Knowledge-driven Subspace Fusion and Gradient Coordination for Multi-modal Learning

Yupei Zhang[1]*, Xiaofei Wang[2]*, Fangliangzi Meng[3], Jin Tang[4], Chao Li[2,5,6]✉

[1] Department of Pathology, The University of Hong Kong
[2] Department of Clinical Neurosciences, University of Cambridge, UK
[3] School of Life Sciences and Technology, Tongji University, China
[4] Zhejiang Lab, China
[5] School of Science and Engineering, University of Dundee, UK
[6] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK
cl647@cam.ac.uk

Abstract. Multi-modal learning plays a crucial role in cancer diagnosis and prognosis. Current deep learning based multi-modal approaches are often limited by their abilities to model the complex correlations between genomics and histopathology data, addressing the intrinsic complexity of tumour ecosystem where both tumour and microenvironment contribute to malignancy. We propose a biologically interpretative and robust multi-modal learning framework to efficiently integrate histopathology images and genomics by decomposing the feature subspace of histopathology images and genomics, reflecting distinct tumour and microenvironment features. To enhance cross-modal interactions, we design a knowledge-driven subspace fusion scheme, consisting of a cross-modal deformable attention module and a gene-guided consistency strategy. Additionally, in pursuit of dynamically optimizing the subspace knowledge, we further propose a novel gradient coordination learning strategy. Extensive experiments demonstrate the effectiveness of the proposed method, outperforming state-of-the-art techniques in three downstream tasks of glioma diagnosis, tumour grading, and survival analysis. Our code is available at https://github.com/helenypzhang/Subspace-Multimodal-Learning.

Keywords: Multi-modal learning · Molecular Pathology · Cancer diagnosis and prognosis.

1 Introduction

Multi-modal integration of genomics and histopathology becomes increasingly important for cancer diagnosis, evidenced by the recent shift of cancer taxonomy criteria to integrating molecular markers with histopathology features [2, 6, 22]. However, joint analysis of multi-modal data at the clinic remains challenging [21, 23]. Automatic algorithms to effectively integrate multi-modal data of genomics

---

* Equal contribution.

and histopathology promise to offer rapid diagnosis and aid in precise cancer treatment.

Deep learning-based digital pathology [4, 24] holds promise for rapid cancer diagnosis based on whole slide images (WSIs) derived from tissue sections [26]. Research [12,20] has shown evidence of associations between genomics and WSIs, indicating that morphology features from WSIs may mirror genomic information. However, it remains a challenge to effectively integrate WSIs with genomic information for cancer diagnosis due to (1) the complexity of multi-modal data, i.e., WSIs in gigapixels and genomic profiles of tens of thousands genes and (2) the intrinsic tumour heterogeneity and complexity of cancers.

Previous studies have developed multi-modal approaches to integrate WSIs with genomics for cancer diagnosis. Among them, late-fusion methods integrate modality-specific features at the prediction layer. For instance, Chen *et al.* [4] integrated genomics and WSIs using the Kronecker Product [19] for cancer diagnosis and prognosis. However, late-fusion models ignore the interactions of the multi-modal data at their earlier learning stage, unable to model the modality interaction for robust feature extraction and utilisation.

By contrast, intermediate-fusion techniques show promise to fuse modality-specific features at various levels before prediction. For instance, Chen *et al.* [5] introduced a co-attention method that maps correlations between genomics and WSIs for survival analysis, enhancing survival predictions by learning dense co-attention mappings between genomics and bag representations of WSIs. Additionally, Zhou *et al.* [27] developed a multi-modal learning framework that delves into cross-modal correlations through inter-modality translation and alignment, offering complementary insights from different modalities for survival analysis.

Despite encouraging results, these models typically map WSIs to genomics in a singular embedding space, which may not fully reflect the complexity of cancer. In tumorigenesis, both tumour cells and tumor microenvironment[1] (TME) contribute to malignancy [11], providing essential insights with distinct morphological features for histopathology assessment. Further, evidence [25, 28] shows that genetic profiles are associated with tumour and microenvironment characteristics. For example, isocitrate dehydrogenase (IDH) wildtype has marked necrosis and microvascular proliferation observed in WSIs. Therefore, decoding tumour- and TME-related morphology features from WSIs and genes from genomics promises to advance multi-modal integration. However, it remains challenging to effective model the regional interactions between WSIs and genomics.

To tackle the challenge, we propose a novel multi-modal learning framework with a decoupled gene-to-histopathology integration strategy, for accurate and interpretable automatic cancer diagnosis. Specifically, our contribution is three-fold:

- For the first time, we explicitly decompose the genomics data into tumour- and TME-related genes for effective integration with histological features.

---

[1] The tumor microenvironment is a complex ecosystem surrounding tumor cells, mainly composed of immune cells, also with other stromal cells and vessels.
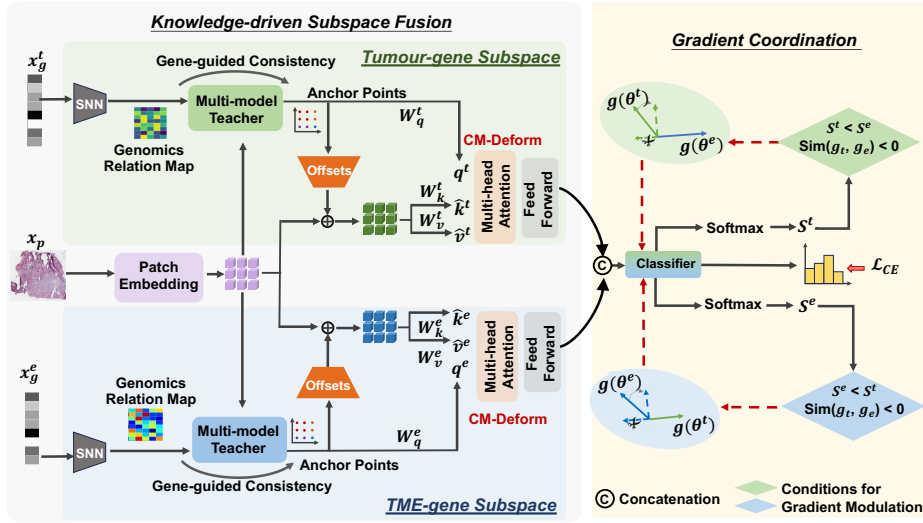
Fig. 1. Overview of the proposed framework. Left: Knowledge-driven subspace fusion scheme. Right: Confidence-guided Gradient Coordination.

– We propose a novel knowledge-driven subspace fusion (KS-Fusion) scheme to effectively enhance the multi-modal interactions. Specifically, in KS-Fusion, we present a cross-modal deformable attention (CM-Deform) module with a gene-guided consistency strategy to capture specific morphological features of the corresponding tumour- and TME-related genes.
– We propose a confidence-guided gradient coordination (CG-Coord) scheme to regulate the multi-modal learning process, which promotes the optimal global performance through dynamic optimization.

Our extensive experiments on the three downstream tasks, i.e., cancer diagnosis, grading, and patient survival prediction, in two public datasets demonstrate that our method outperforms other state-of-the-art (SOTA) methods.

2 Methodology

2.1 Framework

Mounting research [3,8] has revealed the crucial importance of characterizing tumor and TME features for a deeper cancer understanding. We built our method on this biological prior knowledge, learning clinically relevant multimodal features in each subspace. Fig. 1 illustrates the proposed multi-modal learning framework. Overall, to decompose the modal subspace features from genomics and WSIs and model their interactions, we propose a KS-Fusion scheme (left) and a CG-Coord scheme (right). Specifically, as shown in the left part of Fig. 1, categorized genomics features and histopathology features are first fused through

a linear layer to generate a multi-modal teacher, which will be utilized as a query in the deformable attention module. To regulate the deformed offsets, we introduce a batch consistency to regulate the deformed offsets. Then, as illustrated in the right part of Fig. 1, the concatenated multi-modal feature is fed to the classifier for decision prediction, where the proposed CG-Coord scheme can alleviate the conflicts introduced by subspace gradients in training the classifier.

2.2   Knowledge-driven Subspace Fusion

Tumor and TME provide crucial information for cancer analysis while presenting significant variances in genes and WSIs. Existing multi-modal fusion methods lack comprehensive modelling of feature subspaces for genomics profiles and WSIs morphological features. To enhance the biological guidance from niche-specific genomic profiles, we propose a KS-Fusion scheme to capture subspace informative features from both tumour- and TME-related gene profiles. Specifically, in the KS-Fusion scheme, the CM-Deform module along with the gene-guided consistency strategy is designated for efficiently extracting histological features of the corresponding tumour- and TME-related genes simultaneously.

Cross-modal Deformable Attention Module. Overall, we build a two-stream neural network to model the cross-modal subspace interactions within 1) tumour-related genomics x_g^t and the WSIs features x_p; 2) TME-related genomics x_g^e and the WSIs features x_p, where t signifies the tumour niche, e denotes the TME niche, g stands for genomics and p represents WSIs. Specifically, the genomic profile encoder is a spiking neural network (SNN) [10] $\mathcal{G}$. The genomics data x_g is divided into two groups, with x_g^t representing the tumour-related genes and x_g^e representing the TME-related genes.

To realize the cross-modal deformable attention, we first apply a linear layer to obtain the multi-modal teacher features x_{pg}^t and x_{pg}^e, which are then used to guide the offsets generation through an offsets generation network ψ, with two convolution layers and a scaler. Original reference points are a uniform grid of points $p^u \in \mathbb{R}^{H_G \times W_G \times 2}$, given the input feature map $x_{pg}^u \in \mathbb{R}^{H \times W \times C}$, where u represents the subspace and can be uniformly denoted as $u \in \{t, e\}$. For each stream, the multi-head cross-attention can be denoted as:

$$q^u = x_{pg}^u W_q^u, \ \hat{k}^u = \hat{x}_p^u W_k^u, \ \hat{v}^u = \hat{x}_p^u W_v^u, \tag{1}$$

where $q^u, \hat{k}^u, \hat{v}^u$ represent the query, deformed key, and deformed value, respectively the W_q^u, W_k^u, W_v^u are projection networks. The $\hat{x}_p^u = F(x_p^u; \text{norm}(p^u + \Delta p^u))$, $\Delta p^u = \psi(x_{pg}^u)$, and F represents a bilinear interpolation function. For each stream, the output of an attention head is formulated as:

$$z^{m;u} = \text{softmax}(q^{(m);u}\hat{k}^{(m);u\top}/\sqrt{d})\hat{v}^{(m);u}, \tag{2}$$

where m represents the index of the attention head, with the range of 1 to M, and the final output is calculated as:

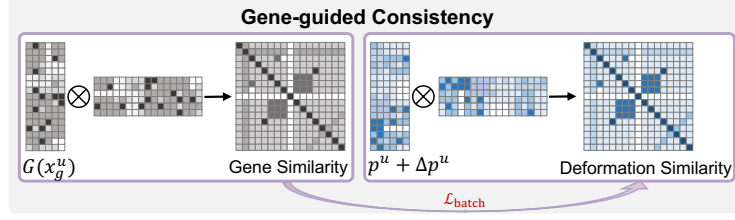$$z^u = \text{concat}(z^{1;u}, ..., z^{M;u})W_o^u, \tag{3}$$

Fig. 2. Illustration of gene-guided consistency strategy, which will regulate the deformation by the sample-wise similarity constraint.

where W_o^u is a projection network. In this way, the informative morphological features are dominated by the multi-modal features, enhancing the correlations and interactions between subspace genomic and histopathology embeddings.

Gene-guided consistency strategy. Furthermore, to ensure the intrinsic consistency between genomics and the deformed WSIs features, we introduce a batch consistency, denoted as gene-guided consistency strategy (Ge-Con), to further regulate offset adjustments in each subspace, which can be formulated as:

$$\mathcal{L}_{\text{batch}}^u = \frac{1}{B}||\mathcal{S}_b(\mathcal{G}(x_g^u)) - \mathcal{S}_b(p^u + \Delta p^u)||_2, \tag{4}$$

where $\mathcal{S}_b(x_1)$ denotes the Gram matrix of feature x_1, representing the correlations among individuals, and B represents the batch size. As shown in Fig. 2, the sample-wise similarity further enhances the guidance of genomics. This gene-knowledge penetration towards WSIs features further enhances the subspace informative feature interaction and interaction, which ultimately facilitates multi-modal subspace fusion.

2.3 Confidence-guided Gradient Coordination

In training the classifier with multi-modal features, it is challenging to obtain a global optimum performance because the subspace gradients of tumour-gene and TME-gene may conflict when applying joint optimization. Therefore, to boost the downstream task by combining these two types of domain knowledge, we design the CG-Coord scheme to obtain the global training optimum via dynamic gradient regulation.

Specifically, assuming that $g(\theta^t)$ and $g(\theta^e)$ are in conflict when $\text{cosine}(g(\theta^t), g(\theta^e))$ is smaller than zero, we adjust the gradient with less prediction confidence score when they conflict with each other. The adjustment process is formulated as:

$$\begin{cases} \tilde{g}(\theta^t) = \Gamma(g(\theta^t), g(\theta^e)), & \sum s^t < \sum s^e, \\ \tilde{g}(\theta^e) = \Gamma(g(\theta^e), g(\theta^t)), & \sum s^e < \sum s^t, \end{cases} \tag{5}$$

where $s^t = \text{softmax}(\mathcal{D}(z^t))[k]$, $s^e = \text{softmax}(\mathcal{D}(z^e))[k]$, and $\mathcal{D}$ is the decoder for the diagnosis and grading tasks. The confidence score for survival is represented

by corresponding C-Index values. In equation 5, $\Gamma(\boldsymbol{x}_1, \boldsymbol{x}_2)$ denotes prjection of the vector $\boldsymbol{x}_1$ to the perpendicular direction of vector $\boldsymbol{x}_2$. Besides, $\sum s^t$ and $\sum s^e$ represent the sum of the corresponding prediction scores on the $k-th$ label from tumour and TME branches in a mini-batch. In this way, gradients of different subspace features will be modulated to avoid conflicts dynamically, thereby achieving harmonious training.

2.4   Training Objectives of Downstream Tasks

In our framework, the training loss functions are tailored for varied downstream tasks, including cancer diagnosis, tumour grading, and prognosis prediction. Specifically, for the cancer diagnosis and grading tasks, we apply the cross-entropy loss as the task-specific objective, and the total training objectives can be formulated as:

$$\mathcal{L}_{\text{diag}} = L_{\text{CE}}(\mathcal{D}(z^t, z^e;\ \theta_{diag}),\ Y_{\text{diag}}) + \alpha\mathcal{L}_{batch}^t + (1-\alpha)\mathcal{L}_{batch}^e, \qquad (6)$$

$$\mathcal{L}_{\text{grad}} = L_{\text{CE}}(\mathcal{D}(z^t, z^e;\ \theta_{grad}),\ Y_{\text{grad}}) + \alpha\mathcal{L}_{batch}^t + (1-\alpha)\mathcal{L}_{batch}^e, \qquad (7)$$

where $\mathcal{L}_{\text{CE}}$ represents the cross-entropy loss, $\mathcal{D}$ represents a classifier, θ_{diag} and θ_{grad} represent the classifier parameters for diagnosis and grading tasks, respectively. Y_{diag} and Y_{grad} represent the diagnosis and grading labels, respectively. The α is a hyper-parameter for balancing subspace gene knowledge penetration. Notably, the hyper-parameter sensitivity analysis study can be found in Section 3.2.

Besides, for the prognosis prediction task, we adopt the NLL (negative log-likelihood) survival loss [27], denoted as $\mathcal{L}_{\text{NLL}}$, as the task-specific objective for the survival outcome prediction, and the total training objective is:

$$\mathcal{L}_{\text{surv}} = \mathcal{L}_{\text{NLL}} + \alpha\mathcal{L}_{batch}^t + (1-\alpha)\mathcal{L}_{batch}^e, \qquad (8)$$

In this way, our method can benefit downstream tasks by promoting subspace cross-modal fusion and harmonious training.

3   Experiments & Results

3.1   Datasets & Implementation Details

Datasets. We conduct experiments on two public datasets, i.e., TCGA GBM-LGG [17] dataset and IvyGAP [14, 15] dataset. The two datasets, focusing on gliomas, are merged as a meta dataset for better performance, including 2,387 tissue samples (668 cases) with paired WSIs and genomics profiles. We randomly split it into training (534 cases), testing (68 cases), and validation (66 cases) sets. WSIs are crop into patches sized at 224px $\times$ 224px of 0.5μm px^{-1}. Following [22], for each WSI, we extract 2500 patches with the biological repeat strategy. Stain normalization [18] is applied to harmonize the data. For genomics process, according to previous studies [1], we first sort the shared gene signatures in the

Table 1. Comparison with SOTA methods on three tasks. p. and g. represent the modality of histopathology and genomics, respectively. Best and second results are highlighted with **bold** and underline.

Methods	p.	g.	Diagnosis, %					Grading, %					Survival, %
			AUC	Acc.	Sen.	Spec.	F1-score	AUC	Acc.	Sen.	Spec.	F1-score	C-Index
AttMIL [7]	✓		80.25	53.81	38.77	81.99	33.30	79.70	61.02	61.36	80.77	60.30	55.59
TransMIL [16]	✓		74.90	51.69	44.06	84.11	39.24	83.14	68.64	67.69	84.21	66.83	67.71
SNN [10]		✓	90.87	76.27	64.79	92.80	64.94	<u>90.52</u>	**83.05**	<u>81.92</u>	<u>91.18</u>	<u>81.72</u>	<u>77.04</u>
Concat	✓	✓	90.89	74.58	63.01	92.35	63.64	89.32	78.39	76.76	88.70	75.00	75.06
Add	✓	✓	<u>91.67</u>	75.85	63.90	92.67	63.56	90.36	<u>82.63</u>	81.60	91.01	81.39	73.42
Bilinear	✓	✓	90.25	<u>79.24</u>	<u>70.33</u>	<u>93.50</u>	<u>69.87</u>	82.49	70.34	68.27	84.55	62.82	73.71
MCAT [5]	✓	✓	80.70	55.08	39.85	84.68	35.81	87.40	59.75	62.63	80.54	50.93	69.64
CMTA [27]	✓	✓	88.14	68.64	59.01	89.87	53.61	87.54	54.66	57.36	78.07	46.85	65.96
Ours	✓	✓	**95.28**	**81.36**	**72.92**	**94.44**	**70.97**	**91.53**	**83.05**	**82.29**	**91.29**	**82.43**	**79.78**

TCGA GBM-LGG [17] and IvyGAP [14,15] datasets according to the expression variance and then select the top 30% Highly Variable Genes[1] (HVGs), to capture important biological information in genes. The genomic profile is enriched with 420 features, encompassing 59 tumour-related genes and 361 TME-related genes. **Implementation Details.** All experiments were conducted using the PyTorch [13] on two NVIDIA RTX A5000 GPUs, with a batch size of 8. Our method was trained for 20 epochs for diagnosis and grading, and 10 epochs for survival prediction. Network optimization was performed using the Adam optimizer [9]. Experiments are conducted three times with different random seeds and we only report the mean values of the evaluation metrics. The key hyper-parameters can be found in Supplementary Table I. Each hyperparameter was tuned to achieve optimal performance on the validation set.

3.2   Performance Evaluation

Baselines and SOTA Comparison Methods. For each task, we compare our model with eight SOTA methods, with three uni-modal methods: AttMIL (Histopathology only) [7], TransMIL (Histopathology only) [16], SNN (Genomics only) [10]) and five multi-modal fusion algorithms: Concat (AttMIL with SNN), Add (AttMIL with SNN), Bilinear (AttMIL with SNN), MCAT [5], and CMTA [27].

Downstream Task I: Glioma Diagnosis. As shown in Table 1, in gliomas diagnosis task, our framework is superior to all SOTA models, achieving AUC of 95.28%, Acc. of 81.36%, Sen. of 72.92%, Spec. of 94.44%, and F1-score of 70.97%. In terms of AUC and Acc., our framework outperforms others by at least 3.61% and 2.12%, respectively, indicating the superior multi-modal learning ability of our method. Furthermore, ROCs can be found in Fig. S1 (top) of supplementary material.

---

[1] Highly Variable Genes indicate high signal-to-noise ratio information within an organism, allowing for more biological information captured with smaller dimensions.

Table 2. Ablation studies on Diagnosis, Grading, and Survival tasks. The best results are highlighted with **bold**.

Methods	Diagnosis, %					Grading, %					Survival, %
	AUC	Acc.	Sen.	Spec.	F1-score	AUC	Acc.	Sen.	Spec.	F1-score	C-Index
w/o Ge-Con	91.55	79.66	69.43	93.89	67.07	90.49	71.61	71.82	85.77	72.04	72.16
w/o CG-Coord	91.27	77.97	67.47	93.45	67.70	89.08	73.73	73.90	86.82	74.10	76.54
Ours	**95.28**	**81.36**	**72.92**	**94.44**	**70.97**	**91.53**	**83.05**	**82.29**	**91.29**	**82.43**	**79.78**

Downstream Task II: Glioma Grading. Gliomas can be categorized into four severity grades. As shown in the middle panel of Table 1, our framework outperforms all SOTA models, achieving AUC of 91.53%, the Acc. of 83.05%, the Sen. of 82.29%, the Spec. of 91.29%, and the F1-score of 82.43%. In terms of AUC and F1-score, our framework outperforms others by at least 1.01% and 0.71%, respectively, suggesting our excellent ability in the grading tasks.

Downstream Task III: Survival Analysis. Following previous studies [4,5], we segment the overall survival time into four intervals based on the quartiles of event times of uncensored patients to compute the discretized-survival C-index. As shown in Table 1, our framework outperforms state-of-the-art models, achieving the C-Index of 79.78%, which outperforms the second-best method (marked in underline) by 2.74%. The results show that our proposed multi-modal learning framework is also powerful in the prognosis prediction task.

Hyper-parameters Sensitivity Analysis. An additional hyper-parameter sensitivity analysis experiment is conducted to investigate the trade-off between tumour and TME subspaces in the proposed KS-Fusion scheme. Specifically, we conduct experiments on the hyperparameter α in Eq. 6 and Eq. 7. As shown in Fig. S1 (bottom) of supplementary material, our method achieves the best performance when α equals 0.5, indicating that we can efficiently utilise both tumour- and TME-related gene features.

Ablation Study. To quantitatively evaluate the effectiveness of our proposed components, we conduct ablation studies for each component on two downstream tasks, i.e., cancer diagnosis and grading. In particular, we apply two ablative baselines of the proposed framework by disabling the gene-guided consistency strategy (denoted as *w/o* Ge-Con) and Confidence-guided Gradient Coordination (denoted as *w/o* CG-Coord). As illustrated in Table 2, the AUC shows an improvement of 3.73% and 4.01% for the Ge-Con and CG-Coord correspondingly on the glioma diagnosis task. In terms of grading, the AUC increases 1.04% and 2.45% for Ge-Con and CG-Coord, respectively, indicating the effectiveness of our gene-guided consistency strategy and CG-Coord scheme.

4   Conclusion

Multi-modal data plays an increasingly important role in recent cancer diagnosis criteria. In order to effectively model the multi-modal data of histopathology

and genomics, we propose a multi-modal learning framework with KS-fusion scheme, reflecting the intrinsic cancer mechanisms of tumour and TME. In the KS-fusion scheme, we propose the CM-Deform module and gene-guided consistency strategy to enhance the multi-modal interaction among tumour-genes, TME-genes and histopathology images. Besides, we also design a CG-Coord scheme to stabilize the multi-modal learning process, via dynamically adjusting the optimization of subspace features. Extensive experiments conducted on three tasks demonstrate the effectiveness of our method, with a significant performance improvement compared to other SOTA methods, promising to promote precision oncology.

References

1. Bhattacharya, S., Dunn, P., Thomas, C.G., Smith, B., Schaefer, H., Chen, J., Hu, Z., Zalocusky, K.A., Shankar, R.D., Shen-Orr, S.S., et al.: Immport, toward repurposing of open access immunological assay data for translational and clinical research. Scientific data **5**(1), 1–9 (2018)
2. Bombonati, A., Sgroi, D.C.: The molecular pathology of breast cancer progression. The Journal of pathology **223**(2), 308–318 (2011)
3. Burgos-Panadero, R., Lucantoni, F., Gamero-Sandemetrio, E., de la Cruz-Merino, L., Álvaro, T., Noguera, R.: The tumour microenvironment as an integrated framework to understand cancer biology. Cancer Letters **461**, 112–122 (2019)
4. Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F.: Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging **41**(4), 757–770 (2020)
5. Chen, R.J., Lu, M.Y., Weng, W.H., Chen, T.Y., Williamson, D.F., Manz, T., Shady, M., Mahmood, F.: Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4025 (2021)
6. Harris, T.J., McCormick, F.: The molecular pathology of cancer. Nature reviews Clinical oncology **7**(5), 251–265 (2010)
7. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
8. Junttila, M.R., De Sauvage, F.J.: Influence of tumour micro-environment heterogeneity on therapeutic response. Nature **501**(7467), 346–354 (2013)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
10. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. Advances in neural information processing systems **30** (2017)

11. Li, C., Wang, S., Liu, P., Torheim, T., Boonzaier, N.R., van Dijken, B.R., Schönlieb, C.B., Markowetz, F., Price, S.J.: Decoding the interdependence of multiparametric magnetic resonance imaging to reveal patient subgroups correlated with survivals. Neoplasia **21**(5), 442–449 (2019)
12. Lin, T.C., Yeh, Y.M., Fan, W.L., Chang, Y.C., Lin, W.M., Yang, T.Y., Hsiao, M.: Ghrelin upregulates oncogenic aurora a to promote renal cell carcinoma invasion. Cancers **11**(3), 303 (2019)
13. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
14. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.G., Smith, K.A., Lankerovich, M., Bertagnolli, D., Bickley, K., et al.: An anatomic transcriptional atlas of human glioblastoma. Science **360**(6389), 660–663 (2018)
15. Shah, N., Feng, X., Lankerovich, M., Puchalski, R.B., Keogh, B.: Data from ivy gap. The Cancer Imaging Archive **10**, K9 (2016)
16. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
17. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: Review the cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia **2015**(1), 68–77 (2015)
18. Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE transactions on medical imaging **35**(8), 1962–1971 (2016)
19. Van Loan, C.F.: The ubiquitous kronecker product. Journal of computational and applied mathematics **123**(1-2), 85–100 (2000)
20. Wang, Q.M., Lv, L., Tang, Y., Zhang, L., Wang, L.F.: Mmp-1 is overexpressed in triple-negative breast cancer tissues and the knockdown of mmp-1 expression inhibits tumor cell malignant behaviors in vitro. Oncology letters **17**(2), 1732–1740 (2019)
21. Wang, X., Jiang, L., Li, L., Xu, M., Deng, X., Dai, L., Xu, X., Li, T., Guo, Y., Wang, Z., et al.: Joint learning of 3d lesion segmentation and classification for explainable covid-19 diagnosis. IEEE transactions on medical imaging **40**(9), 2463–2476 (2021)
22. Wang, X., Price, S., Li, C.: Multi-task learning of histology and molecular markers for classifying diffuse glioma. arXiv preprint arXiv:2303.14845 (2023)
23. Wei, Y., Chen, X., Zhu, L., Zhang, L., Schönlieb, C.B., Price, S., Li, C.: Multimodal learning for predicting the genotype of glioma. IEEE Transactions on Medical Imaging (2023)
24. Xing, X., Chen, Z., Zhu, M., Hou, Y., Gao, Z., Yuan, Y.: Discrepancy and gradient-guided multi-modal knowledge distillation for pathological glioma grading. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 636–646. Springer (2022)
25. Zeng, Y., Zeng, Y., Yin, H., Chen, F., Wang, Q., Yu, X., Zhou, Y.: Exploration of the immune cell infiltration-related gene signature in the prognosis of melanoma. Aging (albany NY) **13**(3), 3459 (2021)
26. Zhang, L., Wei, Y., Fu, Y., Price, S., Schönlieb, C.B., Li, C.: Mutual contrastive low-rank learning to disentangle whole slide image representations for glioma grading. arXiv preprint arXiv:2203.04013 (2022)

27. Zhou, F., Chen, H.: Cross-modal translation and alignment for survival analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21485–21494 (2023)
28. Zhou, M., Zhang, Z., Bao, S., Hou, P., Yan, C., Su, J., Sun, J.: Computational recognition of lncrna signature of tumor-infiltrating b lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer. Briefings in Bioinformatics **22**(3), bbaa047 (2021)