioma, reflecting the up-to-date diagnosis paradigm. (2) We devise a multi-scale disentangling module to generate efficient multi-scale features for both histology and molecular marker prediction. A CPLC-Graph network is proposed to model the relationship of multiple molecular markers. (3) We design a DCC loss and a CMG-Modu training strategy to coordinate the cross-modal interaction between histology and molecular markers for glioma classification. (4) We perform extensive experiments to validate our performance in several tasks of glioma classification, molecular markers and histology prediction.

2. Related work

In traditional pathology, WSI inspection by pathologists remains the gold standard for clinical diagnosis (Deng et al., 2020). However, the conventional diagnostic process is labor-intensive and time-consuming (Deng et al., 2020), with the shortage of experienced pathologists further exacerbating the challenge. DL-based methods (Martínez-Fernandez et al., 2023), which employ biologically-inspired neural networks to utilize the rich information contained in WSIs. are advancing computer-aided pathology diagnosis. In this section, we firstly review DL-based digital pathology methods for cancer diagnosis, and then focus on methods specific to glioma diagnosis, and finally review methods that integrate histology and molecular markers for precision oncology.

2.1. Digital pathology for cancer diagnosis

DL-based digital pathology methods are generally categorized into low (cell)-level modeling and high (tissue)-level modeling. These approaches address downstream tasks such as cell segmentation (Graham et al., 2019) and tissue phenotyping (Bhattacharyya et al., 2024), respectively. Specifically, cell segmentation models primarily focus on assigning pixel-level labels for cells and nuclei (Graham et al., 2019), further used as biomarkers for disease diagnosis. For instance, Graham et al. (2019) proposed Hover-Net, which performs simultaneous nuclear segmentation and classification by leveraging the distances of nuclear pixels to the mass centre. Similarly, Ren et al. (2021) developed a DL-based pipeline to generate an embedded map profiling cell composition, further utilized to extract texture patterns for tumour and immune cells. In contrast to cell-level modeling, tissue-level modeling is more efficient in capturing global phenotype features, providing a more comprehensive understanding of tissue architecture and pathology that is crucial for accurate and effective clinical assessments. Particularly, recent DL-based approaches have significantly improved computer-aided classification of various diseases (Li et al., 2022a), including membranous nephropathy (Hao et al., 2023), interstitial lung disease (Uegami et al., 2022).

DL-based tissue-level modeling has been widely researched in cancer diagnosis, promising to enhance the accuracy and efficiency of cancer classification (Yang et al., 2022; Kanavati et al., 2020). For instance, Lin et al. (2018) devised ScanNet, a fast and dense scanning framework utilizing asynchronous sample prefetching and hard negative mining for efficient breast cancer detection. Additionally, Yang
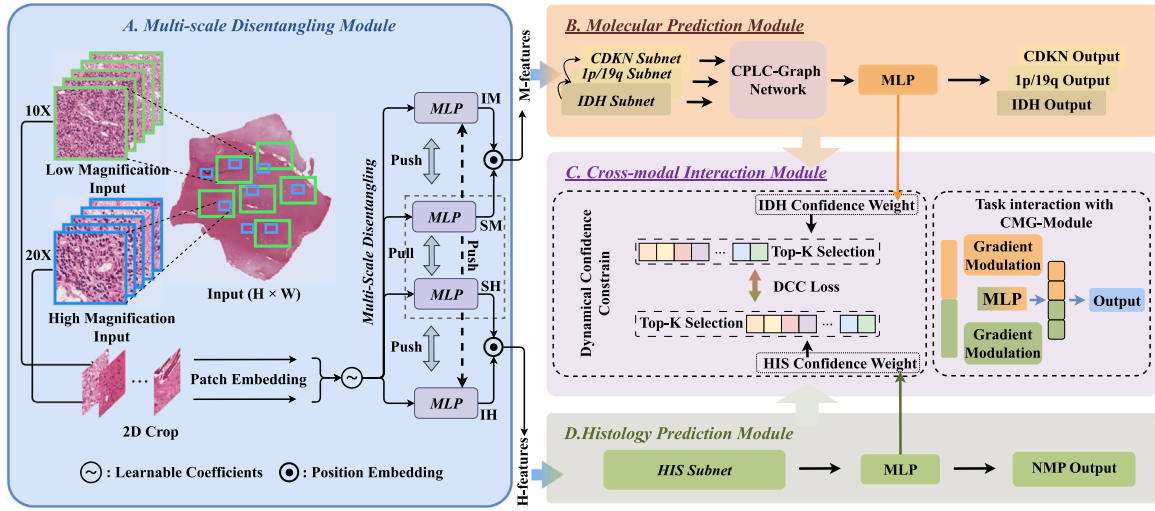
Fig. 1. Framework of our M3C2 method, including (A) multi-scale disentangling module, (B) molecular prediction module, (C) cross-module interaction module and (D) histology prediction module. Note that IM, SM, IH and SH denote independent molecular features, shared molecular features, independent histology features and shared histology features, respectively.

et al. (2022) proposed FA-MSCN, a feature aligned multi-scale convolutional network, to achieve liver tumour diagnosis. FA-MSCN integrates features from multiple magnifications to improve the performance by referencing more neighboring information. Similarly, Kanavati et al. (2020) designed an end-to-end, high-generalizable, weakly supervised deep convolutional network, which achieved great performance in lung cancer classification. Despite the success in many cancers, DL-based algorithms still face significant challenges in brain tumors, particularly glioma, due to its remarkable heterogeneity, complex diagnostic criteria and limited samples (Perry and Wesseling, 2016; Louis et al., 2021).

2.2. Automatic glioma grading with WSIs

In recent years, several models (Zhang et al., 2022b; Pei et al., 2021; Lv et al., 2024) have been proposed to classify glioma based on WSIs. For instance, Zhang et al. (2022b) designed a mutual contrastive low-rank learning (MCL) scheme, where formalin-fixed paraffin-embedded (FFPE)-based and frozen-based WSIs are integrated for better glioma grading performance. Additionally, Pei et al. (2021) proposed a novel DL-based classification method that fuses genomic features with cellularity features for glioma prediction. However, most existing DL methods are based on the obsolete glioma taxonomy criteria (Louis et al., 2007, 2016), which only focuses on histology features, such as NMP. Therefore, it is in demand to develop DL-based methods consistent with the latest glioma diagnosis criteria, which jointly consider histology and molecular markers in the taxonomy pipeline.

Most recently, a few DL-based methods (Hollon et al., 2023; Nasrallah et al., 2023; Lv et al., 2024) have achieved automatic glioma diagnosis using the up-to-date criteria. For example, Hollon et al. (2023) proposed a hierarchical vision transformer-based method to separately predict molecular markers and histology features, which are then combined for final diagnosis. Similarly, Lv et al. (2024) devised a ResNet-50-based DL structure with a novel patch selection strategy for classifying glioma according to the latest criteria. Despite their potentials, significant limitations remain in existing methods: **(i)** Current methods are incapable of achieving efficient interaction between the histology features and molecular markers for clinically interpretable diagnosis; **(ii)** The diagnosis scheme used in existing works are incomplete or inconsistent with the latest criteria. For instance, in Hollon et al. (2023), grade 4 astrocytoma, a newly defined class in WHO 2021 criteria, was ignored in the training process. **(iii)** Most models were trained with a single type of WSIs, i.e., either FFPE or frozen tissue

sections, whereas in real-world practice, both FFPE- and frozen-based WSIs are commonly used for diagnosis in different clinical scenarios. To address these gaps, the proposed M3C2 achieves explicit interaction between histology features and molecular markers aligned with the WHO 2021 criteria and trained on both FFPE and frozen sections.

2.3. Integration of histology and molecular markers

Multi-modal learning for precision oncology. Recently, joint modeling of multi-modal data of histology and molecular markers has achieved progress in precision oncology, advancing AI towards practical clinical application in various cancers (Zhang et al., 2024a; Wei et al., 2023). Multi-modal algorithms are generally categorized into three types: early, intermediate, and late fusion. Early fusion integrates raw data or extracted features before inputting them into the proposed algorithms. Intermediate fusion (Zhou and Chen, 2023; Jaume et al., 2024) modulates multi-modal feature embeddings at specific fusion modules or different layers within the model. Late fusion (Shao et al., 2019) involves modality-specific models that aggregate predictions from each modality for the final classification. For example, Zhou and Chen (2023) proposed an intermediate fusion framework (CMAT), which explores intrinsic associations between histopathology images and gene expression profiles through a cross-modal attention module and a modality alignment and fusion module. Similarly, Jaume et al. (2024) developed a late fusion framework named SURVPATH, a memory-efficient multimodal transformer that integrates gene pathways and histopathology images for pan-cancer survival prediction.

For gliomas, several multi-modal data fusion methods have been proposed (Chen et al., 2020; Qiu et al., 2024). Specifically, Chen et al. (2020) introduced the Pathomic Fusion method that integrates gene expression and WSI for jointly predicting survival and glioma classification. Qiu et al. (2024) devised a disentangled multi-modal framework that can handle both complete and incomplete samples for glioma diagnosis and prognosis. Despite the effectiveness, most current multi-modal methods face practical limitations in clinical scenarios, as they rely on the input of molecular information, which is expensive and time-consuming to obtain. This dependence hampers their applicability and widespread adoption in clinical practice.

To tackle this challenge, Xing et al. (2022) proposed a novel discrepancy and gradient-guided modality distillation framework. In this approach, WSIs and gene expressions are combined during training, while the branch for gene expressions is distilled during testing. However,

despite this innovative approach, modality distillation-based multi-modal learning methods are fall short in generating molecular markers. This limitation renders them impractical for clinical use according to the latest diagnosis criteria.

Multi-task learning for precision oncology Multi-task learning shows promise to improve on existing methods, which can predict both histology features and molecular markers only using WSIs as input. For instance, Coudray et al. (2018) proposed an Inception-v3 (Szegedy et al., 2016)-based DL method to separately predict both classification and mutation from non–small cell lung cancer histopathology images. However, most existing methods perform multiple tasks separately, without explicit interaction between histology and molecular marker predictions. In comparison, we proposed a novel multi-task learning method with efficient modality interaction via the DCC loss and the specially designed gradient modulation strategy. This approach facilitates explicit task interplay, enhancing the predictive power and clinical applicability of the model by integrating histology and molecular marker more effectively.

2.4. Differences from the conference paper

This paper is an extended version of our conference paper (Wang et al., 2023) (Nominated for Best Paper Award) with substantial improvements by: (1) adding a thorough literature review and an extensive discussion. (2) fulfilling the details of the proposed method and improving the original structure as follows. We first propose a multi-scale disentangling module to extract multi-magnification WSI features for both histology and molecular marker predictions. We design a CMG-Modu strategy to harmonize the training process of multiple tasks and achieve cross-modal task interplay. In addition, an attention mechanism is further involved in our framework to extract more relevant features. (3) Consequently, the performance of our method has been improved in all tasks of glioma classification (by 5.6% in accuracy), molecular markers (e.g., by 4.3% in accuracy of IDH) and histology (by 4.3% in accuracy) prediction. (4) We comprehensively validate the model with diverse experimental settings. First, we perform additional ablation experiments to evaluate the newly proposed modules. Secondly, we add an external validation set of 753 WSIs, verifying the generalizability of our method. Thirdly, we evaluate the model performance with varied experimental settings on input magnifications and incorporation of auxiliary tasks. Finally we compare our model with more SOTA methods.

3. Methodology

3.1. Framework

According to the latest glioma diagnosis criteria using both histology and molecular information, it is therefore intuitive to jointly learn the multiple tasks of histology and molecular markers prediction, as well as the final glioma classification, in a unified framework. In this paper, we propose a novel M3C2 method to simultaneously handle these tasks. The framework of M3C2 is shown in Fig. 1. As can be seen, given the cropped multi-scale patches $\{\mathbf{X}_i^h\}_{i=1}^N$ and $\{\mathbf{X}_j^l\}_{j=1}^N \in \mathbb{R}^{N \times H \times W \times 3}$ (with patch number N, height H, width W and 3 channels of RGB) of 20X and 10X WSI magnification as the input, the model can predict (1) molecular markers, including IDH mutation $\hat{l}_{idh} \in \mathbb{R}^2$, 1p/19q co-deletion $\hat{l}_{1p/19q} \in \mathbb{R}^2$ and CDKN HOMDEL $\hat{l}_{cdkn} \in \mathbb{R}^2$, (2) existence of NMP $\hat{l}_{nmp} \in \mathbb{R}^2$ and (3) final diagnosis of glioma $\hat{l}_{glio} \in \mathbb{R}^4$. Note that in our 4-class glioma classification task, class 0 to 3 refer to grade 4 GBM, high grade astrocytoma, low grade astrocytoma and oligodendroglioma, respectively.

The structure of M3C2 consists of 4 modules, including multi-scale disentangling module, molecular prediction module, histology prediction module and cross-modal interaction module. Detailed structures are described as follows.

3.2. Multi-scale disentangling module

Clinical cancer diagnosis typically relies on WSI scanning under multiple magnifications (from 20X to 10X), showing morphological features from cellular-level to tissue-level for precision oncology. Prior biological knowledge (Phan and Rizzoli, 2020) suggests that the molecular markers and histology features are associated with varied multi-scale features. For instance, the molecular markers could be better reflected by high-magnification features with cellular-level details, while histology features are better captured at the low magnification features at the tissue level, providing information on tissue structures. Hence, we propose a novel multi-scale disentangling module. By disentangling the independent and shared features for molecular and histology tasks, we ensure that the embeddings extracted are more relevant and useful for their respective downstream tasks, rather than relying on a single set of undifferentiated features. This module can generate representative embeddings for both histology and molecular prediction tasks via the interaction of multi-magnification WSI features.

Given the multi-scale patches $\{\mathbf{X}_i^l\}_{i=1}^N$ and $\{\mathbf{X}_j^h\}_{j=1}^N$ as input, we firstly extract the patch embeddings $\mathbf{F}^l \in \mathbb{R}^{N \times K}$ and $\mathbf{F}^h \in \mathbb{R}^{N \times K}$ (with embedding dimension K) (Mao et al., 2023) using a pre-trained ResNet-50 (He et al., 2016). The multi-scale extracted embeddings are then fed into the multi-scale disentangling module, a multi-layer perceptron (MLP)-based network for producing molecular- and histology-oriented features, as shown in Fig. 1. Specifically, the embeddings $\mathbf{F}^l$ and $\mathbf{F}^h$ are first weighted with learnable coefficients and then concatenated together and finally processed with four separate MLPs to generate both shared and independent features for the molecular ($\mathbf{S}_m$ and $\mathbf{I}_m$ for shared and independent features) and histology ($\mathbf{S}_h$ and $\mathbf{I}_h$) prediction tasks. Moreover, we further proposed a cross-modal disentanglement loss to minimize the disparity among shared representations while maximize that among independent representations. Formally, the disentanglement loss is defined as follows.

$$\mathcal{L}_{\text{disent}} = \frac{\|\mathbf{S}_m - \mathbf{S}_h\|_2}{\|\mathbf{I}_m - \mathbf{I}_h\|_2 + \|\mathbf{I}_m - \mathbf{U}_m\|_2 + \|\mathbf{I}_h - \mathbf{U}_h\|_2}, \quad (1)$$

3.3. Attention-based hierarchical multi-task multi-instance learning

To further extract molecular and histology information for the multi-modal prediction tasks, we propose an attention-based hierarchical multi-task multi-instance learning (AHMT-MIL) framework. Different from methods using one (Zhang et al., 2022b) or several (Xing et al., 2022) representative patches per slide, AHMT-MIL framework can extract information from N=2,500 patches per WSI on both magnifications via utilizing the MIL learning paradigm. Of note, for WSIs with patch number $< N$ and $> N$, we adopt the biological repeat and 2D average sampling strategy for dimension alignment, respectively. Specifically, developed on the top of vision transformer (Dosovitskiy et al., 2020), the AHMT-MIL framework consists of two modules for the histology and molecular prediction tasks. Fig. 2 illustrates the structure of these two modules, which are further detailed as follows.

Histology prediction module. As illustrated in Fig. 2, the histology prediction module takes the H-features $\mathbf{F}_{his}$ as input and outputs the prediction of the histology features of NMP. Specifically, $\mathbf{F}_{his}$ is firstly processed through three cascaded transformer blocks to extract NMP features, which is then refined with the attention mechanism (Ilse et al., 2018) for more distinctive histology features. Details of the attention mechanism are described below.

Given a bag of N embeddings $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_N\} \in \mathbb{R}^{N \times K}$, the refined embeddings $\mathbf{z} \in \mathbb{R}^{1 \times K}$ can be generated via an MIL pooling as follows:

$$\mathbf{z} = \sum_{n=1}^N a_n \mathbf{f}_n, where \ a_n = \frac{\exp\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{f}_n^\top)\}}{\sum_{j=1}^N \exp\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{f}_j^\top)\}}. \quad (2)$$

In Eq. (2), $\mathbf{w} \in \mathbb{R}^{L \times 1}$ and $\mathbf{V} \in \mathbb{R}^{L \times K}$ are learnable parameters, while $\tanh(\cdot)$ denotes element-wise hyperbolic tangent function. Finally, the
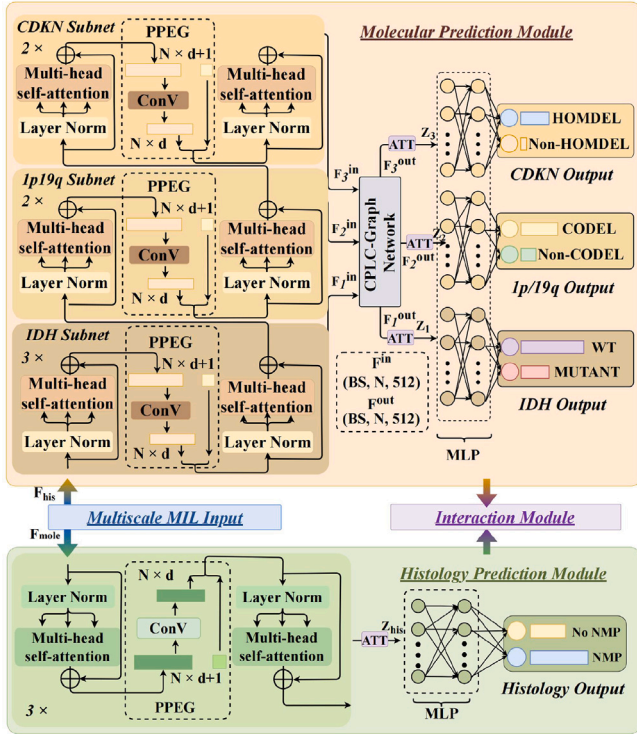
Fig. 2. Detailed structure of the proposed molecular prediction module (above) and the histology prediction module (below).
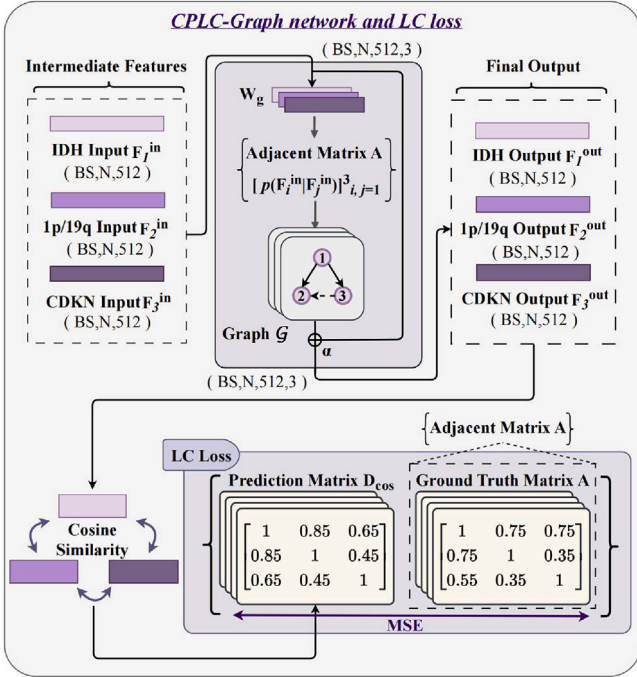

Fig. 3. Detailed structure of the proposed CPLC-Graph network and the LC loss.

refined embeddings $\mathbf{z}$ is fed to the MLP for the histology prediction task.

Molecular prediction module. Taking M-features $\mathbf{F}_{\text{mole}}$ from the multiscale disentangling module as input, the molecular prediction module is developed for the multi-label classification (Liu et al., 2021) of the three molecular markers. As shown in Fig. 2, the molecular prediction

module consists of three sequential subnets, extracting features for IDH, 1p/19q and CDKN markers, respectively. Specifically, each subnet is composed of several cascaded transformer blocks, with the block number of 3, 2 and 2 for IDH, 1p/19q and CDKN, respectively.[2] In addition, similar to histology prediction, attention mechanism is also utilized in the MIL process of for the molecular prediction task, allowing for the extraction of representative patches for these three markers. Finally, we further devised a graph-based network to model the intrinsic correlation of the three molecular markers, which is introduced as follows.

3.4. Co-occurrence probability-based, label-correlation graph

In classifying molecular markers, including IDH, 1p/19q and CDKN, current MLC methods based on label correlation might overlook the co-occurrence of these labels. To address this, we propose a co-occurrence probability-based label-correlation graph (CPLC-Graph) network and a label correlation (LC) loss, which are designed for intra-omic modeling of the co-occurrence probabilities among the three markers.

(1) CPLC-Graph network: As shown in Fig. 3, CPLC-Graph is defined as $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ indicates the nodes and $\mathbf{E}$ represents the edges. Given the intermediate features in predicting the three molecular markers $\mathbf{F}^{\text{in}} = [\mathbf{F}_i^{\text{in}}]_{i=1}^3 \in \mathbb{R}^{3 \times N \times K}$ as input nodes, we then construct a co-occurrence probability based correlation matrix $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ to capture the relationships among these node features. Note that $\mathbf{A}$ is a shared value calculated using our internal large-scale TCGA dataset. Additionally, a weight matrix $\mathbf{W}_g \in \mathbb{R}^{K \times K}$ is utilized to update $\mathbf{F}^{\text{in}}$. The output nodes $\mathbf{F}^{\text{mid}} \in \mathbb{R}^{3 \times N \times K}$ are then computed using a single layer of a graph convolutional network, formulated as follows.

$$\mathbf{F}^{\text{mid}} = \delta(\mathbf{A}\mathbf{F}^{\text{in}}\mathbf{W}_g), \tag{3}$$

$$\text{where} \mathbf{A} = [A_i^j]_{i,j=1}^3, \quad A_i^j = \frac{1}{2}\left(p(\mathbf{F}_i^{\text{in}}|\mathbf{F}_j^{\text{in}}) + p(\mathbf{F}_j^{\text{in}}|\mathbf{F}_i^{\text{in}})\right). \tag{4}$$

In the above equation, $\delta(\cdot)$ represents an activation function and $p(\mathbf{F}_i^{\text{in}}|\mathbf{F}_j^{\text{in}})$ indicates the probability of the status of ith marker given the status of jth marker. Additionally, residual structure is employed to produce the final output $\mathbf{F}^{\text{out}}$ of CPLC-Graph network, which is defined as $\mathbf{F}^{\text{out}} = \alpha\mathbf{F}^{\text{mid}} + (1 - \alpha)\mathbf{F}^{\text{in}}$, where α is a graph balancing hyper-parameter.

(2) LC loss: To fully leverage the co-occurrence probability of different molecular markers, we introduce the LC loss, which constrains the similarity between any two output molecular markers, $\mathbf{F}_i^{\text{out}}$ and $\mathbf{F}_j^{\text{out}}$, to align with their corresponding co-occurrence probability A_i^j. This approach aims to better capture the intrinsic associations among various genomic alterations. Formally, the LC loss is defined as follows:

$$\mathcal{L}_{\text{LC}} = \mathcal{MSE}(\mathbf{A}, \mathbf{D}_{\text{cos}}), \tag{5}$$

$$\text{where } \mathbf{D}_{\text{cos}} = [D_{cos}^{i,j}]_{i,j=1}^3, \quad D_{cos}^{i,j} = \frac{(\mathbf{F}_i^{\text{out}})^{\top}\mathbf{F}_j^{\text{out}}}{\left\|\mathbf{F}_i^{\text{out}}\right\| \left\|\mathbf{F}_j^{\text{out}}\right\|}. \tag{6}$$

In (5), $\mathcal{MSE}$ is the function of mean square error, while $D_{cos}^{i,j}$ denotes the cosine similarity of features $\mathbf{F}_i^{\text{out}}$ and $\mathbf{F}_j^{\text{out}}$.

3.5. Cross-modal interaction module

To achieve efficient interaction between histology and molecular markers prediction tasks, we further proposed the cross-modal interaction module with the specially designed CMG-Modu training strategy and dynamic confidence constrain loss, which are detailed as follows.

---

[2] Of note, IDH is a relative upstream genetic event, so that it is in the forefront of the feature extracting sequence.
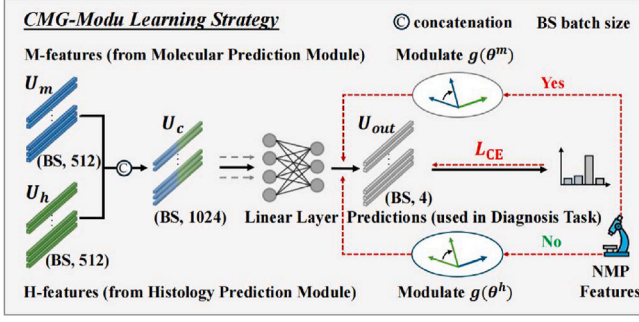
Fig. 4. Illustration of the CMG-Modu learning strategy.

3.5.1. CMG-Modu learning strategy

According WHO 2021 criteria, the molecular markers and histology features are closely integrated for the glioma diagnosis. For instance, patients previously diagnosed as high grade based on *histology features of NMP* are consistently considered as high grade in the updated criteria. In contrast, the grade of patients without NMP features turns to be determined by *molecular markers* in the latest diagnostic scheme (Berger et al., 2022). Inspired by this clinical prior, we propose the cross-modal gradient modulation (CMG-Modu) learning strategy, as shown in Fig. 4, to facilitate the interplay between the molecular markers and histology prediction tasks. Generally, the optimization process is dynamically calibrated by the gradient of the dominant task, i.e., either predicting molecular markers or histology features of NMP. Detailed design of the CMG-Modu is as follows.

As illustrated in Fig. 4, we firstly concatenate the extracted histology features $\mathbf{U}_h$ and molecular features $\mathbf{U}_m$ to generate the multi-modal features $\mathbf{U}_c$, which is further fed to a fully connected layer for the glioma classification. During the training process, the gradients in optimizing the multi-modal prediction tasks are modulated by the dominant task in terms of the labels of NMP, thus calibrating the prediction bias. Specifically, we modulate the gradients of molecular markers prediction task ($g(\theta^h)$) using those of histology prediction task when the histology label within a minibatch[3] is NMP positive, and vice versa. Formally, the modulated gradients for each task is defined as follows:

$$\begin{cases} \tilde{g}(\theta^h) = \rho(\Phi(g(\theta^h), g(\theta^m))), & NMP\ negative, \\ \tilde{g}(\theta^m) = \rho(\Phi(g(\theta^m), g(\theta^h))), & NMP\ positive, \end{cases} \quad (7)$$

where $\rho(\cdot)$ is a scaling function to fix gradient scale (Wang et al., 2024b), while $\Phi(\mathbf{a}, \mathbf{b})$ represents the projection of vector $\mathbf{a}$ onto the direction perpendicular to vector $\mathbf{b}$, which can be calculated as:

$$\text{Proj}_{\perp b}\mathbf{a} = \mathbf{a} - \text{Proj}_b\mathbf{a} \quad (8)$$

where $\text{Proj}_b\mathbf{a}$ is the projection of vector $\mathbf{a}$ onto vector $\mathbf{b}$:

$$\text{Proj}_b\mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2}\mathbf{b} \quad (9)$$

Moreover, the detailed model learning process with CGM-Modu strategy can be found in Algorithm 1. As can be seen, the cross-modal optimization reflecting the up-to-date diagnostic criteria can relieve the modality-specific bias and enhances the final glioma classification performance, as validated in Section 4.4.

---

[3] Note that the histology label within a minibatch is obtained by majority voting.

Algorithm 1 Training process with CMG-Modu strategy

Input: Paired histopathology features and molecular markers features $\{\mathbf{U}_h, \mathbf{U}_m\}$ for up-to-date brain tumor diagnosis; Model parameters $\Theta = \{\theta^h, \theta^m\}$; Multi-modal classifier $\mathcal{D}$; Task label Y; Histopathology-based grading label T.

Output: The optimal model parameters Θ^*
1: **while** Θ doesn't reach convergence **do**
2:    **for** each $\mathbf{u}_h$ and $\mathbf{u}_m \in \{\mathbf{U}_h, \mathbf{U}_m\}$ **do**
3:       Concatenate $\mathbf{u}_h$ and $\mathbf{u}_m$ to get features $\mathbf{u}_c$
4:       Minimize $\mathcal{L}_{\text{CE}}(Y, \mathcal{D}(\mathbf{u}_c; \theta^h, \theta^m)$
5:       **if** T belongs to high grade **then**
6:          Calibrate molecular gradients using Eq. (7)
7:       **else**
8:          Calibrate histopathology gradients using Eq. (7)
9:       **end if**
10:   **end for**
11: **end while**
12: The $\mathcal{D}$ is well-trained by CMG-Modu.

3.5.2. Dynamic confidence constrain

Additionally, we have further developed a dynamic confidence constrain (DCC) strategy to model the interaction between molecular markers and histological features. The DCC incorporates known clinical co-occurrence relationships between molecular and histology features, such as IDH wildtype and NMP. IDH wildtype is used as a major diagnostic marker for glioblastoma in the current clinical paradigm, while historically, the diagnosis of glioblastoma is purely based on NMP. In addition, studies have shown that over 95% of patients with NMP are IDH wildtype (Alzial et al., 2022), highlighting the strong co-occurrence relationship between the two markers. This relationship underpins the potential performance benefits of leveraging this co-occurrence for molecular prediction. This design not only enhances the interpretability of molecular predictions (Fig. 6) but also improves prediction performance. Specifically, inspired by Zhang et al. (2022a), the confidence weights for IDH wildtype $\mathbf{C}_{\text{wt}} = [c_{\text{wt}}^n]_{n=1}^N \in \mathbb{R}^{N\times 1}$ can be defined as:

$$\mathbf{C}_{\text{wt}} = (\mathbf{F}_1^{\text{out}} \cdot \hat{\mathbf{z}}_1) \otimes \mathbf{w}_{\text{wt}}, \text{ where } \hat{\mathbf{z}}_1 \in \mathbb{R}^{N\times K} \xleftarrow{\text{repeat}} \mathbf{z}_1 \in \mathbb{R}^{1\times K} \quad (10)$$

In (10), $\cdot$ denotes the dot product, $\otimes$ represents Hadamard product, $\mathbf{z}_1$ is refined IDH embeddings with attention mechanism and $\mathbf{w}_{\text{wt}}$ is the weights in the MLP for IDH wildtype. We then reorder $[c_{\text{wt}}^n]_{n=1}^N$ to $[\hat{c}_{\text{wt}}^n]_{n=1}^N$ based on their values. Similarly, we obtain $\mathbf{C}_{\text{nmp}} = [\hat{c}_{\text{nmp}}^n]_{n=1}^N$ for NMP confidence weights.

Using ordered confidence weights, we constrain the prediction networks of histology and molecular markers to concentrate on the WSI areas important for both predictions, thereby modeling cross-modal interactions. Specifically, the confidence constraint is achieved by a novel DCC loss, which focuses on the top M important patches for both predictions. Formally, the DCC loss in pth training epoch is defined as:

$$\mathcal{L}_{\text{DCC}} = \frac{1}{2M_p} \sum_{m=1}^{M_p} \left(S(\hat{c}_{wt}^m, \mathbf{C}_{\text{nmp}}) + S(\hat{c}_{\text{nmp}}^m, \mathbf{C}_{\text{wt}}) \right), \quad (11)$$

where $S(\hat{c}_{wt}^m, \mathbf{C}_{\text{nmp}})$ is the indicator function that takes the value 1 when the kth important patches of IDH widetype is among the top M_p important patches for NMP, and vice versa. In addition, to enhance the learning process with DCC loss, we employ a curriculum-learning-based training strategy dynamically focusing on hard-to-learn patches, identified as those with higher decision importance weight, as patches with lower confidence weight, such as those with fewer nuclei, are typically easier to learn for both tasks. Hence, M_p is further defined as

$$M_p = M_0 \beta^{\lfloor \frac{p}{p_0} \rfloor}. \quad (12)$$

Table 1
Mean values in terms of percentage for glioma classification metrics by our and other methods over the internal and external validation datasets.

Approaches	Attributes			Evaluation on internal dataset					Evaluation on external dataset				
	MTL[a]	CMM[b]	For glioma	Acc.	Sen.	Spec.	AUC	F$_1$-score	Acc.	Sen.	Spe.	AUC	F$_1$-score
VGG				71.1	71.1	85.2	84.9	71.1	89.4	89.4	81.6	91.8	89.4
Inception				60.2	60.2	84.3	81.4	60.2	67.2	67.2	99.5	84.5	67.2
DenseNet				63.2	63.2	**91.6**	84.8	63.2	68.3	68.3	99.3	73.7	68.3
AlexNet				53.3	53.3	89.0	71.0	53.3	80.2	80.2	99.6	89.9	80.2
ResNet				65.5	65.5	52.7	80.4	65.5	55.3	55.3	63.5	94.6	55.3
ABMIL			✓	69.4	69.4	82.3	85.9	69.4	93.1	93.1	81.7	97.9	93.1
CLAM			✓	63.5	63.5	91.1	83.7	63.5	82.2	82.2	**99.9**	86.4	82.2
TransMIL			✓	66.1	66.1	84.0	85.6	66.1	82.3	82.3	69.3	93.1	82.3
Charm			✓	57.6	57.6	83.0	79.6	57.6	54.3	54.3	81.0	64.4	54.3
Deepglioma			✓	55.3	55.3	84.9	79.9	55.3	66.9	66.9	99.7	84.6	66.9
MCAT		✓	✓	68.0	68.0	89.3	87.7	68.0	–	–	–	–	–
CMTA		✓	✓	66.3	66.3	88.8	87.0	66.3	–	–	–	–	–
Wang et al. [c]	✓	✓	✓	73.0	73.0	82.3	87.2	73.0	95.7	95.7	51.2	99.7	95.7
Ours	✓	✓	✓	**78.6**	**78.6**	85.4	**88.7**	**78.6**	**98.1**	**98.1**	69.5	**99.7**	**98.1**

[a] MTL refers to multi-task learning.
[b] CMM denotes cross-modal modeling.
[c] Wang et al. is the conference version of our method.
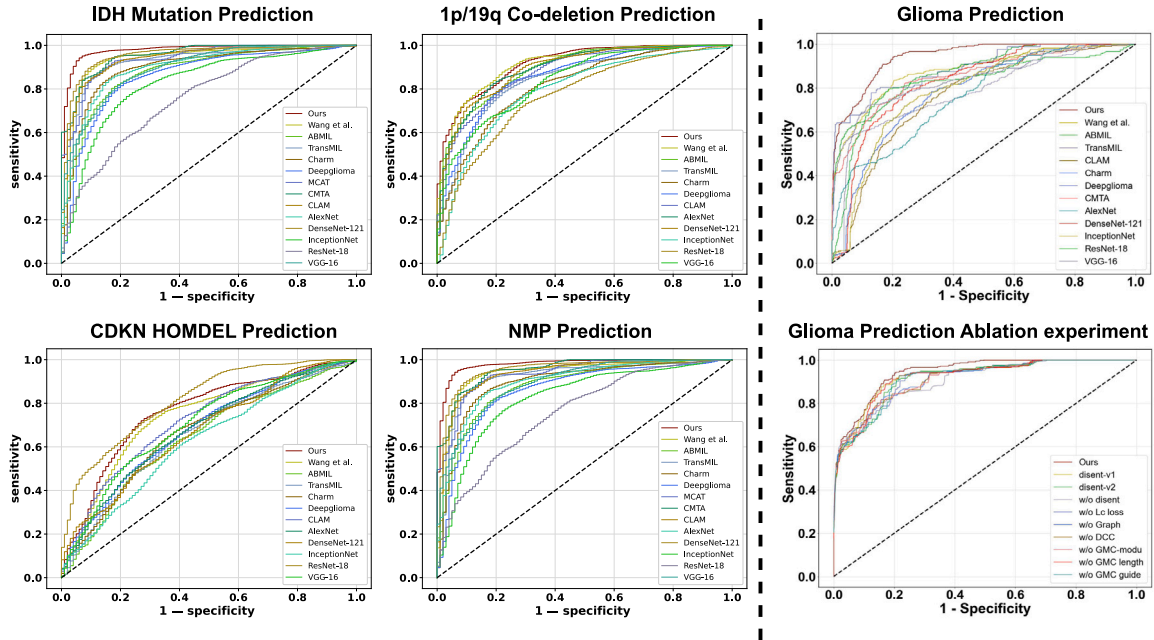
Fig. 5. ROCs of our model, comparison and ablation models for predicting IDH, 1p/19q, CDKN, NMP and Glioma.

In (12), M_0 and p_0 are hyper-parameters to adjust $\mathcal{L}_{DCC}$ in training process.

4. Experiment

4.1. Datasets and training labels

Dataset preparation. Overall, three public datasets, i.e., TCGA GBM-LGG (Tomczak et al., 2015), CPTAC (Verdugo et al., 2022) and IvY-GAP (Eberhart and Bar, 2020), are involved in our study. In all datasets, we remove the WSIs of low quality or lack of labels, consistent with settings in Lu et al. (2021). Totally, we include 3,578 WSIs from 1,054 cases. Following (Nguyen et al., 2023), the two multi-regional TCGA GBM-LGG and CPTAC datasets are further merged as the meta dataset, i.e., internal dataset, for better learning performance. Additionally, IvYGAP dataset is used as an external validation dataset. See more

details of dataset preparation in section 1 of supplementary material.

Training labels. Original labels for molecular markers and histology of WSIs are obtained from the three used datasets. According to the up-to-date WHO criteria (Louis et al., 2021), we generate the classification labels for each case as grade 4 glioblastoma (defined as IDH wildtype), oligodendroglioma (defined as IDH mutant and 1p/19q co-deletion), grade 4 astrocytoma (defined as IDH mutant, 1p/19q non co-deletion with CDKN HOMDEL or NMP), or low-grade astrocytoma (other cases). Detailed diagnosis pipeline of glioma based on WHO 2021 can be found in Figure 1 in supplementary material.

4.2. Implementation details

The proposed M3C2 model is trained on the internal training set for 50 epochs, with a batch size of 6 and a learning rate of 0.003 using the Adam optimizer (Kingma and Ba, 2014), alongside weight decay. Key hyperparameters are detailed in Table 1 of the supplementary material.

Table 2
Mean values in terms of percentage for predicting molecular markers and histology over the internal dataset.

Task	Metrics	Ours	Wang et al.	ABMIL	TransMIL	CLAM	Charm	Deepglioma	MCAT	CMTA	Alex.	Dense.	Inception	Res.	VGG
IDH	Acc.	**88.8**	84.5	80.9	84.2	84.2	82.6	80.9	–	–	83.6	74.0	75.7	68.4	83.2
	Sen.	82.6	78.5	76.0	84.3	78.5	82.6	76.9	–	–	79.3	**92.6**	79.3	39.7	81.0
	Spec.	**92.9**	88.5	84.2	84.2	88.0	82.5	83.6	–	–	86.3	61.7	73.2	87.4	84.7
	AUC	**95.0**	92.1	90.5	89.7	92.7	87.8	87.3	–	–	91.9	86.5	78.9	82.0	88.9
	F_1	**85.5**	80.2	76.0	81.0	79.8	79.1	76.2	–	–	79.3	73.9	72.2	50.0	79.4
1p19q	Acc.	80.6	**82.2**	79.0	78.3	78.6	73.7	79.9	–	–	81.6	78.6	74.3	72.0	72.0
	Sen.	**81.7**	63.4	46.8	40.2	63.4	76.8	72.0	–	–	52.4	57.3	62.2	48.3	84.1
	Spec.	80.2	89.2	87.1	92.3	84.2	72.5	82.9	–	–	**92.3**	86.5	78.8	78.9	67.6
	AUC	**90.9**	89.1	81.6	76.4	85.1	82.0	83.7	–	–	86.8	85.4	76.6	69.0	81.5
	F_1	**69.4**	65.8	47.3	50.0	61.5	61.2	65.9	–	–	60.6	59.1	56.7	43.9	61.9
CDKN	Acc.	**72.7**	71.1	63.5	63.8	65.1	64.1	62.2	–	–	62.8	61.5	60.5	71.6	64.8
	Sen.	76.3	79.4	59.0	68.8	**81.9**	58.1	46.3	–	–	75.6	50.6	41.3	48.3	58.8
	Spec.	68.8	61.8	69.7	58.3	46.5	70.8	79.9	–	–	48.6	73.6	**81.9**	78.4	71.5
	AUC	**76.0**	75.0	66.2	65.7	73.6	67.8	70.7	–	–	66.8	71.8	67.5	72.8	70.7
	F_1	**74.6**	74.3	65.0	66.7	71.2	63.1	56.3	–	–	68.2	58.1	52.4	43.6	63.7
NMP	Acc.	**94.4**	90.1	82.4	87.5	86.2	84.5	80.9	86.0	86.6	85.2	89.8	78.3	69.1	81.6
	Sen.	93.4	88.0	**93.6**	87.4	82.0	82.6	82.6	75.9	77.8	80.2	90.4	73.7	68.9	71.9
	Spec.	**95.6**	92.7	69.4	87.6	91.2	86.9	78.8	90.7	90.7	91.2	89.1	83.9	69.3	93.4
	AUC	**98.1**	94.8	91.3	92.1	94.7	90.0	86.5	95.7	96.7	92.8	95.9	84.6	76.0	93.6
	F_1	**94.8**	90.7	85.1	88.5	86.7	85.4	82.6	77.4	78.5	85.6	90.7	78.8	71.0	81.0

All hyperparameters are fine-tuned to achieve optimal performance on the internal validation set. Experiments are conducted on a computer equipped with an AMD EPYC 7H12 CPU @1.70 GHz, 1TB RAM, and 8 Nvidia Tesla V100 GPUs. Our method is implemented in PyTorch within a Python environment.

4.3. Performance evaluation

Glioma classification. We compare our method with thirteen other SOTA methods, including five commonly-used image classification methods (AlexNet Krizhevsky et al., 2012, ResNet-18 He et al., 2016, InceptionNet Szegedy et al., 2015, MnasNet Tan et al., 2019, DenseNet-121 Huang et al., 2017), three SOTA MIL framework (ABMIL Ilse et al., 2018, TransMIL Shao et al., 2021 and CLAM Lu et al., 2021), two SOTA multi-modal learning methods (MCAT Chen et al., 2021, CMTA Zhou and Chen, 2023) three SOTA methods specially designed for glioma classification based on WHO 2021 (Charm Nasrallah et al., 2023, Deep-glioma Khalighi et al., 2024 and DeepMO-Glioma Wang et al., 2023). Note that DeepMO-Glioma is the conference version of our method. All the compared methods are trained by the classification labels of glioma, which are defined by integrating molecular markers with histology, thereby demonstrating the fairness of our comparison. In our experiments, we apply five metrics[4] to assess glioma classification: accuracy, sensitivity, specificity, AUC and F_1-score. The middle panel of Table 1 shows that M3C2 performs the best on the internal dataset, achieving at least 5.6%, 5.6%, 1.5% and 5.6% improvement over other models in accuracy, sensitivity, AUC and F_1-score, respectively. These significant improvements are mainly due to the following aspects. (1) We design a multi-scale disentangling module to extract multi-magnification WSI features for both the molecular marker and histology prediction tasks (2) We propose a CPLC-Graph network for efficient intra-omic modeling of the intrinsic correlation of the molecular markers. (3) We devise the DCC loss and CMG-Modu training strategy for the cross-modal interaction of histology and molecular markers. The corresponding analysis can be found in the ablation study of Section 4.4. Additionally, Fig. 5 plots the ROCs of all models, demonstrating the superior performance of our model over other comparison models.

Performance of auxiliary tasks of molecular markers and histology prediction. From Table 2, we observe that M3C2 achieves an AUC of 95.0%, 90.9%, 76.0%, and 98.1% for IDH mutation, 1p/19q

co-deletion, CDKN HOMDEL and NMP prediction, respectively, significantly outperforming all comparison models. Of note, the compared multi-modal learning methods, i.e., MCAT and CMTA, take the molecular markers as input and thus cannot perform the molecular markers prediction task. Fig. 5 presents the ROCs of all models, showcasing the superior performance of our model compared to others.

Generalizability validation. We compare our model with SOTA methods on the external validation dataset, i.e., IvYGAP (Eberhart and Bar, 2020), without fine-tuning. Results are shown in the right panel of Table 1. We can observe from the table that our method achieves the increment of 98.1%, 99.7% and 98.1% over the best SOTA method in accuracy, AUC and F_1-score, respectively, suggesting our great robustness and generalizability.

Network interpretability. An additional visualization experiment is conducted based on patch decision scores to test the interpretability of our method. Fig. 6 shows the visualization maps of our method predicting molecular markers and histology. As shown in this figure, our method generates more consistent model decision maps for predicting IDH wildtype (molecular marker) with NMP positive (histology feature) compared to other molecular markers. This finding aligns with the current diagnostic criteria where glioblastoma is predominantly IDH wildtype and NMP positive in histology, suggesting our success in modeling cross-modal co-occurrence. Consequently, this indicates the interpretability of integrating molecular markers with histology for clinical diagnosis.

Subgroup analysis. To validate the effectiveness of our method on different WSI materials, such as frozen and FFPE sections, we perform additional subgroup analysis on multiple tasks including glioma classification, molecular markers prediction and histology prediction. Specifically, the results are presented in Table 3. As shown in this table, the performance of frozen sections is slightly inferior, yet still comparable, to that of FFPE sections for the molecular markers and histology prediction tasks. This is because, compared to frozen sections, FFPE sections preserve tissue architecture and cellular morphology more effectively (Ozyoruk et al., 2021; Zhang et al., 2022b). Additionally, on the glioma classification task, the performance is similar between the two materials, with an AUC of 88.1% for FFPE and 89.2% for frozen sections, highlighting the practical applicability of our method in addressing varied clinical needs, especially for different WSI materials.

4.4. Results of ablation experiments

We ablate different components of our M3C2 method to thoroughly analyze their effects on glioma classification and the prediction of

---

[4] Note that we use micro-average for all metrics in glioma classification.

Table 3
Subgroup analysis in terms of WSI materials on multiple tasks of molecular markers and histology prediction, as well as glioma classification.

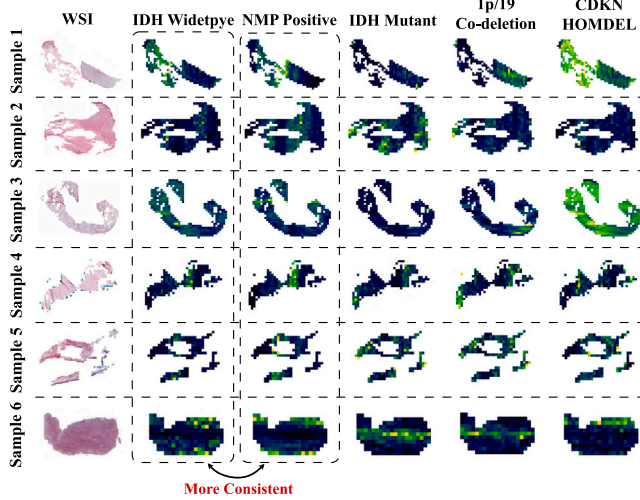| Material	Task	Acc.	Sen.	Spec.	AUC	F1-score
FFPE	Diag	82.1	82.1	86.4	88.1	82.1
	IDH	90.1	89.2	90.9	96.1	89.8
	1p19q	80.1	96.6	69.6	92.2	79.2
	CDKN	80.1	76.5	83.1	82.8	77.6
	Grade	88.7	77.9	97.6	98.4	86.2
Frozen	Diag	75.2	75.2	78.7	89.2	75.2
	IDH	87.6	72.3	94.3	93.1	78.2
	1p19q	81.0	83.5	87.7	83.4	70.8
	CDKN	65.4	76.1	59.2	66.6	72.5
	Grade	78.4	66.7	100.0	98.1	80.0

Fig. 6. Visualization maps of M3C2 predicting molecular markers of IDH mutation, 1p/19q co-deletion and CDKN HOMDEL, as well as histology of NMP.

Fig. 7. Ablation study of CMG-Modu training strategy on the glioma classification task over internal and external validation datasets.

molecular markers and histology features.

Multi-scale disentangling. To evaluate the effectiveness of the proposed multi-scale disentangling module, we conduct ablation experiments on the disentanglement loss $\mathcal{L}_{\text{disent}}$ as follows: (1) *w/o* the constraint between the modality independent features of molecular markers ($\mathbf{I}_m$) and histology ($\mathbf{I}_h$), denoted as $\mathcal{L}_{\text{disent}-v1}$; (2) *w/o* the constraint between the independent and shared features within each modality, denoted as $\mathcal{L}_{\text{disent}-v2}$; (3) *w/o* the disentanglement loss. The results on the glioma classification task and the auxiliary tasks are shown in Table 4 and Table 5, respectively. All three models perform worse than M3C2 on both the internal and external validation datasets, suggesting that the designed multi-scale disentanglement loss can enhance the overall model performance.

CPLC-Graph network. Table 4 shows that, by setting graph balancing weight α to 0 for the proposed CPLC-Graph, the accuracy, sensitivity, specificity, AUC and F1-score decreases by 9.2%, 9.2%, 4.1%, 2.3% and 9.2%, respectively. Similar results are observed for the prediction tasks of molecular markers and histology (Table 5). Also, the ROC of removing the CPLC-Graph network is shown in Fig. 5. These results indicate the utility of the proposed CPLC-Graph network.

LC loss. We further evaluate the effectiveness of our LC loss, by simply removing the LC loss when training our M3C2 method. Table 4 shows that the performance after removing LC loss decreases in four main metrics, causing a reduction of 16.1%, 16.1%, 2.2% and 16.1%, in accuracy, sensitivity, AUC and F1-score, respectively. Similar results for the tasks of molecular marker and histology prediction are observed in the Table 5 with the ROC in Fig. 5, indicating the effectiveness of the
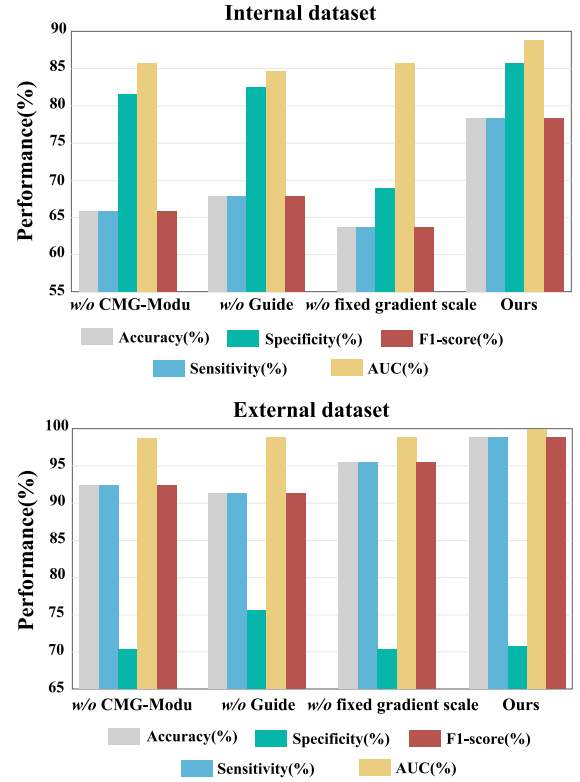
LC loss.

DCC loss. From Table 4, we observe that the proposed DCC loss improves the performance in terms of accuracy by 11.2%. Similar results can be found for sensitivity, AUC and F1-score. From Table 5, we observe that the AUC decreases 8.9%, 2.6%, 0.9% and 5.1% for the prediction of IDH, 1p/19q, CDKN and NMP, respectively, when removing the DCC loss. Such performance is also found in comparing the ROC curves in Fig. 5, suggesting the importance of the DCC loss for all the tasks.

CMG-Modu training strategy. Finally, we validate the effectiveness of the CMG-Modu training strategy, by modifying our training strategy as follows: (1) *w/o* fixed gradient scale - directly project the modulated gradient to the perpendicular direction of the dominant gradient without scaling; (2) *w/o* guide - remove the guidance of histology grade during gradient modulation; (3) *w/o* CMG-Modu - remove the whole CMG-Modu training strategy. The experimental results are shown in Fig. 7, in which all three modified training strategies perform worse than M3C2, indicating the effectiveness of our CMG-Modu training strategy.

Overall, as shown in the left panel of Table 4, the glioma classification performance of our M3C2 slightly decreases when ablating each individual module. For example, the AUC value decreases by only 0.3% and 1.0% when ablating the dynamic confidence constrain (DCC) loss and disentanglement loss on the internal validation dataset, respectively, indicating the relative independence of these modules for glioma classification. Similar results can also be observed on the external validation dataset, as shown in the right panel of Table 4.

Moreover, in Table 4, the superior performance observed on the external dataset compared to the internal dataset is primarily attributed to differences in their class distributions, i.e., the external dataset consists exclusively of two classes: grade 4 GBM and high-grade astrocytoma. The reduced number of classes in the external dataset contributes to the

Table 4
Ablation study on the multi-modal disentanglement loss, CPLC-Graph network, LC loss and DCC loss on the internal and external validation datasets.

Ablation	Internal					External				
	Acc.	Sen.	Spec.	AUC	F1-score	Acc.	Sen.	Spec.	AUC	F1-score
$\mathcal{L}_{\text{disent}-v1}$	65.8	65.8	85.3	86.1	65.8	94.5	94.5	57.3	99.7	94.5
$\mathcal{L}_{\text{disent}-v2}$	67.8	67.8	83.1	86.1	67.8	90.4	90.4	69.3	99.1	90.4
w/o $\mathcal{L}_{\text{disent}}$	64.8	64.8	83.2	87.7	64.8	93.7	93.7	57.3	99.3	93.7
w/o Graph	69.4	69.4	81.3	86.4	69.4	93.1	93.1	**75.5**	99.6	93.1
w/o LC loss	62.5	62.5	86.4	86.5	62.5	90.2	90.2	93.7	98.9	90.2
w/o DCC	67.4	67.4	**90.2**	88.4	67.4	91.3	91.3	81.6	97.5	91.3
Ours	**78.6**	**78.6**	85.4	**88.7**	**78.6**	**98.1**	**98.1**	69.5	**99.7**	**98.1**

Table 5
Ablation study on the multi-modal disentanglement loss, CPLC-Graph network, LC loss, DCC loss and CMG-Modu training strategy on the auxiliary tasks of molecular markers and histology prediction over internal dataset.

Task	Metrics	Ours	$\mathcal{L}_{\text{disent}-v1}$	$\mathcal{L}_{\text{disent}-v2}$	w/o $\mathcal{L}_{\text{disent}}$	w/o Graph	w/o LC loss	w/o DCC	w/o GMC-Modu	w/o Guide	w/o fixed gradient scale
IDH	Acc.	**88.8**	80.6	90.4	80.9	81.6	82.6	78.9	83.2	82.6	85.5
	Sen.	82.6	**93.4**	68.8	81.0	81.0	82.6	71.1	80.2	87.6	89.3
	Spec.	**92.9**	72.1	90.9	80.9	82.0	82.5	84.2	85.2	79.2	83.1
	AUC	**95.0**	89.0	92.7	87.8	87.2	90.0	86.1	90.5	88.2	93.5
	F$_1$-score	**85.5**	79.3	26.8	77.2	77.8	79.1	72.9	79.2	80.0	83.1
1p19q	Acc.	80.6	74.3	73.4	76.0	73.7	75.7	**80.9**	78.3	74.0	79.9
	Sen.	81.7	82.9	**84.1**	51.2	72.0	85.4	84.1	57.3	74.4	59.8
	Spec.	80.2	71.2	69.4	85.1	74.3	72.1	79.7	86.0	73.9	**87.4**
	AUC	**90.9**	83.1	81.6	81.9	82.5	83.0	88.3	84.3	84.2	86.3
	F$_1$-score	69.4	63.6	63.0	53.5	59.6	65.4	**70.4**	58.7	60.7	61.6
CDKN	Acc.	**72.7**	68.8	61.2	66.8	68.4	68.1	67.4	70.4	69.4	72.0
	Sen.	76.3	71.3	**95.0**	61.3	75.0	63.1	91.9	81.3	75.0	78.1
	Spec.	68.8	66.0	23.6	72.9	61.1	**73.6**	40.3	58.3	63.2	65.3
	AUC	**76.0**	72.8	72.7	75.5	71.8	75.2	75.1	73.8	74.6	73.8
	F$_1$-score	**74.6**	70.6	72.0	66.0	71.4	67.6	74.8	74.3	72.1	74.6
NMP	Acc.	**94.4**	84.2	87.8	88.2	87.2	92.1	84.9	90.1	87.5	91.1
	Sen.	**93.4**	82.0	85.6	83.2	94.0	93.4	92.2	88.6	83.2	91.0
	Spec.	**95.6**	86.9	90.5	94.2	78.8	90.5	75.9	92.0	92.7	91.2
	AUC	**98.1**	93.1	93.3	93.8	93.1	96.3	93.0	94.7	93.5	96.2
	F$_1$-score	**94.8**	85.1	88.5	88.5	89.0	92.9	87.0	90.8	88.0	91.8

relatively higher performance. To ensure a fair comparison, we have recalculated all experimental results using only the two classes (grade 4 GBM and high-grade astrocytoma) from the internal dataset. These results are provided in Supplementary Table 3. As shown, the model's performance on the external dataset is comparable to its performance on the internal dataset, further demonstrating the generalizability of our model.

4.5. Analysis on experimental settings

Influence of input magnification. We evaluate the performance of our M3C2 method on the glioma diagnosis task under different input magnification settings, i.e., using only one magnification of 10X and 20X, as well as simply concatenating and adding features of the two magnifications. Table 6 tabulates the results of glioma classification under different magnification settings. As shown in this table, the glioma prediction performance is improved using our proposed multi-scale disentangling module, compared to other settings of single magnification or multi-magnification concatenation and addition. For instance, the classification results of our method are 78.6%, 78.6%, 85.4%, 88.7% and 78.6% in accuracy, sensitivity, specificity, AUC and F1-score, whereas those of multi-magnification concatenation are 69.7%, 69.7%, 81.6%, 88.6% and 69.7%, respectively. To summarize, the above results imply the performance improvement of our method in efficient multi-magnification fusion.

Moreover, additional experimental results on the auxiliary tasks of molecular markers and histology prediction with different input magnification settings can be found in Table 2 in the supplementary material.

Table 6
Mean values in terms of percentage for glioma classification accuracy by our and other magnification-related baseline methods.

Internal dataset					
Scale	Acc.	Sen.	Spec.	AUC	F1-score
10X	68.8	68.8	**88.7**	88.4	68.8
20X	66.4	66.4	88.4	88.5	66.4
10X+20X (concat)	69.7	69.7	81.6	88.6	69.7
10X+20X (add)	74.3	74.3	74.1	88.6	74.3
Ours	**78.6**	**78.6**	85.4	**88.7**	**78.6**
External dataset					
Scale	Acc.	Sen.	Spec.	AUC	F1-score
10X	95.8	95.8	39.1	99.6	95.8
20X	93.1	93.1	63.4	99.3	93.1
10X+20X (concat)	97.3	97.3	39.1	99.7	97.3
10X+20X (add)	96.5	96.5	57.3	99.6	96.5
Ours	**98.1**	**98.1**	69.5	**99.7**	**98.1**

Impact of auxiliary tasks for glioma classification. In our method, we perform multi-task learning with the primary task of glioma classification and the auxiliary tasks of molecular and histology prediction. Thus, we conduct additional experiments to evaluate the effectiveness of the involved auxiliary tasks on the glioma classification performance, via ablating the molecular and/or histology prediction tasks. As shown in Table 7, the glioma classification accuracy decreases 9.5%, 10.5% and 10.8% on the internal dataset when training without histology prediction module, molecular prediction module and both the two modules, respectively. Similar results can be found on the external validation dataset. This indicates that our multi-task learning strategy

Table 7

Mean values in terms of percentage for glioma classification accuracy by our method with and without auxiliary tasks.

Internal dataset					
Structure	Acc.	Sen.	Spec.	AUC	F1-score
w/o histology	69.1	69.1	80.6	87.5	69.1
w/o markers	68.1	68.1	74.3	88.3	68.1
w/o both tasks	67.8	67.8	70.2	87.1	67.8
Ours	**78.6**	**78.6**	**85.4**	**88.7**	**78.6**
External dataset					
Structure	Acc.	Sen.	Spec.	AUC	F1-score
w/o histology	93.4	93.4	69.5	99.5	93.4
w/o markers	92.1	92.1	**87.7**	99.6	92.1
w/o both tasks	92.6	92.6	57.2	99.3	92.6
Ours	**98.1**	**98.1**	69.5	**99.7**	**98.1**

can facilitate the glioma classification performance via incorporating the two auxiliary tasks.

5. Discussion

Recently, cancer diagnosis criteria has shifted from solely relying on histology features to integrating molecular markers and histology features, especially for brain cancers such as diffuse glioma. It is promising to develop DL-based methods to facilitate clinical diagnosis of glioma. Some recent works, such as Charm (Nasrallah et al., 2023) and Deepglioma (Khalighi et al., 2024), have been specially designed for glioma classification based on WHO 2021 criteria. Yet, these existing methods suffer from either incomplete diagnosis pipeline or inability of cross-modal modeling of molecular and histology features. In this context, multi-modal learning (MML) (Zhang et al., 2024a) seems to be a potential solution for the integrated diagnosis of glioma. Traditional MML methods take molecular markers as input, which are not predicted from the WSIs, making them less practical in real-world applications.

In contrast, we propose a multi-task learning method (M3C2) to jointly predict histology features, molecular markers, and glioma classes, incorporating efficient task interaction designs, including the DCC loss and CMG-Modu learning strategy. The DCC loss constrains the histology features of NMP and molecular features of IDH to be close, based on the clinical observation that over 95% of patients with NMP are IDH wildtype (Alzial et al., 2022). many histology-molecular pairs exhibit strong correlations beyond the widely recognized associations between IDH wild-type and NMP. For example, the 1p/19q co-deletion is strongly associated with the characteristic histological features of oligodendroglioma, such as the *fried egg* cellular appearance (Wesseling et al., 2015). Similarly, microvascular proliferation and H3 K27me3 loss are associated pairs in ependymomas. Beyond gliomas, many cancers also exhibit histology-molecular correlations. For instance, in lung mucinous adenocarcinoma, the presence of mucinous histology strongly correlates with KRAS mutations, representing another critical histology-molecular relationship frequently used in clinical practice. Therefore, the proposed DCC loss is not limited to this specific pair and can be generalized not only to other glioma subtypes but also to various cancers, where such histology-molecular correlations play a critical role in diagnosis. Moreover, the CMG-Modu allows gradient guidance from the decisive modality, either histology features or molecular markers, during training. In this strategy, histology prediction serves as the reference for molecular prediction when the WSI shows NMP features, and vice versa. This is motivated by the fact that, under the revised criteria, patients with NMP are still considered as high grade, whereas the grade of patients without NMP features is determined by molecular markers. Furthermore, ablation study results (Table 4 and Fig. 7) validate the effectiveness of our task interaction design.

In addition to the task interaction design, we also propose a multi-scale disentangling module for extracting multi-magnification histology features, which can be applied broadly to pathological image analysis for various diseases. This is particularly relevant in clinical settings, where multi-magnification histopathology images play a vital role in precision medicine, as each magnification level provides unique diagnostic insights. Specifically, high-magnification images capture fine-grainedcellular-level features, such as intricate cell morphology, nuclear atypia, and mitotic activity, which are essential for detailed cellular characterization. Low-magnification images capture tissue-level features, providing a holistic view of tissue architecture, structural organization, and spatial relationships among tissue niches. These cellular-level and tissue-level features offer complementary perspectives and enable a comprehensive understanding of disease, enhancing diagnostic accuracy. By integrating both scales, our framework supports robust and clinically meaningful insights, making it broadly applicable to various types of cancer and other diseases that require detailed histopathological evaluation. To validate the impact of our multi-magnification approach, we conduct experiments (Table 6) comparing our method to single magnification and simple combination methods. The significant performance improvement demonstrates the effectiveness of our multi-magnification design.

To further explore the superiority of each modality on specific groups of patients, we divide the patients into four groups based on their glioma classes. For each group, we generate a histogram where the *y*-axis reflects the superiority of each modality, defined as the ratio of the importance of histology features to molecular markers in the final glioma grading task. Specifically, the importance value of each modality in the final glioma grading task is derived from the gradient scale of the histology and molecular prediction task in our cross-modal gradient modulation strategy. The histograms of the four groups are shown in Figure 3 in supplementary material. As demonstrated, the superiority of each modality varied across groups, which could provide more interpretability for the joint model and suggest significant potential for personalized treatment in the future.

Our method has great potential in real-world clinical scenarios, including the computer-assisted diagnosis of glioma and prediction of crucial molecular markers. The blueprint of the clinical application of our M3C2 method is illustrated in Figure 2 of the supplementary material. As shown, compared to the current clinical glioma diagnosis process , our proposed AI system (M3C2) can make diagnosis with less expenses, especially in obtaining molecular markers. In addition, our method can be further deployed in hospitals to offer real-time, AI-assisted glioma diagnosis based on WSIs, potentially improving the diagnosing efficiency and further accelerating the treatment planning for patients. Moreover, our M3C2 method can be used in many other clinical scenarios, such as the rapid diagnosis during neuro-therapy with frozen sections, thereby increasing the rate of successful surgeries. Overall, these potential real-world applications imply how our M3C2 can contribute to precision neuro-oncology.

Despite the high performance of classifying glioma, our M3C2 still has some weakness. For example, we only pay attention to predicting molecular markers essential for glioma diagnosis, without considering other molecular markers important for treatment planning and disease prognosis, e.g., MGMT (O-6-methylguanine-DNA methyltransferase) methylation. Furthermore, our method is dependent on the completeness of the multi-modal data of histology and molecular markers. However, certain modalities, e.g., molecular markers, could be missing in real-world scenarios. For instance, the molecular markers of CDKN might be inaccessible for patients with NMP features, especially for those from retrospective study. One possible solution for incomplete modalities is to propose a more modal-adaptable approach, such as incorporating modality-imputation techniques or leveraging weakly supervised learning algorithms.

In future work, it is essential to explore additional potential biomarkers, such as MGMT methylation, EGFR (epidermal growth

factor receptor) amplification, ATRX (alpha thalassemia/mental retardation syndrome X-linked) mutation, to enhance the diagnostic and prognostic capabilities of our study. These markers could provide more comprehensive insights into tumour behavior and patient treatment outcomes. Furthermore, using more WSI magnifications, e.g., 40X and 5X, can also promise to facilitate the efficient extraction of morphology feature for better model performance. Finally, according to the most recent WHO cancer diagnostic criteria, the diagnosis of many other cancers, including endometrial cancer (Imboden et al., 2021), renal neoplasia (Trpkov et al., 2021) and thyroid carcinomas (Volante et al., 2021), has also shifted from purely relying on histology features to integrating molecular and histology features. Hence, our proposed M3C2 for cancer classification framework can be applicable for these cancers by jointly predicting molecular markers and histology features while modeling their interactions for cancer classification.

6. Conclusion

The paradigm of pathology diagnosis of diffuse gliomas has shifted to integrating molecular makers with histology features. In this paper, we target on classifying glioma under the latest diagnosis criteria, via jointly learning the tasks of molecular marker and histology prediction, as well as the final glioma classification. Inputting multi-magnification histology WSIs, our model incorporates a novel AHMT-MIL framework with multi-scale disentangling to extract both cellular-level and tissue-level information for the downstream tasks. Moreover, a CPLC-Graph network is devised for intra-omic interactions, while a DCC loss and a CMG-Modu training strategy are further designed for inter-omic interactions. Our experiments demonstrate that M3C2 achieves superior and more robust performance over other state-of-the-art methods, opening a new avenue of for digital pathology based on WSIs in the era of molecular pathology.

CRediT authorship contribution statement

Xiaofei Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hanyu Liu:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Yupei Zhang:** Writing – original draft, Validation, Methodology, Conceptualization. **Boyang Zhao:** Writing – review & editing, Visualization, Validation, Data curation. **Hao Duan:** Writing – review & editing, Visualization, Data curation. **Wanming Hu:** Writing – review & editing, Methodology, Investigation, Data curation. **Yonggao Mou:** Writing – review & editing, Methodology, Data curation. **Stephen Price:** Supervision, Resources, Funding acquisition. **Chao Li:** Conceptualization, Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that they have nothing to disclose.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xiaofei Wang reports financial support was provided by China Scholarship Council. Stephen Price reports financial support was provided by National Institute for Health and Care Research. Chao Li reports financial support was provided by Guarantors Of Brain. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2025.103505.

Data and code availability

This paper uses three publicly available datasets, including TCGA GBM-LGG dataset (https://portal.gdc.cancer.gov/), CPTAC dataset (https://portal.gdc.cancer.gov/) and IvYGAP dataset (https://glioblastoma.alleninstitute.org/static/home/). The source code is available at https://github.com/LHY1007/M3C2/.

References

Alzial, G., Renoult, O., Paris, F., Gratas, C., Clavreul, A., Pecqueur, C., 2022. Wild-type isocitrate dehydrogenase under the spotlight in glioblastoma. Oncogene 41 (5), 613–621.

Berger, T.R., Wen, P.Y., Lang-Orsini, M., Chukwueke, U.N., 2022. World health organization 2021 classification of central nervous system tumors and implications for therapy for adult-type gliomas: a review. JAMA Oncol..

Bhattacharyya, R., Roy, P., Banerji, S., Mitra, S., 2024. Efficient grading of prostate cancer WSI with deep learning. In: Medical Imaging 2024: Digital and Computational Pathology, vol. 12933, SPIE, pp. 20–26.

Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F., 2020. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans. Med. Imaging 41 (4), 757–770.

Chen, R.J., Lu, M.Y., Weng, W.-H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F., 2021. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: 2021 IEEE/CVF International Conference on Computer Vision. ICCV, pp. 3995–4005. http://dx.doi.org/10.1109/ICCV48922.2021.00398.

Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A., 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature Med. 24 (10), 1559–1567.

Deng, S., Zhang, X., Yan, W., Chang, E.I.C., Fan, Y., Lai, M., Xu, Y., 2020. Deep learning in digital pathology image analysis: a survey. Front. Med. 14, 470–487.

Ding, K., Zhou, M., Metaxas, D.N., Zhang, S., 2023. Pathology-and-genomics multimodal transformer for survival outcome prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 622–631.

Dosovitskiy, A., et al., 2020. An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Eberhart, C.G., Bar, E.E., 2020. Spatial enrichment of cellular states in glioblastoma. Acta Neuropathol. 140 (1), 85–87.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.

Hao, F., Liu, X., Li, M., Han, W., 2023. Accurate kidney pathological image classification method based on deep learning and multi-modal fusion method with application to membranous nephropathy. Life 13 (2), 399.

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.

Hollon, T., Jiang, C., Chowdury, A., Nasir-Moin, M., Kondepudi, A., Aabedi, A., Adapa, A., Al-Holou, W., Heth, J., Sagher, O., 2023. Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging. Nature Med. 29 (4), 828–832.

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.

Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning. In: International Conference on Machine Learning. PMLR, pp. 2127–2136.

Imboden, S., Nastic, D., Ghaderi, M., Rydberg, F., Siegenthaler, F., Mueller, M.D., Rau, T.T., Epstein, E., Carlson, J.W., 2021. Implementation of the 2021 molecular ESGO/ESTRO/ESP risk groups in endometrial cancer. Gynecol. Oncol. 162 (2), 394–400.

Jaume, G., Vaidya, A., Chen, R.J., Williamson, D.F., Liang, P.P., Mahmood, F., 2024. Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11579–11590.

Jose, L., Liu, S., Russo, C., Cong, C., Song, Y., Rodriguez, M., Di Ieva, A., 2023. Artificial intelligence–assisted classification of gliomas using whole slide images. Arch. Pathol. Lab. Med. 147 (8), 916–924.

Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F., Yamazaki, K., Takeo, S., Iizuka, O., Tsuneki, M., 2020. Weakly-supervised learning for lung carcinoma classification using deep learning. Sci. Rep. 10 (1), 9297.

Khalighi, S., Reddy, K., Midya, A., Pandav, K.B., Madabhushi, A., Abedalthagafi, M., 2024. Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. NPJ Precis. Oncol. 8 (1), 80.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25.

Li, X., Li, C., Rahaman, M.M., Sun, H., Li, X., Wu, J., Yao, Y., Grzegorzek, M., 2022a. A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. Artif. Intell. Rev. 55 (6), 4809–4878.

Li, X., Wu, H., Li, M., Liu, H., 2022b. Multi-label video classification via coupling attentional multiple instance learning with label relation graph. Pattern Recognit. Lett. 156, 53–59.

Lin, H., Chen, H., Dou, Q., Wang, L., Qin, J., Heng, P.-A., 2018. Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide image. In: 2018 IEEE Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 539–546.

Liu, W., Wang, H., Shen, X., Tsang, I.W., 2021. The emerging trends of multi-label learning. IEEE Trans. Pattern Anal. Mach. Intell. 44 (11), 7955–7974.

Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P., 2007. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. 114, 97–109.

Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta Neuropathol. 131, 803–820.

Louis, D.N., et al., 2021. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro-Oncol. 23 (8), 1231–1251.

Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. 5 (6), 555–570.

Lv, Q., Liu, Y., Sun, Y., Wu, M., 2024. Insight into deep learning for glioma IDH medical image analysis: A systematic review. Medicine 103 (7), e37150.

Mao, Y., Jiang, L., Chen, X., Li, C., 2023. Disc-diff: Disentangled conditional diffusion model for multi-contrast mri super-resolution. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 387–397.

Martínez-Fernandez, E., Rojas-Valenzuela, I., Valenzuela, O., Rojas, I., 2023. Computer aided classifier of colorectal cancer on histopatological whole slide images analyzing deep learning architecture parameters. Appl. Sci. 13 (7), 4594.

Molinaro, A.M., Taylor, J.W., Wiencke, J.K., Wrensch, M.R., 2019. Genetic and molecular epidemiology of adult diffuse glioma. Nat. Rev. Neurol. 15 (7), 405–417.

Nasrallah, M.P., Zhao, J., Tsai, C.C., Meredith, D., Marostica, E., Ligon, K.L., Golden, J.A., Yu, K.H., 2023. Machine learning for cryosection pathology predicts the 2021 WHO classification of glioma. Med 4 (8), 526–540.

Nguyen, T., Khadka, R., Phan, N., Yazidi, A., Halvorsen, P., Riegler, M.A., 2023. Combining datasets to improve model fitting. In: 2023 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–9.

Ozyoruk, K.B., Can, S., Gokceler, G.I., Basak, K., Demir, D., Serin, G., Hacisalihoglu, U.P., Kurtuluş, E., Darbaz, B., Lu, M.Y., et al., 2021. Deep learning-based frozen section to FFPE translation. arXiv preprint arXiv:2107.11786.

Pei, L., Jones, K.A., Shboul, Z.A., Chen, J.Y., Iftekharuddin, K.M., 2021. Deep neural network analysis of pathology images with integrated molecular data for enhanced glioma classification and grading. Front. Oncol. 11, 668694.

Perry, A., Wesseling, P., 2016. Histologic classification of gliomas. Handb. Clin. Neurol. 134, 71–95.

Phan, N.T., Rizzoli, S.O., 2020. High-resolution molecular imaging and its applications in brain and synapses. In: Volume Microscopy: Multiscale Imaging with Photons, Electrons, and Ions. Springer, pp. 37–58.

Qiu, L., Zhao, L., Zhao, W., Zhao, J., 2024. Dual-space disentangled-multimodal network (DDM-net) for glioma diagnosis and prognosis with incomplete pathology and genomic data. Phys. Med. Biol. 69 (8), 085028.

Ren, E., Kim, S., Mohamad, S., Huguet, S.F., Shi, Y., Cohen, A.R., Piddini, E., Salas, R.C., 2021. Deep learning-enhanced morphological profiling predicts cell fate dynamics in real-time in hPSCs. BioRxiv, 2021-07.

Schmitz, R., Madesta, F., Nielsen, M., Krause, J., Steurer, S., Werner, R., Rösch, T., 2021. Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. Med. Image Anal. 70, 101996.

Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., Zhang, Y., 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Adv. Neural Inf. Process. Syst. 34, 2136–2147.

Shao, W., Han, Z., Cheng, J., Cheng, L., Wang, T., Sun, L., Lu, Z., Zhang, J., Zhang, D., Huang, K., 2019. Integrative analysis of pathological images and multi-dimensional genomic data for early-stage cancer prognosis. IEEE Trans. Med. Imaging 39 (1), 99–110.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2818–2826.

Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., Le, Q.V., 2019. Mnasnet: Platform-aware neural architecture search for mobile. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2820–2828.

Tomczak, K., Czerwińska, P., Wiznerowicz, M., 2015. Review the cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp. Oncol./ WspÓ l Czesna Onkol. 2015 (1), 68–77.

Trpkov, K., Hes, O., et al., 2021. New developments in existing WHO entities and evolving molecular concepts: The genitourinary pathology society (GUPS) update on renal neoplasia. Mod. Pathol. 34 (7), 1392–1424.

Uegami, W., Bychkov, A., Ozasa, M., Uehara, K., Kataoka, K., Johkoh, T., Kondoh, Y., Sakanashi, H., Fukuoka, J., 2022. MIXTURE of human expertise and deep learning—developing an explainable model for predicting pathological diagnosis and survival in patients with interstitial lung disease. Mod. Pathol. 35 (8), 1083–1091.

Verdugo, E., Puerto, I., Medina, M.Á., 2022. An update on the molecular biology of glioblastoma, with clinical implications and progress in its treatment. Cancer Commun. 42 (11), 1083–1111.

Volante, M., Lam, A.K., Papotti, M., Tallini, G., 2021. Molecular pathology of poorly differentiated and anaplastic thyroid cancer: what do pathologists need to know? Endocr. Pathol. 32, 63–76.

Wang, X., Huang, X., Price, S., Li, C., 2024a. Cross-modal diffusion modelling for super-resolved spatial transcriptomics. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 98–108.

Wang, X., Price, S., Li, C., 2023. Multi-task learning of histology and molecular markers for classifying diffuse glioma. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 551–561.

Wang, P., Zhang, H., Zhu, M., Jiang, X., Qin, J., Yuan, Y., 2024b. MGIML: Cancer grading with incomplete radiology-pathology data via memory learning and gradient homogenization. IEEE Trans. Med. Imaging.

Wei, Y., Chen, X., Zhu, L., Zhang, L., Schönlieb, C.B., Price, S., Li, C., 2023. Multi-modal learning for predicting the genotype of glioma. IEEE Trans. Med. Imaging 42 (11), 3167–3178.

Wesseling, P., van den Bent, M., Perry, A., 2015. Oligodendroglioma: pathology, molecular mechanisms and markers. Acta Neuropathol. 129, 809–827.

Xing, X., Chen, Z., Zhu, M., Hou, Y., Gao, Z., Yuan, Y., 2022. Discrepancy and gradient-guided multi-modal knowledge distillation for pathological glioma grading. In: MICCAI 2022. Springer, pp. 636–646.

Yang, T.L., Tsai, H.W., Huang, W.C., Lin, J.C., Liao, J.B., Chow, N.H., Chung, P.C., 2022. Pathologic liver tumor detection using feature aligned multi-scale convolutional network. Artif. Intell. Med. 125, 102244.

Zhang, L., Fritah, S., Nazarov, P.V., Kaoma, T., Van Dyck, E., 2023. Impact of IDH mutations, the 1p/19q co-deletion and the G-CIMP status on alternative splicing in diffuse gliomas. Int. J. Mol. Sci. 24 (12), 9825.

Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y., 2022a. Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812.

Zhang, L., Wei, Y., Fu, Y., Price, S., Schönlieb, C.B., Li, C., 2022b. Mutual contrastive low-rank learning to disentangle whole slide image representations for glioma grading. arXiv preprint arXiv:2203.04013.

Zhang, Y., Xu, Y., Chen, J., Xie, F., Chen, H., 2024a. Prototypical information bottlenecking and disentangling for multimodal cancer survival prediction. arXiv preprint arXiv:2401.01646.

Zhang, J., Yan, R., Perelli, A., Chen, X., Li, C., 2024b. Phy-diff: Physics-guided hourglass diffusion model for diffusion MRI synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 345–355.

Zhou, F., Chen, H., 2023. Cross-modal translation and alignment for survival analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21485–21494.