Multimodal Outer Arithmetic Block Dual Fusion of Whole Slide Images and Omics Data for Precision Oncology

Omnia Alwazzan, Amaya Gallagher-Syed, Thomas O. Millner, Sebastian Brandner, Ioannis Patras, *Senior Member, IEEE*, Silvia Marino, Gregory Slabaugh, *Senior Member, IEEE*

***Abstract*— The integration of DNA methylation data with a Whole Slide Image (WSI) offers significant potential for enhancing the diagnostic precision of central nervous system (CNS) tumor classification in neuropathology. While existing approaches typically integrate encoded omic data with histology at either an early or late fusion stage, the potential of reintroducing omic data through dual fusion remains unexplored. In this paper, we propose the use of omic embeddings during early and late fusion to capture complementary information from local (patch-level) to global (slide-level) interactions, boosting performance through multimodal integration. In the early fusion stage, omic embeddings are projected onto WSI patches in latent-space, which generates embeddings that encapsulate per-patch molecular and morphological insights. This effectively incorporates omic information into the spatial representation of the WSI. These embeddings are then refined with a Multiple Instance Learning gated attention mechanism which attends to diagnostic patches. In the late fusion stage, we reintroduce the omic data by fusing it with slide-level omic-WSI embeddings using a Multimodal Outer Arithmetic Block (MOAB), which richly intermingles features from both modalities, capturing their correlations and complementarity. We demonstrate accurate CNS tumor subtyping across 20 fine-grained subtypes and validate our approach on benchmark datasets, achieving improved survival prediction on TCGA-BLCA and competitive performance on TCGA-BRCA compared to state-of-the-art methods. This dual fusion strategy enhances interpretability and classification performance, highlighting its potential for clinical diagnostics.**

***Index Terms*— Multimodal Deep Learning, Dual Fusion, Digital Pathology, Omics, Survival Prediction**

I. Introduction

EPIGENETICS describes a raft of molecular mechanisms which affect gene expression without changing the DNA sequence itself. DNA methylation is an epigenetic process where methyl groups are added to specific sites on DNA molecules, typically at cytosine bases followed by guanine (CpG sites) [1]. Such a process often leads to gene silencing or reduced gene expression by preventing transcription factors from binding to DNA, thereby inhibiting gene activation, but affects gene expression in many complex ways. The DNA methylation landscape of a cell is dependent on its developmental course and function, and this landscape can be grossly disrupted in the setting of cancer. In central nervous system (CNS) tumors, DNA methylation patterns offer valuable insights, enabling the differentiation of tumor subtypes, prediction of clinical outcomes, and guidance on treatment strategies [2]. Diagnosis solely based on the Whole Slide Image (WSI) can be challenging due to the overlapping appearance of different tumor subtypes, resulting in high inter-observer variability [3].

Given the aggressive nature of malignant CNS tumors and their associated poor survival rates, there is a critical need to improve diagnostic precision [4], expedite diagnostic timeframes and identify targets for future personalized treatments. Importantly, evaluation of digitized WSI is becoming increasingly utilized in these clinical diagnostic pathways. Incorporating DNA methylation profiling for diagnosis, one study in pediatric patients demonstrated altered subtype classifications in 35% of cases, potentially impacting treatment decisions for 4% of pediatric patients [5], while another study in an adult population showed diagnosis was changed in 25%, refined in 4% and confirmed in 25% of cases [6]. This demand has driven the development of CNS tumor classifiers based on DNA methylation array data, providing significant advancements in neuropathology [3]. The World Health Organization (WHO) has also responded by incorporating molecular profiling alongside traditional histology into the latest CNS tumor classification guidelines, which defines 40 tumor types and subtypes based on key molecular characteristics [7]. Such findings have inspired further research into automated integrative molecular morphology classification systems using artificial intelligence (AI) algorithms, including machine learning (ML) and deep learning (DL) approaches, to improve tumor diagnosis and prognosis [3], [8]–[13].

The authors appreciate the support of the University of Jeddah and the Saudi Arabia Cultural Bureau. This paper utilised Queen Mary's Andrena HPC facility. This work also acknowledges the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330), a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George's University Hospitals NHS Foundation Trust and St George's University of London.

The authors are affiliated with Queen Mary University of London. Corresponding author (e-mail:o.alwazzan@qmul.ac.uk).

Sebastian Brandner is with the Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, and Department of Neurodegenerative Disease, Queen Square, Institute of Neurology, University College London, London, United Kingdom (e-mail: s.brandner@ucl.ac.uk).

Unimodal Approaches. In the single modality domain, Capper et al. [9] pioneered a methylation-based classification system on 2801 CNS tumor samples using an ML approach [9]. For each CNS tumor subtype included in the classifier, also known as the methylation class, the classifier generated a predicted probability (calibrated score) that summed to 1 [14], with tumors with a score below 0.3 classified as "no match". Capper's method now serves as an essential aid in the routine diagnostic workup of CNS tumors [3]. Hwang et al. [15] have developed an image-based DL model using DNA methylation data to predict the origin of cancers of unknown primary (CUPs). By employing a vision transformer to organ-specific DNA methylation images, their approach shows significant potential for enhancing CUP diagnosis and informing treatment strategies. DNA methylation profiling not only aids in precise classification but also supports surgical strategies specific to CNS tumors, improving surgical outcomes and overall patient care [4]. Djirackor et al. [4] utilize ML algorithms to classify brain tumors in real-time by taking the DNA methylation data of a new tumor sample and comparing it to a database of known methylation signatures, assigning a classification based on the closest match. This allows for fast intraoperative decision-making by providing molecular insights during surgery.

Multimodal Approaches. In the multimodal domain, Hoang et al. [12] developed "Deploy", a DL model designed to predict DNA methylation beta values from WSIs. Additionally, Zheng et al. [16] demonstrated that classical ML algorithms can link DNA methylation profiles of cancer samples with morphometric features from WSIs, showing improved model performance when genes are grouped into methylation clusters. Sturm et al. [17] explore the use of a multiomic approach — integrating genomics, transcriptomics, and epigenomics data — to improve the diagnostic accuracy of pediatric brain tumors, which are often challenging to classify due to overlapping histological features. However, few studies [12] have combined epigenetic data with WSIs, primarily due to the significant data size and complexity of both modalities, which require extensive preprocessing and developing advanced fusion techniques to address their heterogeneity.

Multimodal DL approaches combining histology and omic data for improved survival prediction have gained considerable attention in recent years [8], [10], [11], [13], [18]–[25]. Several studies [10], [13], [22], [26] have highlighted the value of different fusion stages (early or late), particularly emphasizing early fusion for its ability to create an explainable framework from heterogeneous data. However, Zhang et al. [20] argue that early or late fusion methods can partially overlook modality-specific information, potentially leading to a decline in quantitative or qualitative performance. Accordingly, they propose a prototypical information bottleneck framework to maintain modality-specific information while simultaneously reducing redundancy. Furthermore, a joint distribution of feature embeddings is used to calculate the mutual information between omic and WSI modalities.

Despite the promising performance of multimodal approaches in medical diagnostics, significant challenges remain in effectively integrating and analyzing diverse data types, particularly in the context of CNS tumor subtyping. Our assessment of existing methods reveals three key gaps:

1) **Limitations of WSI-only diagnosis**: While WSI serves as a primary diagnostic tool for pathologists, accurately identifying fine-grained tumor subtypes based on morphology alone is challenging due to visual feature overlap among CNS subtypes. This similarity increases the risk of including irrelevant tumor regions and highlights the need for complementary data sources to achieve more precise subtyping [3].

2) **Shortcomings of DNA methylation classification**: DNA methylation classifiers have demonstrated high accuracy in tumor subtyping, but they lack the ability to connect this accuracy to specific regions within WSIs. The main limitation is that these classifiers focus solely on DNA methylation profiles without considering the spatial context provided by WSIs. This limits their capacity to capture how epigenetic patterns contribute to the morphological characteristics observed in specific regions, ultimately reducing the interpretability and comprehensive understanding that could be achieved through integration with WSI data.

3) **Scarcity of integrative models for CNS tumor subtyping**: This challenge is particularly acute in the context of central nervous system tumors, where the inherent heterogeneity and large size of both DNA methylation arrays (typically 850k one-dimensional vectors) and WSIs (represented with multi-dimensional matrices up to 150k x 150k pixels) have received limited attention. The lack of effective multimodal fusion methods in this domain presents an opportunity to leverage the powerful discriminative capacity of DNA methylation data to improve subtyping and enhance the clinical translation of these advanced imaging and molecular techniques.

A. Motivations and Contributions

Addressing these challenges, we propose a novel dual fusion approach to improve CNS tumor subtyping by integrating DNA methylation data with WSIs. Our dual fusion strategy is specifically motivated by the following considerations:

- **Local interpretability via early fusion**: By injecting encoded omics features into patch-level WSI embeddings, early fusion enables the model to learn fine-grained, spatially-informed representations. This is especially valuable in the absence of patch-level labels, as it facilitates interpretability through attention-based heatmaps highlighting molecularly salient regions.
- **Global contextualization via late fusion**: After aggregating patch-level features, late fusion captures interactions across the entire slide. This provides the model with a holistic tumor context, which is critical given the spatial and molecular heterogeneity of CNS tumors.
- **Complementarity of fusion stages**: By integrating both strategies, our framework benefits from the dominant modality (DNA methylation), which offers rich and discriminative features for tumor identification, while anchoring its predictions in interpretable, spatially-resolved morphological features.

Thus, we design MOAD-FNet, a **M**ultimodal **O**uter **A**rithmetic **D**ual **F**usion **Net**work that combines two fusion variants: early fusion focused on capturing essential local interactions and late fusion for broader, richer cross-modal global context, ultimately providing complementary insights and improving the model's decision-making process. This dual fusion strategy provides comprehensive, holistic integration of cross-modal data, maximizing the strengths of each fusion type. Our main contributions are as follows:

- We introduce a novel dual fusion network that seamlessly incorporates both early and late fusion approaches, enabling detailed integration of molecular and imaging data at patch and slide levels.
- Our approach develops a unique Multimodal Outer Arithmetic Block (MOAB) fusion strategy that enhances cross-modal feature interaction and improves the model's ability to capture complex tumor subtype features.
- Our method, MOAD-FNet, is the first imaging-omics pipeline to leverage the NHNN BRAIN UK dataset, demonstrating exceptional performance in brain tumor subtyping. Extensive ablation studies on TCGA datasets validate its robustness, achieving state-of-the-art survival prediction on the BLCA dataset and competitive results on the BRCA dataset, showcasing its versatility across multimodal oncology tasks.

Our approach builds on prior observations that early fusion alone may overlook global relationships, while late fusion lacks spatial grounding and interpretability [13], [27]. Our dual fusion strategy addresses both limitations, offering a principled and effective solution well suitted to the problem of multimodal CNS tumor subtyping.

Note that this paper extends our previous conference paper [28] by implementing a dual fusion architecture, working with three additional imaging-omics datasets, recently proposed state-of-the-art (SOTA) backbones, and shows how heatmaps can be generated, providing interpretability.

II. Study Design

A. Datasets

The following sections briefly provide an overview of the datasets used to evaluate MOAD-FNet.

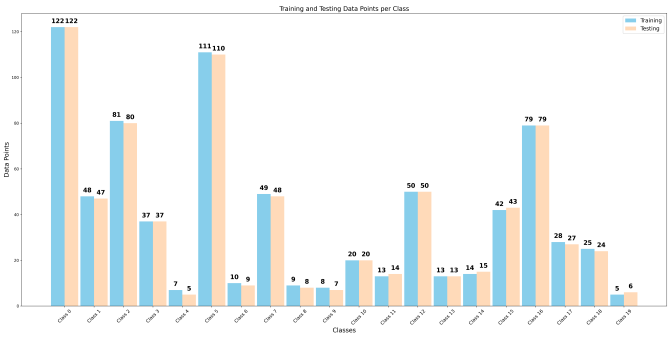

Fig. 2. Distribution of training and testing data points across 20 classes/subtypes. The bar chart illustrates the number of patients allocated to training and testing sets for each class, highlighting the balance of data used for model development and evaluation.

1) Slide Subtyping - NHNN BRAIN UK Brain Tumor Dataset: We obtained WSIs and matched DNA methylation data from the NHNN, University College London Hospital (UCLH) through the UK Brain Archive Information Network [29][1]. The dataset includes (H&E)-stained WSIs from 1,504 patients, covering 20 DNA-methylation-based subtypes of high- and low-grade glial tumors. The data exhibits significant class imbalance across its 20 subtypes, as shown in Fig. 2. To ensure a balanced evaluation, we perform 2-fold cross-validation by splitting the dataset into two equal parts (50%-50%). In each fold, one part is used for training and the other for testing, ensuring all slides are evaluated while maintaining a strict separation between the two sets based on patient IDs. We report the average performance across both folds for all metrics. For WSIs, we used QuPath [30] for tissue segmentation and tiling, generating 1M patches (4K–10K per slide). For the tissue methylation profile, we processed raw IDAT files to extract DNA methylation values at CpG sites using the Illumina EPIC v1 BeadChip array aligned to the GRCh38 (hg38) human genome reference. Specifically, we used an R-based pipeline that employed the `wateRmelon` package [31] for signal extraction and normalization. To ensure robust quantification, and following the workflow described in [32], we removed probes mapped to sex chromosomes (chrX, chrY, and chrM) to avoid gender-specific technical biases. Normalization was then performed using the `dasen` method on autosomal CpGs only. Additionally, we filtered out probes associated with known single nucleotide polymorphisms (SNPs), which are known to interfere with probe binding and methylation signal reliability. The resulting matrix comprised high-confidence methylation measurements across 850,000 CpG sites for 1,504 patients. Next, we calculated M-values to quantify methylation levels, resulting in a matrix of 850K CpG sites for 1,504 patients. The M-value for a given CpG site is defined as:

$$M = \log_2 \left(\frac{I_M}{I_U} \right) \tag{1}$$

where I_M and I_U represent the intensities of the methylated and unmethylated signals, respectively. To mitigate the risk of overfitting associated with high-dimensional omic inputs (850K), we employed a range of feature selection methods to reduce dimensionality. Specifically, we used variance, coefficient of variation (CV), median absolute deviation (MAD), and interquartile range (IQR), selecting the top 8K most variable CpG sites by intersecting these methods. We also experimented with a clustering-based selection of 4K CpG sites and a Random Forest-based selection of 10K sites, as suggested in the literature [9], [32], [33]. However, the 4K feature set reduced performance, and the 10K set showed no further improvement. Therefore, we adopted the 8K CpG feature subset in our final model.

2) Survival prediction - TCGA Datasets: To showcase MOAD-FNet's versatility and compare it with other omic-WSI fusion methods [8], [10], [13], [22], [34], we evaluate its performance on The Cancer Genome Atlas (TCGA) datasets for Bladder Urothelial Carcinoma (BLCA) $(n = 359)$ and

[1]https://www.southampton.ac.uk/brainuk

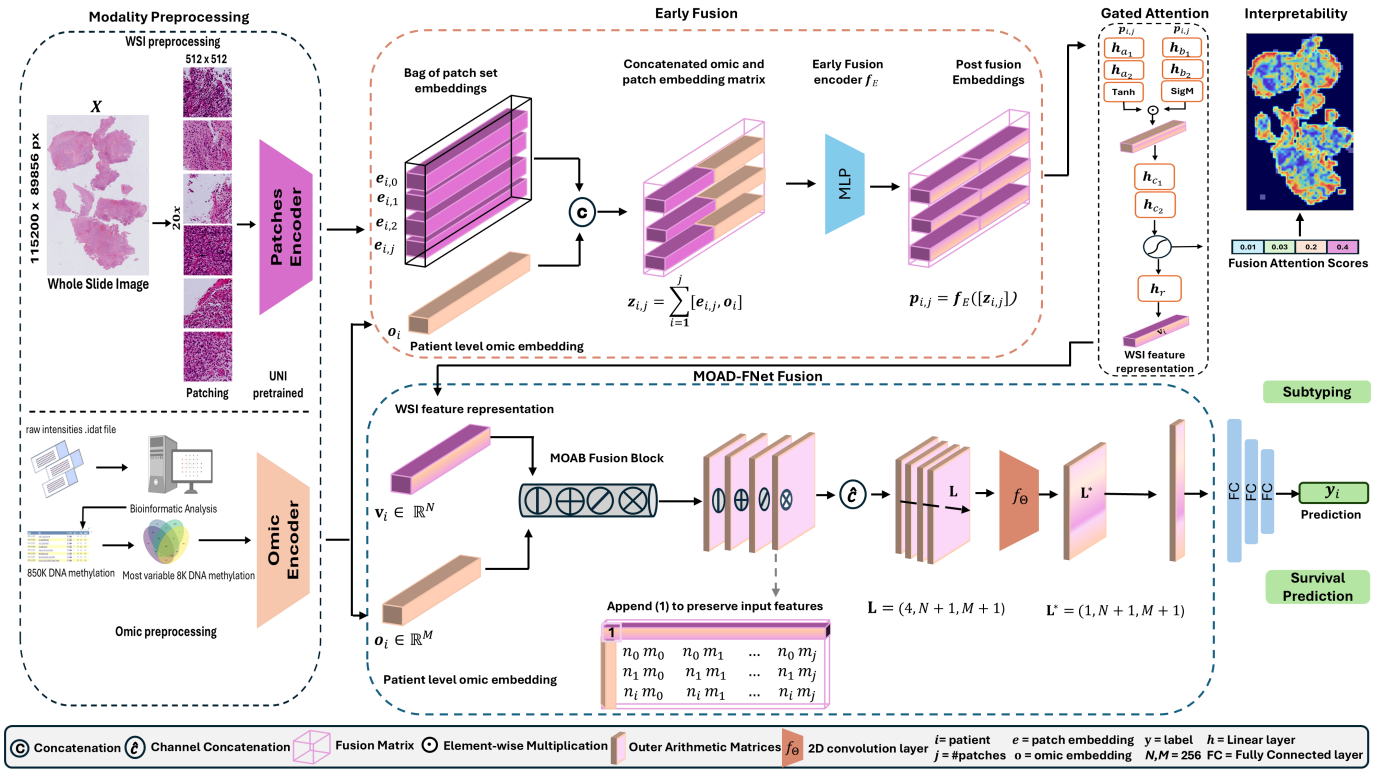

Fig. 1. Overview of the proposed MOAD-FNet framework. Data engineering and encoding for each modality are performed in the preprocessing block. The early fusion block (top) receives encoded inputs from both modalities, where omic data is concatenated to form a matrix $\mathbf{z_{i,j}}$ which is processed by an MLP encoder that learns a joint mapping, resulting in output $\mathbf{p_{i,j}}$. A gated attention via multiple instance learning (MIL) scores patch importance, providing heatmap interpretability, and producing a WSI feature $\mathbf{v_i}$. Next, the MOAD-FNet fusion block (bottom) reintroduces omic features $\mathbf{o_i}$ alongside the $\mathbf{v_i}$ feature representation from the early fusion block as input to the MOAB fusion block. This block performs four outer arithmetic operations to create fusion representations, which are further reduced with $\boldsymbol{f_\theta}$ before being sent to the final subtyping classifier.

Breast Invasive Carcinoma (BRCA) ($n = 869$), using the survival prediction task outlined by Jaume et al. [13]. We utilized a coupled set of 331 biological pathways derived from 4,999 distinct genes, as previously defined in [13]. These pathway-derived features are based on gene-level data, which were selected based on their known involvement in key biological processes associated with cancer progression and clinical outcomes relevant to the BLCA and BRCA datasets. We followed identical data splits provided by Jaume et al. [13], which are publicly available in the official SurvPath GitHub repository[2]. These include the TCGA-BLCA and TCGA-BRCA folders. Their setup uses 5-fold cross-validation, with each fold stratified by sample site to reduce potential batch effects. This ensured consistency with prior work and allowed for fair comparison across models.

B. Multimodal Outer Arithmetic Dual Fusion Network

We designed a multimodal fusion framework, shown in Fig. 1, that integrates omic data and a WSI through combined early and late fusion stages. Our proposed method, MOAD-FNet, is aimed at multimodal brain tumor subtyping, and survival prediction in both lung and breast cancer. In the following subsections, we highlight key components of the MOAD-FNet framework.

[2]https://github.com/mahmoodlab/SurvPath/tree/main/splits/5foldcv/

1) Omic Encoder: To construct the omic encoder, we followed the established practice of using a Self-Normalizing Neural (SNN) [35] consisting of two fully connected layers, where each layer applies an Exponential Linear Unit (ELU) activation function followed by Alpha Dropout (0.25). The SNN compresses the 8K CpG sites into an encoded omic feature $\mathbf{o}_i \in \mathbb{R}^{d^o}$ $d^o = 256$. Note that for fair comparison with other methods [13], [18], [20], [22], we used the same encoder tokenizing the genes into a group of pathways.

2) Whole Slide Image Encoder: For a WSI $\mathbf{X}$, we extract a collection of patches, represented as $\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{ij}$, where i represents the patient/slide index and j indicates the patch index that varies across slides. We extract non-overlapping patches from tissue areas at a 20× magnification (about 0.5 µm/pixel resolution). Subsequently, we utilized a SOTA image-only encoder [11] UNI to obtain the patch embeddings $\mathbf{e}_{ij}$. Using UNI $f_{enc}(.)$, we derive a set of low-dimensional patch embeddings for each patient, where $\mathbf{e}_{ij} = f_{enc}(\mathbf{x}_{ij}) \in \mathbb{R}^{d^e}$, $d^e = 1024$, serving as input to our pipeline.

3) Fusion Stages: The fusion scheme is the key contribution of this work. Our motivation sparks from the enhanced features obtained from the dense modeling conducted by multimodal early fusion methods in [10], [13], [22], [26]. Hence, we divide this subsection into two: early fusion and late fusion.

Early Fusion. Given a matrix of patch embeddings $\mathbf{e}_{ij}$ and the omic feature vector $\mathbf{o}_i$ for patient i, the encoded omic feature vector is cloned to match the number of patches in

the WSI. This results in a tensor of shape (N_i, d_o), where N_i is the number of patches in WSI i, and d_o is the dimension of the omic feature. The combined feature set for WSI i is represented as:

$$\mathbf{z}_{ij} = [\mathbf{e}_{ij}, \mathbf{o}_i] \tag{2}$$

Here, $\mathbf{z}_{ij} \in \mathbb{R}^{d_e+d_o}$ represents the concatenated [.] feature vector for patch j in WSI i, and the resulting shape becomes $(N_i, d_e + d_o)$. Next, a Multilayer Perceptron (MLP) encoder, denoted as f_E, is applied to each concatenated patch embedding and omic feature pair to learn their joint representation:

$$\mathbf{p}_{ij} = f_E(\mathbf{z}_{ij}) \tag{3}$$

This operation is performed for all patches j of the ith WSI. By incorporating omic features $\mathbf{o}_i$ into patch embeddings via early fusion, we enrich the representation with complementary molecular information. Leveraging an Attention-based Deep Multiple Instance Learning (ABMIL) approach [36], we capture patch-level discriminative features that synergistically combine molecular and morphological insights, enabling more precise identification of critical regions. The resulting embedding is projected to a slide-level representation $\mathbf{v}_i \in \mathbb{R}^{256}$, dimensionally aligned with the original omic feature $\mathbf{o}_i$ to facilitate subsequent late fusion. In the survival prediction tasks where biological pathways were used, we followed the methodology in [13] by encoding the 331 pathways using the SNN. Each pathway was processed independently, and the resulting embeddings were stacked to form a comprehensive pathway representation. This representation was passed through an MLP to learn a unified omic embedding. The resulting vector was then concatenated with each WSI patch embedding to enable early fusion.

This MLP-based early fusion approach was chosen for its simplicity, computational efficiency, and scalability to large WSIs. Since each patch–omic pair is processed independently, the method scales linearly with the number of patches ($\mathcal{O}(N)$), making it well-suited for high-resolution slides that may contain tens of thousands of patches.

Late Fusion, motivated by [13] in modeling the interaction between omic to histology, histology to omic, and omic to omic, we mimic a similar behavior by employing our novel multimodal outer arithmetic fusion block (MOAB) [28] within a dual fusion approach, marking a new direction in combining MOAB with dual fusion to enhance modality integration. MOAB inputs ($\mathbf{v}_i$,$\mathbf{o}_i$) will be fused through four operations: outer product, outer division, outer subtraction, and outer addition. MOAB extracts various interactions while simultaneously preserving the $\mathbf{v}_i$ input feature by appending one to each input embedding when performing the outer product and division fusion, and zero in the case of the outer subtraction and addition fusion. To clarify, $\mathbf{v}_i \in \mathbb{R}^N$ and $\mathbf{o}_i \in \mathbb{R}^M$. We append a 1 to each of these column vectors, $\mathbf{v}_{i1} = [1; \mathbf{v}_i], \mathbf{o}_{i1} = [1; \mathbf{o}_i]$. Thus, vectors $\mathbf{v}_{1i} \in \mathbb{R}^{N+1}$ and $\mathbf{o}_{1i} \in \mathbb{R}^{M+1}$, where i indexes the sample. Then, their outer product is defined as:

$$(\mathbf{v}_{i1} \otimes \mathbf{o}_{i1})_{nm} = \mathbf{v}_{1i,n} * \mathbf{o}_{1i,m}, \tag{4}$$

where $n \in [1, N+1]$ and $m \in [1, M+1]$ index elements of the extended feature vectors. This yields a $(N+1) \times (M+1)$ matrix that intermingles every element of $\mathbf{v}_{i1}$ with every element of $\mathbf{o}_{i1}$. The appended 1 in both $\mathbf{v}_{i1}$ and $\mathbf{o}_{i1}$ ensures the original unimodal features $\mathbf{v}_i$ and $\mathbf{o}_i$ appear in the outer product matrix. Similarly, we define outer addition, subtraction, and division as:

$$(\mathbf{v}_{i0} \oplus \mathbf{o}_{i0})_{nm} = \mathbf{v}_{0i,n} + \mathbf{o}_{0i,m}, \tag{5}$$

$$(\mathbf{v}_{i0} \ominus \mathbf{o}_{i0})_{nm} = \mathbf{v}_{0i,n} - \mathbf{o}_{0i,m}, \tag{6}$$

$$(\mathbf{v}_{i1} \oslash \mathbf{o}_{i1})_{nm} = \mathbf{v}_{1i,n} \div (\mathbf{o}_{1i,m} + \epsilon), \tag{7}$$

where ϵ is a small number (set to 1e$-$10), and $\mathbf{v}_{i0} = [0; \mathbf{v}_i], \mathbf{o}_{i0} = [0; \mathbf{o}_i]$.

The four matrices produced by MOAB are concatenated along the channel dimension to form a multimodal tensor $\mathbf{L} \in \mathbb{R}^{4\times257\times257}$. We hypothesize that channel fusion will maintain the proximity of closer points and will use fewer parameters compared to a typical concatenation. By combining features across the channel dimension, we greatly decrease the dimension by compressing the feature representation from $(257\times257)^4$ to $(257)^2$. Following the same parameters in [28], a 2D convolution layer is subsequently performed to leverage associated interactions, resulting in a singular condensed multimodal feature tensor: $\mathbf{L}^* \in \mathbb{R}^{1\times257\times257}$. Last we flatten $\mathbf{L}^*$ and apply Leaky ReLU followed by a linear predictor for brain tumor subtyping prediction. MOAB provides simple yet effective operations to fuse multimodal data, similar to [13], but without relying on approximations. The use of four arithmetic operations is intended to capture diverse types of feature interactions between modalities. While the operations are not based on specific biological priors, they are empirically motivated to enable the model to explore diverse types of interactions across modalities. We speculate that these operations capture complementary interactions, namely, addition amplifies signals when both modalities activate similarly, subtraction highlights modality-specific differences, multiplication encodes how features co-vary in magnitude and sign, while division introduces a notion of relative scale between features. Their combination enables MOAB to learn rich and complementary interactions across the fused representation space. This finding also aligns with observations made in our prior work [37], where the four arithmetic operations conclusively demonstrated optimal performance in multimodal fusion tasks.

4) Interpretability of WSI and DNA Methylation Modalities:

a) WSI Interpretability via Attention Maps.: Employing ABMIL [18] on the post-fusion embedding $\mathbf{p}_{ij}$ enables the visualization of attention scores, which provide an enriched perspective on the tumor's visual and omics characterization. This, in turn, enhances ABMIL's capacity to prioritize patches that are biologically relevant. We define our gated attention

computation as follows:

$$\begin{aligned} \mathbf{h}_{ij} &= \tanh(\mathbf{W}_p \mathbf{p}_{i,j} + \mathbf{b}_p) \odot \sigma(\mathbf{W}_g \mathbf{p}_{i,j} + \mathbf{b}_g), \\ a_{ij} &= \frac{\exp(\mathbf{w}^T \mathbf{h}_{ij})}{\sum_1^j \exp(\mathbf{w}^T \mathbf{h}_{ij})}, \\ \mathbf{v}_i &= f_\rho \left(\sum_{j=1}^{j} \mathbf{a}_{ij} \mathbf{h}_{ij} \right), \end{aligned} \tag{8}$$

Here, $\mathbf{W}_p$ and $\mathbf{W}_g$ are learnable parameters for the input and gating functions, respectively, with $\mathbf{b_p}$ and $\mathbf{b}_g$ as their corresponding biases. The symbol $\odot$ denotes element-wise multiplication, and σ represents the sigmoid activation function. The gated embedding for patch j in slide i is given by $\mathbf{h}_{ij}$, while $\mathbf{a}_{ij}$ represents the attention weight for patch j, normalized across all patches. We leverage $\mathbf{a}_{ij}$ at inference to generate the heatmap shown in Fig. 5. To further enhance our late fusion stage, we weight $\mathbf{v}_i$ with $\mathbf{a}_{ij}$ using an MLP f_ρ which consists of a linear transformation, an activation function (ReLU), and dropout. f_ρ is applied to the pooled embedding $\mathbf{v}_i$, increasing its representational power, making it a more refined input for the late fusion stage. This enriched embedding encapsulates a high-level representation of the WSI, effectively integrating morphological and molecular insights from the attention mechanism.

b) DNA Methylation Interpretability via SHAP: While WSI features lend themselves to spatial interpretation via attention heatmaps, DNA methylation data contribute non-spatial but highly discriminative molecular context. In our dual fusion architecture, the 256-dimensional DNA methylation embedding ($\mathbf{o}_i$) is concatenated with each patch feature and passed through an additional MLP before being aggregated by AB-MIL. This multi-stage transformation prevents direct attribution of predictions to individual CpG sites within the fused model. To address this, we assess the interpretability of the DNA methylation modality using SHAP (SHapley Additive exPlanations) applied to the standalone DNA methylation classifier. This model uses the same 256-dimensional embedding ($\mathbf{o}_i$) produced by the initial SNN encoder used in the fusion setup, allowing for consistency in representation. SHAP allows for both local and global interpretability by assigning attribution scores to each input feature (i.e., CpG site) relative to the model's output. Local explanations reveal which CpG sites contribute most to individual predictions, while global aggregation across samples highlights features with consistent predictive influence. This approach supports the identification of biologically meaningful methylation patterns and provides insights into the molecular mechanisms underlying different tumor subtypes.

III. Experiments

A. Performance Metrics

For the subtyping task, we assessed performance utilizing various metrics: F1-Macro, F1-Micro, Precision, and Recall/Sensitivity. F1-Macro is a significant metric for our quantitative analysis as it independently computes the F1 score for each class and subsequently averages them, assigning equal weight to each class irrespective of its size, thereby ensuring that the performance of minority classes is not overwhelmed by that of majority classes. For the survival prediction task, we follow the implementations of [13], [20], [22] where survival analysis is defined as an estimation of the probability of an event occurring within a given survival time, while accounting for right-censored data. Censorship status is represented as $c \in \{0, 1\}$, where $c = 0$ denotes an observed event (e.g., death) and $c = 1$ indicates the patient's last known follow-up. In line with previous work we discretize the time-to-event into non-overlapping time intervals $(t_{i-1}, t_i]$, based on the quartiles of survival times denoted as y_i. This formulation transforms the problem into a classification task with censorship information, where each patient is represented by $(\mathbf{L}_{\text{logits}}, y_i, c)$.

Next, we use the dual fusion embedding generated by MOAB-FNet, $\mathbf{L}_{\text{logits}}$ to predict the discretized bin corresponding to a time interval t_i. We define the discrete hazard function as:

$$f_{\text{hazard}}(y_i \mid \mathbf{L}_{\text{logits}}) = \sigma(y_i),$$

where σ is the sigmoid activation function. Intuitively, $f_{\text{hazard}}(y_i \mid \mathbf{L}_{\text{logits}})$ represents the probability that the patient experiences the event (e.g., death) during the interval $(t_{i-1}, t_i]$. The discrete survival function is then defined as:

$$f_{\text{surv}}(y_i \mid \mathbf{L}_{\text{logits}}) = \prod_{k=1}^{i-1} \left(1 - f_{\text{hazard}}(y_k \mid \mathbf{L}_{\text{logits}})\right),$$

which represents the probability that the patient survives up to the interval $(t_{i-1}, t_i]$. The Negative Log-Likelihood (NLL) survival loss is formally defined as:

$$\begin{aligned} \mathcal{L}\left(\{\mathbf{L}_{\text{logits}}, y, c\}_{i=1}^{N_{\text{total}}}\right) = -\sum_{i=1}^{N_{\text{total}}} \Big[& c_i \log\left(f_{\text{surv}}(y_i \mid \mathbf{L}_{\text{logits}})\right) \\ & + (1 - c_i) \log\left(f_{\text{surv}}(y_i \mid \mathbf{L}_{\text{logits}}) - 1[y_i]\right) \\ & + (1 - c_i) \log\left(f_{\text{hazard}}(y_i \mid \mathbf{L}_{\text{logits}})\right) \Big], \end{aligned} \tag{9}$$

where (N_{total}) is the total number of samples in the dataset, and (k) corresponds to the total number of discretized labels.

B. Training Configuration

We recognize the challenges posed by class imbalance and limited sample sizes in rare CNS tumor subtypes. To ensure the stability and reliability of our subtyping model, we applied several targeted strategies:

- **Multiple runs with random seeds:** To further evaluate the stability of MOAD-FNet, we initially conducted multiple runs using different random seeds with a 75/25 training–testing split, strictly stratified by patient ID to avoid data leakage. The resulting macro F1-scores across runs were 79, 74, and 73; micro F1-scores were 88, 87, and 82. However, we observed inflated performance, particularly in rare classes (e.g., those with only 11 total samples), where the test set could include as few as two instances. In such cases, correctly classifying both can disproportionately boost metrics, while misclassifications have limited impact, potentially skewing the evaluation. We have also conducted multiple runs with different random seeds using the 50/50 setup and

observed minimal variation across runs. Based on these findings, we adopted a 2-fold cross-validation strategy to ensure greater stability and fairness in performance reporting, especially for underrepresented subtypes.

- **Training protocols across tasks:** MOAD-FNet was trained separately on BLCA and BRCA for the survival task, and on the NHNN BRAIN UK dataset for the subtyping task. Each task used its own training setup and loss function, ensuring a fair comparison with task-specific baselines and demonstrating the versatility of our dual fusion strategy.
- **Parameter efficiency and model size control:** MOAB achieves parameter efficiency by replacing fully connected fusion layers with structured arithmetic operations. While the arithmetic component is parameter-free, the full MOAB block includes a lightweight channel fusion step implemented via a 1×1 convolution, followed by a linear projection. These layers enable the model to learn meaningful cross-modal interactions with only a moderate increase in parameter count.
- **Early fusion architecture:** Each 1024-dimensional WSI patch embedding was concatenated with a 256-dimensional omic embedding and passed through an MLP consisting of two linear layers (each with 1024 units), followed by layer normalization and dropout ($p = 0.1$), enabling effective cross-modal interaction without added complexity.
- **Data augmentation:** For WSIs, we employed random horizontal and vertical flips, and color jittering (brightness, contrast, saturation, hue) to boost model generalization and robustness to input variation.
- **Regularization techniques:** To enhance generalization, we applied dropout and weight decay. Specifically, we used a dropout rate of 0.2 with the Adam optimizer configured at a learning rate of 1e-5 and L1 regularization (weight decay) of 1e-4. Within the MOAB fusion block, LeakyReLU activation with a dropout of 0.2 further contributed to model robustness.
- **Hyperparameter settings:** All models were trained for 50 epochs, as training performance stabilized and test set results remained consistent across both folds, indicating convergence without overfitting. Key hyperparameters included dropout (0.2), hidden dimensions (256 for ABMIL and 512 for MOAB), and batch size (1). The learning rate was selected empirically: higher values (e.g., 1e-3 and 1e-2) led to overfitting, while 1e-5 showed more stable convergence.

C. Domain-Specific Prior Work and Ablation Studies

To evaluate our method on subtyping and survival prediction tasks, we replicate and adapt recent SOTA methods by incorporating MOAB as a replacement for the fusion technique initially used in these methods. For the survival prediction task, we conducted a comparative analysis using a consistent feature extractor across all modalities, including WSI and omic data, utilising the recent TCGA ID samples from [13]. We uniformly implemented training hyperparameters and loss functions across all models displayed in Tables I and III. To this end, this section is divided into two parts. SOTA baseline models detailing unimodal and fusion models employed for each task, together with a description of the ablation studies conducted.

1) State-of-the-Art Baseline Models: We evaluate our approach against SOTA techniques in the unimodal and multimodal setting for both subtyping and survival prediction tasks.

Unimodal Subtyping baselines. We utilize MLP and SNN [35] as baseline models for the DNA methylation data. For WSIs, we utilize Attention-based Multiple Instance Learning (ABMIL) [18], which implements gated weighted attention pooling to determine the importance of patches, as well as Transformer-based Multiple Instance Learning (TransMIL) [34] which employs the Nyström attention mechanism to evaluate correlations among WSI patches. Through rigorous experimentation, we found that ABMIL outperforms the TransMIL baseline in both performance and computational efficiency, also noted in [38], making it the preferred baseline for MOAD-FNet.

Multimodal Subtyping baselines. We adapt Attention Challenging Multiple Instance Learning (ACMIL) [23] to work in a multimodal setting. ACMIL is an approach designed to address overfitting in single-modality WSI classification by using multiple attention branches and a composite loss (cross-entropy + diversity loss) to distribute attention across the WSI. We also used TransMIL with two late fusion variants, concatenation (Cat) and Kronecker product (KP) [34]. Furthermore, we compare against MCAT [10] and SurvPath [13], both of which perform multimodal tokenization to extract histology and biological pathway tokens. MCAT and SurvPath employ a transformer-based early fusion approach, which concatenates the resulting vectors. To maintain their architectural consistency, we opted not to replace concatenation with the MOAB late fusion block, as their complex architectures could be destabilized by it. However, to test our dual fusion hypothesis, we input our early fusion input representation $\mathbf{p}_{ij}$ into MCAT and SurvPath, thus enabling these models to leverage omic-WSIs embeddings at both the early and late fusion stages. By comparing the performance of SurvPath and MCAT with $\mathbf{p}_{ij}$ to the original implementation (SurvPath* and MCAT*), we provide evidence that dual fusion outperforms single-stage fusion methods.

Unimodal Survival baselines. In addition to the unimodal subtyping baselines, we employ Sparse-MLP [13], which tokenizes transcriptomics into biological pathway tokens encoding specific cellular functions for the downstream analysis.

Multimodal Survival baselines. We integrated MOAD-FNet across the same models used for subtyping tasks, adding the prototypical information bottlenecking and disentangling (PIBD) method [20]. It introduces a disentanglement mechanism to separate modality-specific versus shared information. For PIBD, we restricted MOAB to a late fusion setting, respecting PIBD's initial modality-specific separation.

2) Ablation studies: We conducted three ablation experiments to comprehensively evaluate MOAD-FNet. We first removed the MOAB fusion block, using only ABMIL. Here, $\mathbf{p}_{ij}$ served as the input to ABMIL, and the resulting feature embedding $\mathbf{v}_i$ was directly fed to the classifier layer to assess the distinct impact of both ABMIL and the DNA modality. In the second experiment, we removed the early fusion block (illustrated in Fig. 1), making $\mathbf{e}_{ij}$ the input to the gated attention block, which is then followed by MOAD-FNet. Last, we employed a task-agnostic encoder ConvNext.v2 [39], pretrained on ImageNet, to extract features from the WSI and tested MOAD-FNet with this setup. For the survival prediction task, we evaluated the MOAD-FNet approach using the two most common baseline fusion models: ABMIL [36] and TransMIL [34]. We used the late fusion techniques: concatenation and Kronecker product and compared these against MOAB within both fusion settings.

D. Data and Code Availability

For brain tumor subtyping, we obtained data from NHNN through a rigorous application process to BRAIN UK, securing anonymized H&E slides of tissue samples and epigenetic data. For survival prediction, WSIs are publicly available through the TCGA repository, with corresponding omic data from [13]. Source code will be made available upon acceptance.

IV. Results

Subtyping results. In Tables I and II, we present the results for the subtyping task on the NHNN BRAIN UK dataset. MOAD-FNet integrated with ABMIL consistently demonstrates superior performance in brain tumor subtyping. Specifically, Table I shows that the SNN model performs well across metrics using omics-only data, achieving an F1-Macro score of $0.726(\pm 0.003)$, marginally outperforming the MLP. In contrast, WSI-only models show relatively low performance across metrics, with ABMIL achieving an F1-Macro of just $0.247(\pm 0.004)$ and TransMIL performing slightly worse at $0.217(\pm 0.019)$ likely due to the issue of indistinguishable patterns between most subtypes and the presence of underrepresented rare classes. These results indicate that omics data alone offers strong predictive power, however further improvements and interpretability are limited without incorporating WSI data. On the other hand, multimodal models substantially outperformed unimodal ones. For instance, ABMIL-MOAD-FNet achieved the best scores across all metrics, with an F1-Macro of $0.745(\pm 0.025)$ and an F1-Micro of $0.820(\pm 0.031)$. This represents an improvement of 0.027 in F1-Macro compared to the second-best multimodal model, TransMIL-MOAD-FNet, further emphasizing the effectiveness of the MOAD-FNet architecture in leveraging multimodal data for richer and more informative representations. To assess statistical significance, we performed a Wilcoxon rank-sum test [40] comparing the F1-Macro scores of ABMIL-MOAD-FNet with all other multimodal models in Table I. The test yielded a p-value of 0.043, indicating a statistically significant difference at the 0.05 significance level. These results demonstrate that integrating MOAD-FNet with advanced multimodal architectures leads to significant performance gains over single-stage fusion methods. This underscores the critical role of dual fusion strategies in effectively combining complementary features from multiple modalities, ultimately driving superior predictive accuracy and robustness.

TABLE I

Subtyping prediction Results on the NHNN UK BRAIN dataset. We show MOAD-FNet combined with SOTA baseline models. The best performance is highlighted in **bold**. Concatenation and Kronecker products are denoted (Cat) and (KP). The gray rows correspond to models integrating MOAD-FNet or using the early fusion embeddings $\mathbf{p_{ij}}$. These demonstrate improved performance compared to standalone baseline models.

Model	F1-Macro	F1-Micro	Precision	Recall
Omics				
SNN	0.726 (0.003)	0.819 (0.013)	0.799 (0.015)	0.715 (0.023)
MLP	0.690 (0.012)	0.794 (0.021)	0.741 (0.018)	0.684 (0.031)
WSI				
ABMIL	0.247 (0.004)	0.442 (0.004)	0.280 (0.022)	0.026 (0.250)
TransMIL	0.217 (0.019)	0.434 (0.003)	0.240 (0.024)	0.230 (0.013)
Multimodal				
ACMIL MOAD-FNet	0.501 (0.011)	0.729 (0.031)	0.602 (0.032)	0.490 (0.014)
TransMIL (Cat)	0.711 (0.005)	0.799 (0.005)	0.744 (0.003)	0.707 (0.002)
TransMIL (KP)	0.483 (0.003)	0.741 (0.016)	0.485 (0.007)	0.506 (0.008)
TransMIL MOAD-FNet	0.724 (0.005)	0.816 (0.002)	0.759 (0.009)	0.719 (0.001)
MCAT	0.402 (0.046)	0.661 (0.021)	0.408 (0.045)	0.414 (0.039)
MCAT p_{ij}	0.432 (0.015)	0.703 (0.017)	0.441 (0.031)	0.446 (0.011)
SURVPATH	0.424 (0.008)	0.697 (0.012)	0.478 (0.038)	0.425 (0.013)
SURVPATH p_{ij}	0.531 (0.008)	0.761 (0.008)	0.632 (0.005)	0.520 (0.007)
ABMIL (Cat)	0.718 (0.013)	0.806 (0.002)	0.764 (0.018)	0.717 (0.023)
ABMIL (KP)	0.447 (0.024)	0.737 (0.020)	0.481 (0.043)	0.464 (0.032)
ABMIL MOAD-FNet	**0.745** (0.025)	**0.820** (0.013)	**0.769** (0.016)	**0.745** (0.035)

Qualitative results shown in Fig. 7 further illustrate MOAD-FNet's effectiveness in classifying most tumor subtypes, showcasing the strong impact of its intermingled features. To assess class separation accuracy in the t-SNE representations, we calculated silhouette scores for early and late fusion, as they displayed similar patterns. The late fusion t-SNE achieved a silhouette score of 0.33, while MOAD-FNet scored 0.37, indicating that it provides more accurate class separation. This is also evident in the box plot in Fig. 6B, where MOAD-FNet displays fewer low outliers and a more compact distribution than early and late fusion.

In Fig. 7A, we show the confusion matrix corresponding to our model MOAD-FNet, while in Fig. 7B and C, we see the results for late and early fusion, respectively. These demonstrate that our dual-fusion MOAD-FNet method excels in classifying subtypes, especially for minority classes. Notably, classes 4, 9, and 13, which was misclassified with late fusion (Fig. 7B), were correctly classified by MOAD-FNet. Conversely, late fusion performed better for classes 2 and 16, highlighting the challenges of handling heterogeneous multimodal data. This also underscores the need for tailored fusion techniques to address different data complexities. Furthermore, Fig. 6C1 and C2 illustrate that MOAD-FNet's dual fusion approach considerably enhances classification of all glioblastoma and diffuse glioma subtypes, even in cases with highly overlapping morphological features.

We present ablation results in Table II. The findings highlight the necessity for advanced fusion techniques to effectively integrate complementary features across modalities. Notably, the full MOAD-FNet model achieved the highest F1-Macro score of 0.745 (± 0.035), demonstrating a notable improvement over both early fusion, with a gain of 0.101, and

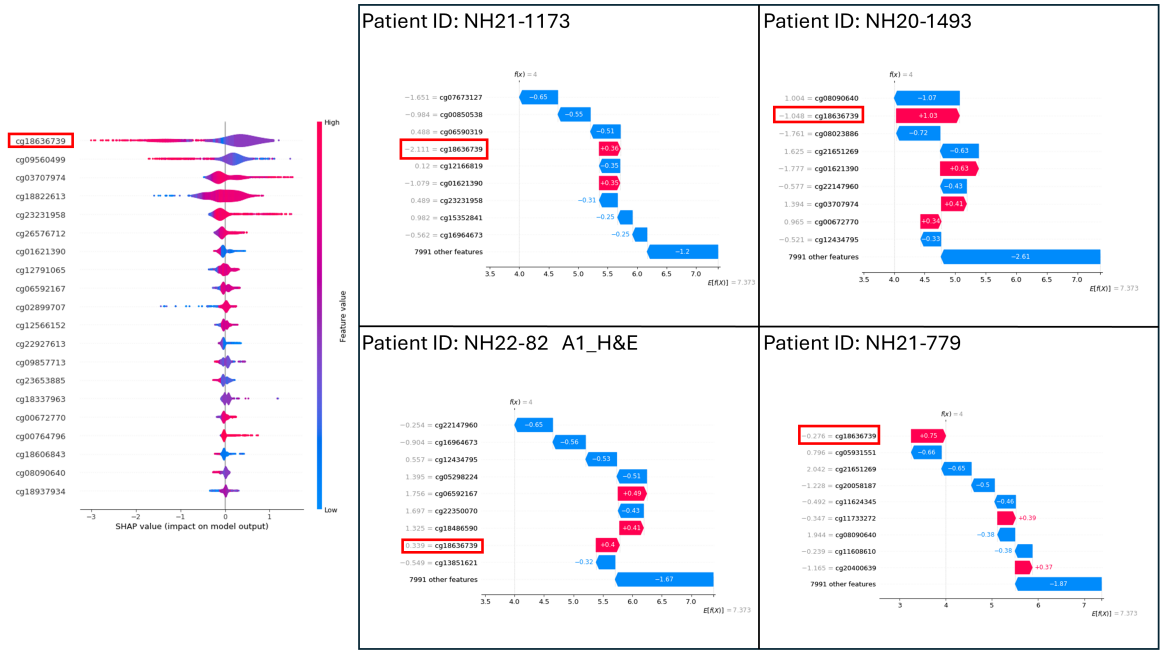

Fig. 3. Per-patient SHAP explanations for four individuals predicted as “high-grade diffuse glioma of the midline/posterior fossa; H3/IDH-wildtype.” Shared influential CpG features (e.g., `cg18636739`) suggest consistency in feature importance across patients.

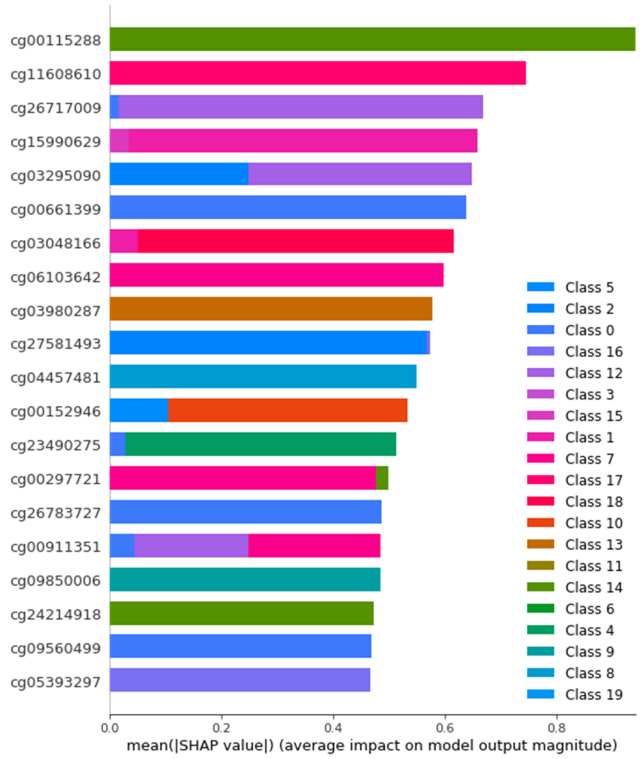

Fig. 4. Top 20 CpG sites ranked by average SHAP value across all subtypes. Each bar is colored by the class in which the feature had the greatest impact.

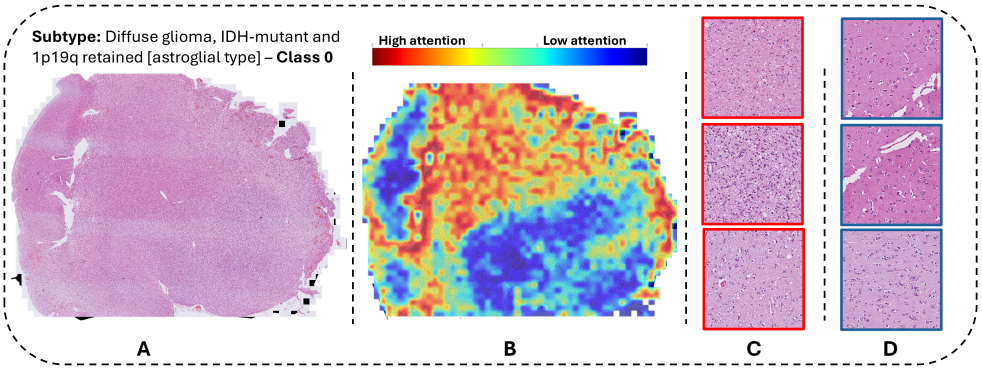

Fig. 5. Visual representation of attention heatmap generated by MOAD-FNet for a diffuse glioma, IDH-mutant and 1p19q-retained (astroglial type) tumor (Class 0). (A) The original histology slide is displayed. (B) The heatmap shows areas of high attention (red) and low attention (blue), with regions of diagnostic relevance highlighted. (C) Representative patches with high attention are bordered in red, potentially indicating hallmark features of astroglial differentiation and cellular atypia crucial for diagnosis. (D) Representative patches with low attention are bordered in blue, reflecting regions of low tumor infiltration. The color bar illustrates the attention scale from high (red) to low (blue).

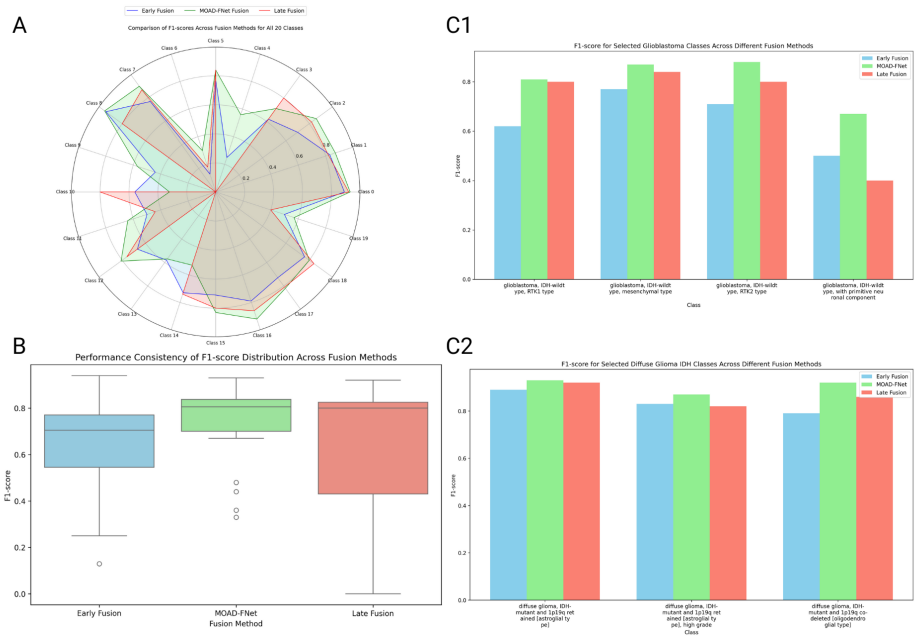

Fig. 6. Comparison of F1-scores across different fusion methods for glioma classification. **(A)** Radar chart illustrating F1-score performance across all 20 glioma subtypes, comparing Early Fusion, MOAD-FNet, and Late Fusion methods. **(B)** Box plot summarizing F1-score distributions, highlighting the variability and consistency of each fusion method. **(C1)** Bar plot showing F1-score comparisons for selected *glioblastoma* subtypes across the three fusion methods. **(C2)** Bar plot showing F1-score comparisons for selected *diffuse glioma, IDH-mutant* subtypes across the same fusion methods.

late fusion, with a gain of 0.055. Surprisingly, the task-agnostic ConvNeXt encoder achieved an F1-Macro of0.732 (± 0.012), marginally lower than the pretrained UNI encoder. This suggests that the core improvement in the full model stems not from the backbone architecture but from the dual fusion mechanism that effectively integrates complementary features across modalities. To further interpret the DNA methylation modality, we analyzed feature attributions generated from the standalone methylation classifier using the SHAP. Fig. 3 shows SHAP plots for four patients from Class 4, where CpG site `cg18636739` consistently appeared as a top contributor. This suggests a stable, subtype-specific signal learned by the model. We confirmed this by aggregating SHAP values across all Class 4 cases, where the same CpG remained the top-ranked feature.

Fig. 4 summarizes the top 20 CpG sites with the highest SHAP values across all subtypes, highlighting subtype-specific patterns used by the model.

Additionally, we investigated the biological context of `cg18636739`, which was highlighted by SHAP analysis. According to available annotations, `cg18636739` is located within the *CSDA* gene (also known as *ZNF9*) [41], situated on chromosome 12 [42]. *CSDA* encodes a zinc finger protein involved in RNA binding and gene regulation. Although `cg18636739` itself has not been extensively studied, its location within a gene associated with transcriptional control and cancer progression suggests that its prominence may reflect relevant epigenetic regulation [43], [44]. These findings

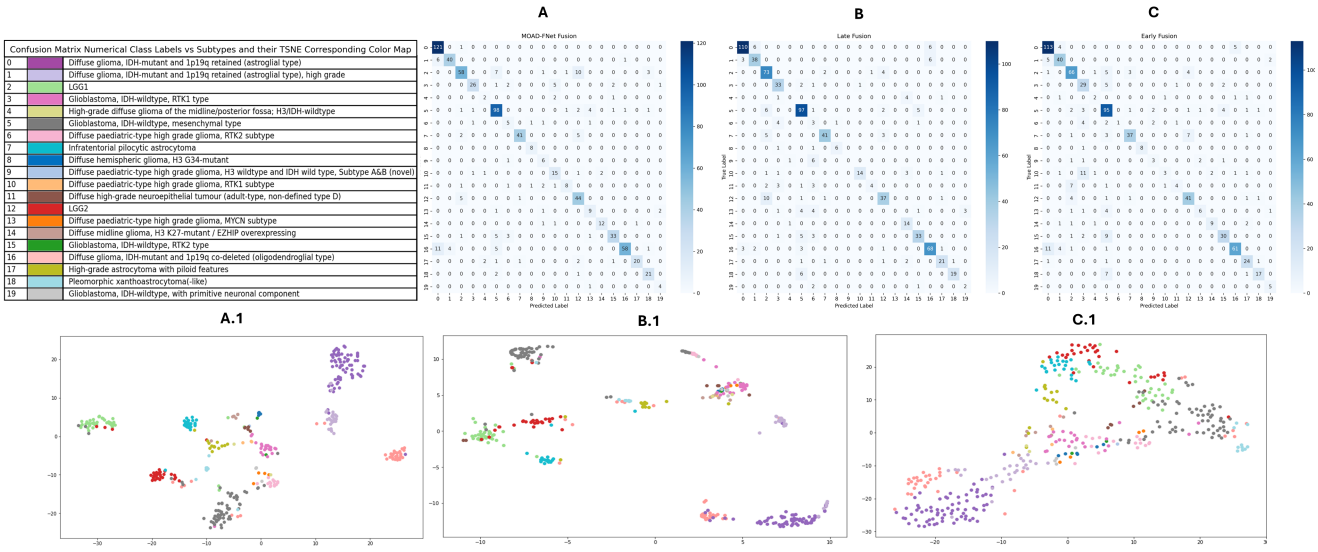

Fig. 7. Comparison of confusion matrices and t-SNE visualizations for three fusion strategies: (A) MOAD-FNet, (B) Late Fusion, and (C) Early Fusion for brain tumor subtyping. The corresponding t-SNE plots are labeled as (A.1), (B.1), and (C.1), respectively.

support the utility of DNA methylation data in enhancing both the accuracy and interpretability of CNS tumor subtyping.

TABLE II

SUBTYPING PREDICTION ABLATION STUDY OF MOAD-FNET SHOWING THE PERFORMANCE OF EARLY, LATE, AND DUAL FUSION METHODS USING THE NHNN BRAIN UK DATASET.

Ablation	Model Description	F1-Macro
ABMIL	Early fusion with $\mathbf{p}_{ij}$	0.644 (0.010)
ABMIL - MOAB	Late fusion with $\mathbf{e}_{ij}$	0.690 (0.030)
MOAD-FNet	ConvNeXt encoder	0.732 (0.012)
MOAD-FNet	**Full model**	0.745 (0.035)

Survival prediction results. The results shown in Table III demonstrate that integrating MOAD-FNet with existing SOTA methods (indicated by *) derives consistent performance improvement compared to baseline models. For instance, our approach achieves the highest c-index of $0.691(\pm 0.069)$ for BLCA and $0.726(\pm 0.049)$ for BRCA, surpassing all other methods in predicting patient disease-specific survival for BLCA while performing on par with the top-performing model MMP [22] for BRCA. Interestingly, the best results on BRCA were obtained by PIBD with MOAB, achieving a strong c-index of $0.749(\pm 0.062)$ [22].

Note that we did not test MOAD-FNet with MMP [27] because MMP already incorporates two early fusion stages (transformer and optimal transport). Adding MOAB would bring the total to four fusion stages, potentially introducing additional complexity and noise without clear performance benefits.

To further evaluate the effectiveness of MOAD-FNet in the survival prediction task, we conducted additional ablation studies presented in Table IV. The results demonstrate that the ABMIL model with Dual Fusion (DF) consistently delivers superior performance across both the BRCA and BLCA datasets, particularly when paired with the MOAB aggregation method. For the BRCA dataset, ABMIL-DF with MOAB achieves the

TABLE III

SURVIVAL PREDICTION RESULTS OF MOAD-FNET WITH BASELINES FOR PREDICTING PATIENT DISEASE-SPECIFIC SURVIVAL USING THE C-INDEX. THE BEST PERFORMANCE IS HIGHLIGHTED IN **BOLD**, AND THE SECOND-BEST PERFORMANCE IS UNDERLINED. OVER FIVE RUNS, THE STANDARD DEVIATION IS REPORTED IN BRACKETS. METHODS MARKED * ARE RE-IMPLEMENTED.

Model	BLCA (↑)	BRCA (↑)
WSI		
ABMIL* [36]	0.572 (0.084)	0.573 (0.097)
TransMIL* [34]	0.579 (0.052)	0.611 (0.011)
Omics		
MLP*	0.660 (0.060)	0.569 (0.084)
SNN*	0.671 (0.058)	0.574 (0.011)
S-MLP*	0.658 (0.053)	0.598 (0.014)
Multimodal		
PIBD [20]	0.667 (0.061)	0.736 (0.072)
PIBD* - MOAB	0.684 (0.046)	<u>0.749</u> (0.062)
MMP [22]	0.628 (0.064)	**0.753** (0.096)
MMP [22]	0.635 (0.064)	0.738 (0.096)
ACMIL* - MOAD-FNet [23]	0.658 (0.068)	0.661 (0.082)
TransMIL* - MOAD-FNet	0.661 (0.053)	0.675 (0.068)
SurvPath [13]	0.625 (0.056)	0.655 (0.089)
SurvPath* - p_{ij}	0.660 (0.047)	0.665 (0.006)
MBFusion [19]	−−	0.644 (0.020)
ED-GNN [24]	−−	0.672 (0.059)
MoME [45]	<u>0.686</u> (0.041)	−−
MuGI [46]	0.681 (0.056)	−−
ABMIL MOAD-FNet(Ours)	**0.691** (0.069)	0.726 (0.049)

TABLE IV

SURVIVAL PREDICTION ABLATION. COMPARISON OF C-INDEX FOR TGGA BRCA AND BLCA DATASETS USING ABMIL AND TRANSMIL MODELS WITH TWO FUSION STAGES, LATE (LF) AND DUAL (DF), ACROSS THREE AGGREGATION METHODS.

BRCA c-index (↑)				
Model	**Fus-S**	**Cat**	**KP**	**MOAB**
TransMIL	LF	0.606 (0.064)	0.602 (0.089)	0.620 (0.115)
	DF	0.623 (0.032)	0.645 (0.067)	0.675 (0.068)
ABMIL	LF	0.625 (0.075)	0.616 (0.090)	0.711 (0.095)
	DF	0.634 (0.023)	0.640 (0.058)	**0.726** (0.049)
BLCA c-index (↑)				
Model	**Fus-S**	**Cat**	**KP**	**MOAB**
TransMIL	LF	0.599 (0.081)	0.582 (0.053)	0.600 (0.108)
	DF	0.624 (0.068)	0.595 (0.052)	0.661 (0.053)
ABMIL	LF	0.558 (0.082)	0.571 (0.063)	0.677 (0.054)
	DF	0.622 (0.063)	0.561 (0.054)	**0.691** (0.069)

highest c-index of $0.726(\pm 0.049)$, outperforming TransMIL-DF with MOAB. Similarly, for the BLCA dataset, ABMIL-DF with MOAB achieves the highest c-index of $0.691(\pm 0.069)$, surpassing TransMIL-DF with MOAB. The Late Fusion (LF) results follow a similar pattern, with ABMIL consistently outperforming TransMIL across all aggregation methods. These findings underscore the effectiveness of ABMIL, particularly with the Dual Fusion strategy and MOAB aggregation.

Limitations. While our method demonstrates promising performance across multiple tasks, it's important to acknowledge some limitations. MOAD-FNet faces two primary challenges: first, balancing early and late fusion presents a trade-off; although MOAD-FNet leverages both stages to capture the strengths of each, finding the optimal balance can be complex, with some cases exhibiting performance variations depending on the fusion stage. Second, our late fusion block incorporating MOAB operates on latent space vectors derived from the early fusion stage, where omics features have already been blended. Consequently, identifying the specific CpG feature with the most profound impact on the outcome becomes challenging.

V. Conclusion

The MOAD-FNet framework advances CNS tumor subtyping by effectively integrating DNA methylation and WSI data through a novel dual fusion approach. Our framework addresses the under-explored potential of combining these modalities, with the potential for distinguishing subtle morphological differences between tumor subtypes from DNA methylation profiling. The early fusion stage, implemented through an MLP-based mapping of WSI and methylation features, enables interpretable visualization while maintaining low dimensionality. The late fusion stage, enriched by outer arithmetic operations with MOAB, captures complex inter-modal relationships beyond simple addition. MOAD-FNet demonstrates superior performance across all evaluation metrics and exhibits robust scalability with different architectures. The framework's consistent success in both tumor subtyping and survival prediction establishes its practical utility for precision oncology. These results highlight how integration of multiple data modalities can enhance diagnostic accuracy while preserving clinical interpretability, thereby advancing automated approaches to CNS tumor classification.

References

[1] A. Lopomo and F. Coppedè, "Epigenetic signatures in the diagnosis and prognosis of cancer," in *Epigenetic Mechanisms in Cancer*. Elsevier, 2018, pp. 313–343.

[2] W.-W. Liang, R. J.-H. Lu, R. G. Jayasinghe, S. M. Foltz, E. Porta-Pardo, Y. Geffen, M. C. Wendl, R. Lazcano, I. Kolodziejczak, Y. Song *et al.*, "Integrative multi-omic cancer profiling reveals dna methylation patterns associated with therapeutic vulnerability and cell-of-origin," *Cancer cell*, vol. 41, no. 9, pp. 1567–1585, 2023.

[3] R. Drexler, F. Brembach, J. Sauvigny, F. L. Ricklefs, A. Eckhardt, H. Bode, J. Gempt, K. Lamszus, M. Westphal, U. Schüller *et al.*, "Unclassifiable cns tumors in dna methylation-based classification: clinical challenges and prognostic impact," *Acta Neuropathologica Communications*, vol. 12, no. 1, p. 9, 2024.

[4] L. Djirackor, S. Halldorsson, P. Niehusmann, H. Leske, D. Capper, L. P. Kuschel, J. Pahnke, B. J. Due-Tønnessen, I. A. Langmoen, C. J. Sandberg *et al.*, "Intraoperative dna methylation classification of brain tumors impacts neurosurgical strategy," *Neuro-Oncology Advances*, vol. 3, no. 1, p. vdab149, 2021.

[5] J. C. Pickles, A. R. Fairchild, T. J. Stone, L. Brownlee, A. Merve, S. A. Yasin, A. Avery, S. W. Ahmed, O. Ogunbiyi, J. G. Zapata *et al.*, "Dna methylation-based profiling for paediatric cns tumour diagnosis and treatment: a population-based study," *The lancet child & adolescent health*, vol. 4, no. 2, pp. 121–130, 2020.

[6] Z. Jaunmuktane, D. Capper, D. T. Jones, D. Schrimpf, M. Sill, M. Dutt, N. Suraweera, S. M. Pfister, A. von Deimling, and S. Brandner, "Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre," *Acta neuropathologica communications*, vol. 7, pp. 1–18, 2019.

[7] H. L. Smith, N. Wadhwani, and C. Horbinski, "Major features of the 2021 who classification of cns tumors," *Neurotherapeutics*, vol. 19, no. 6, pp. 1691–1704, 2022.

[8] R. J. Chen, M. Y. Lu, D. F. Williamson, T. Y. Chen, J. Lipkova, Z. Noor, M. Shaban, M. Shady, M. Williams, B. Joo *et al.*, "Pan-cancer integrative histology-genomic analysis via multimodal deep learning," *Cancer Cell*, vol. 40, no. 8, pp. 865–878, 2022.

[9] D. Capper, D. T. Jones, M. Sill, V. Hovestadt, D. Schrimpf, D. Sturm, C. Koelsche, F. Sahm, L. Chavez, D. E. Reuss *et al.*, "Dna methylation-based classification of central nervous system tumours," *Nature*, vol. 555, no. 7697, pp. 469–474, 2018.

[10] R. J. Chen, M. Y. Lu, W.-H. Weng, T. Y. Chen, D. F. Williamson, T. Manz, M. Shady, and F. Mahmood, "Multimodal co-attention transformer for survival prediction in gigapixel whole slide images," in *Proceedings of the IEEE/CVF ICCV*, 2021, pp. 4015–4025.

[11] R. J. Chen, T. Ding, M. Y. Lu, D. F. Williamson, G. Jaume, A. H. Song, B. Chen, A. Zhang, D. Shao, M. Shaban *et al.*, "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, vol. 30, no. 3, pp. 850–862, 2024.

[12] D.-T. Hoang, E. D. Shulman, R. Turakulov, Z. Abdullaev, O. Singh, E. M. Campagnolo, H. Lalchungnunga, E. A. Stone, M. P. Nasrallah, E. Ruppin *et al.*, "Prediction of dna methylation-based tumor types from histopathology in central nervous system tumors with deep learning," *Nature Medicine*, pp. 1–10, 2024.

[13] G. Jaume, A. Vaidya, R. J. Chen, D. F. Williamson, P. P. Liang, and F. Mahmood, "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 (CVPR)*, 2024, pp. 11 579–11 590.

[14] D. Capper, D. Stichel, F. Sahm, D. T. Jones, D. Schrimpf, M. Sill, S. Schmid, V. Hovestadt, D. E. Reuss, C. Koelsche *et al.*, "Practical implementation of dna methylation and copy-number-based cns tumor diagnostics: the heidelberg experience," *Acta neuropathologica*, vol. 136, pp. 181–210, 2018.

[15] J. Hwang, Y. Lee, S.-K. Yoo, and J.-I. Kim, "Image-based deep learning model using dna methylation data predicts the origin of cancer of unknown primary," *Neoplasia*, vol. 55, p. 101021, 2024.

[16] H. Zheng, A. Momeni, P.-L. Cedoz, H. Vogel, and O. Gevaert, "Whole slide images reflect dna methylation patterns of human tumors," *NPJ genomic medicine*, vol. 5, no. 1, p. 11, 2020.

[17] D. Sturm, D. Capper, F. Andreiuolo, M. Gessi, C. Kölsche, A. Reinhardt, P. Sievers, A. K. Wefers, A. Ebrahimi, A. K. Suwala *et al.*, "Multiomic neuropathology improves diagnostic accuracy in pediatric neuro-oncology," *Nature medicine*, vol. 29, no. 4, pp. 917–926, 2023.

[18] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. Velázquez Vega, D. J. Brat, and L. A. Cooper, "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. E2970–E2979, 2018.

[19] Z. Zhang, W. Yin, S. Wang, X. Zheng, and S. Dong, "Mbfusion: Multimodal balanced fusion and multi-task learning for cancer diagnosis and prognosis," *Computers in Biology and Medicine*, vol. 181, p. 109042, 2024.

[20] Y. Zhang, Y. Xu, J. Chen, F. Xie, and H. Chen, "Prototypical information bottlenecking and disentangling for multimodal cancer survival prediction," *ICLR*, 2024.

[21] O. Ogundipe, Z. Kurt, and W. L. Woo, "Deep neural networks integrating genomics and histopathological images for predicting stages and survival time-to-event in colon cancer," *Plos one*, vol. 19, no. 9, p. e0305268, 2024.

[22] A. H. Song, R. J. Chen, G. Jaume, A. J. Vaidya, A. S. Baras, and F. Mahmood, "Multimodal prototyping for cancer survival prediction," *ICML*, 2024.

[23] Y. Zhang, H. Li, Y. Sun, S. Zheng, C. Zhu, and L. Yang, "Attention-challenging multiple instance learning for whole slide image classification," *ECCV*, 2024.

[24] V. Ramanathan, P. Pati, M. McNeil, and A. L. Martel, "Ensemble of prior-guided expert graph models for survival prediction in digital pathology," in *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, 2024, pp. 262–272.

[25] H. Liu, Y. Shi, Y. Xu, A. Li, and M. Wang, "Agnostic-specific modality learning for cancer survival prediction from multiple data," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[26] Y. Xu and H. Chen, "Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction," in *Proceedings of the IEEE/CVF ICCV*, 2023, pp. 21 241–21 251.

[27] A. H. Song, R. J. Chen, T. Ding, D. F. Williamson, G. Jaume, and F. Mahmood, "Morphological prototyping for unsupervised slide representation learning in computational pathology," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 11 566–11 578.

[28] O. Alwazzan, A. Khan, I. Patras, and G. Slabaugh, "Moab: Multi-modal outer arithmetic block for fusion of histopathological images and genetic data for brain tumor grading," in *International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023.

[29] J. A. Nicoll, T. Bloom, A. Clarke, D. Boche, and D. Hilton, "Brain uk: Accessing nhs tissue archives for neuroscience research," *Neuropathology and Applied Neurobiology*, vol. 48, no. 2, p. e12766, 2022.

[30] P. Bankhead, M. B. Loughrey, J. A. Fernández, Y. Dombrowski, D. G. McArt, P. D. Dunne, S. McQuaid, R. T. Gray, L. J. Murray, H. G. Coleman *et al.*, "Qupath: Open source software for digital pathology image analysis," *Scientific reports*, vol. 7, no. 1, pp. 1–7, 2017.

[31] R. Pidsley, C. C. Y Wong, M. Volta, K. Lunnon, J. Mill, and L. C. Schalkwyk, "A data-driven approach to preprocessing illumina 450k methylation array data," *BMC genomics*, vol. 14, pp. 1–10, 2013.

[32] J. I. Orozco, T. A. Knijnenburg, A. O. Manughian-Peter, M. P. Salomon, G. Barkhoudarian, J. R. Jalas, J. S. Wilmott, P. Hothi, X. Wang, Y. Takasumi *et al.*, "Epigenetic profiling for the molecular classification of metastatic brain tumors," *Nature communications*, vol. 9, no. 1, p. 4627, 2018.

[33] R. Gomes, N. Paul, N. He, A. F. Huber, and R. J. Jansen, "Application of feature selection and deep learning for cancer prediction using dna methylation markers," *Genes*, vol. 13, no. 9, p. 1557, 2022.

[34] Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, X. Ji *et al.*, "Transmil: Transformer based correlated multiple instance learning for whole slide image classification," *Advances in Neural Information Processing systems*, vol. 34, pp. 2136–2147, 2021.

[35] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Advances in NeurIPS*, vol. 30, 2017.

[36] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *ICML*. PMLR, 2018, pp. 2127–2136.

[37] O. Alwazzan, I. Patras, and G. Slabaugh, "Foaa: Flattened outer arithmetic attention for multimodal tumor classification," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 2024, pp. 1–5.

[38] G. Jaume, L. Oldenburg, A. Vaidya, R. J. Chen, D. F. Williamson, T. Peeters, A. H. Song, and F. Mahmood, "Transcriptomics-guided slide representation learning in computational pathology," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 9632–9644.

[39] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 976–11 986.

[40] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, p. 6086, 2024.

[41] Ensembl, "Gene summary: Ensg00000060138 (ybx3) – homo sapiens," 2025, accessed June 2025. [Online]. Available: https://useast.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000060138

[42] Ensembl Project Team, "Location view: Chromosome 12:10851688–10875922 (grch37)," 2025, accessed June 2025. [Online]. Available: https://grch37.ensembl.org/Homo_sapiens/Location/View?r=12:10851688-10875922

[43] W. Chen, Y. Wang, Y. Abe, L. Cheney, B. Udd, and Y.-P. Li, "Haploinsuffciency for znf9 in znf9+/- mice is associated with multiorgan abnormalities resembling myotonic dystrophy," *Journal of molecular biology*, vol. 368, no. 1, pp. 8–17, 2007.

[44] S. M. Gadalla, R. M. Pfeiffer, S. Y. Kristinsson, M. Björkholm, O. Landgren, and M. H. Greene, "Brain tumors in patients with myotonic dystrophy: a population-based study," *European journal of neurology*, vol. 23, no. 3, pp. 542–547, 2016.

[45] C. Xiong, H. Chen, H. Zheng, D. Wei, Y. Zheng, J. J. Sung, and I. King, "Mome: Mixture of multimodal experts for cancer survival prediction," in *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, 2024, pp. 318–328.

[46] L. Long, J. Cui, P. Zeng, Y. Li, Y. Liu, and Y. Wang, "Mugi: Multi-granularity interactions of heterogeneous biomedical data for survival prediction," in *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. Springer, 2024, pp. 490–500.