Multimodal Cross-Task Interaction for Survival Analysis in Whole Slide Pathological Images

Songhan Jiang[1], Zhengyu Gan[1], Linghan Cai[1], Yifeng Wang[2], and Yongbing Zhang[1](✉)

[1] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
[2] School of Science, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
ybzhang08@hit.edu.cn

Abstract. Survival prediction, utilizing pathological images and genomic profiles, is increasingly important in cancer analysis and prognosis. Despite significant progress, precise survival analysis still faces two main challenges: (1) The massive pixels contained in whole slide images (WSIs) complicate the process of pathological images, making it difficult to generate an effective representation of the tumor microenvironment (TME). (2) Existing multimodal methods often rely on alignment strategies to integrate complementary information, which may lead to information loss due to the inherent heterogeneity between pathology and genes. In this paper, we propose a Multimodal Cross-Task Interaction (MCTI) framework to explore the intrinsic correlations between subtype classification and survival analysis tasks. Specifically, to capture TME-related features in WSIs, we leverage the subtype classification task to mine tumor regions. Simultaneously, multi-head attention mechanisms are applied in genomic feature extraction, adaptively performing genes grouping to obtain task-related genomic embedding. With the joint representation of pathological images and genomic data, we further introduce a Transport-Guided Attention (TGA) module that uses optimal transport theory to model the correlation between subtype classification and survival analysis tasks, effectively transferring potential information. Extensive experiments demonstrate the superiority of our approaches, with MCTI outperforming state-of-the-art frameworks on three public benchmarks. https://github.com/jsh0792/MCTI.

Keywords: Survival analysis · Multiple instance learning · Multi-task learning · Transport-Guided attention

1 Introduction

Survival analysis is a crucial topic in clinical prognosis research, aiming to predict the time elapsed from a known origin to events of interest, such as death

S. Jiang, Z. Gan, and L. Cai—Contributed equally to this work.

and disease recurrence [1,10,18,20]. Accurate survival prediction is significant for clinical management and decision-making, benefiting patients by enabling healthcare professionals to tailor personalized treatment plans. Traditionally, survival analysis is time-consuming and labor-intensive to make a predictive diagnosis by pathologists [8,12]. With the development of deep learning, survival analysis based on whole slide images (WSIs) and genomic profiles [4,6,15,27] has shown massive potential for facilitating disease progression and treatment.

Given the gigapixel resolution of WSIs (e.g., $40,000 \times 40,000$ pixels), the pathological image analysis is often formulated as a weakly supervised task using multiple instance learning (MIL). A WSI is developed as a bag containing multiple instances (patches) within the MIL framework. Existing MIL approaches [11,14,16] generally employ a two-stage architecture: initially using a deep neural network to extract instance features and subsequently aggregating them through a pooling function to obtain a bag representation utilized in downstream tasks. In the task of survival analysis, to better extract pathological features, WSISA [28] adopts k-means clustering in the MIL framework to capture representative patches. DeepAttnMISL [26] further introduces an attention pooling [11] that adaptively aggregates the selected instances for improving bag representation. However, these methods could not effectively extract the tumor microenvironment (TME) [24,21,2] contained in WSIs, as they ignore areas that may have critical information, like tumor cells and lymphocyte infiltration [9], which are highly relevant to survival analysis. On the other hand, the subtype classification task requires the network to capture the tumor regions involved in WSI. Consequently, introducing a subtype classification task could potentially enrich the TME-related features, promoting survival analysis performance.

Genes expression corresponds to some morphological characteristics of pathological TME [17,19], which is crucial for improving survival analysis. Most related works focus on solving the alignment problem among different modalities [3,4,5,15]. Pathomic Fusion [3] develops a tensor fusion module to fuse pathological and genomic features. MCAT [4] uses a multimodal co-attention module to identify instances from pathological images using genomic features as queries. MoCAT [5] builds interactions between pathology and genomics through optimal transport, aligning genomic representations to pathological features. However, the alignment process inevitably loses modality-specific information and semantic differences between pathological images and genomic profiles. Unlike these methods, this paper explores the task correlation between subtype classification and survival analysis, optimizing the joint representation of genes and pathology through task interaction, aiming to enhance the effectiveness of survival analysis.

This paper proposes a Multimodal Cross-Task Interaction (MCTI) framework that integrates the subtype classification task for improving survival analysis. Specifically, MCTI leverages the tumor localization ability of attention-based multiple instance learning framework in the context of subtype classification task, effectively enriching TME-related features. Meanwhile, multi-head attention mechanisms are used to process genes for adaptively grouping and embedding. With the joint representation of pathological image and genes, we

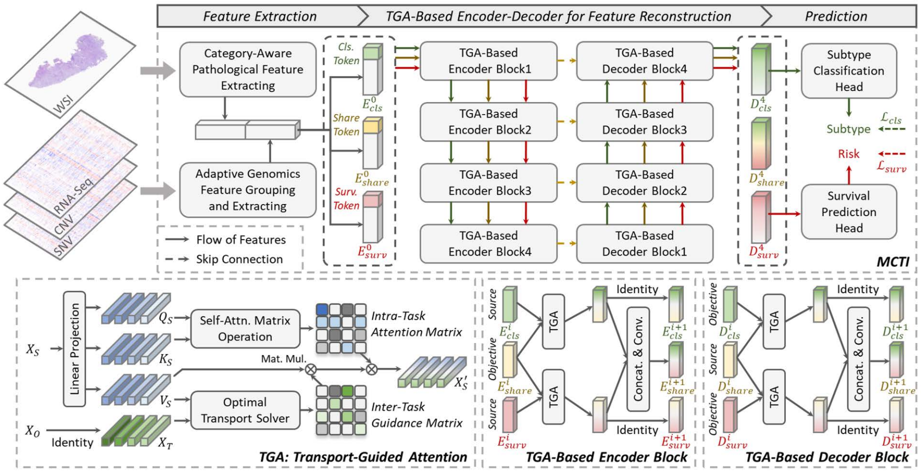

Fig. 1: Illustration of Multimodal Cross-Task Interaction (MCTI) framework. "Mat. Mul." denotes matrix multiplication, "Attn." is attention, "Conv." is a convolutional layer, and "Concat." refers to concatenation operation.

perform a Transport-Guided Attention (TGA) based encoder-decoder for feature reconstruction, where TGA considers the correlation between both tasks, effectively achieving the information interaction for benefiting survival analysis. In short, the contributions of this paper are threefold. (1) We propose a Multimodal Cross-Task Interaction (MCTI) framework that leverages pathological images and genomic data for survival analysis. The framework ingeniously utilizes the subtype classification task to mine valuable disease-positive instances for the survival analysis task, significantly enhancing the model's perception of the tumor microenvironment. (2) We introduce a novel Transport-Guided Attention (TGA) module based on optimal transport theory, highlighting the correlation between subtype classification and survival analysis tasks, effectively performing information interaction for both tasks and optimizing the unified feature representation of pathology and genes. (3) Extensive experiments validate the effectiveness of our approaches, and MCTI outperforms the state-of-the-art frameworks on three public benchmarks. Codes will be publicly available.

2 Methodology

2.1 Preliminary

Survival Analysis. Let $D = \{D_1, D_2, ..., D_N\}$ represent the clinical data of N patients. Each patient data can be represented by a $D_i = (P_i, G_i, c_i, t_i, Y_i)$, where P_i is the set of WSIs, G_i is the set of genomic profiles, $c_i \in \{0, 1\}$ is the right uncensorship status, t_i is overall survival time, and Y_i denotes the subtype of cancer. We aim to construct a survival prediction model to estimate $f_{hazard}(T =$

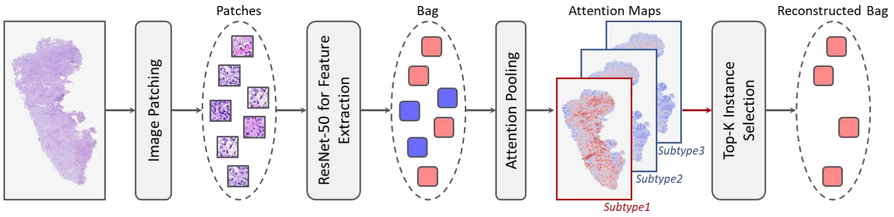

Fig. 2: Overall structure diagram of WSI bag construction in MCTI.

$t|T \geq t, X)$, where T is a random variable and t represents the time point of the occurrence of the death event. And the cumulative value of the risk function is output as the risk score: $f_{surv}(T \geq t, X) = \prod_{u=1}^{t}(1 - f_{hazard}(T = t|T \geq t, X))$.

Subtype Classification. The subtype classification task is often formulated as a MIL problem, where the patches extracted from WSI are considered as instances of the bag, and only the slide-level label Y can be obtained. For a bag B containing n patches, it can be formulated as $B = \{(x_1, y_1), ..., (x_n, y_n)\}$, where x_i represents the i^{th} patch, and y_i represents the corresponding label for the patch. Mainstream MIL frameworks [11,14] generally adopt a suitable transformation f and a permutation-invariant transformation g to obtain the predicted label $\hat{Y}$ of B, given by $\hat{Y} = g\left(f\left(x_1\right), ...f\left(x_n\right)\right)$.

2.2 Overall Framework

The overall framework of our proposed method, Multimodal Cross-Task Interaction (MCTI), is shown in Fig. 1. Firstly, we extract pathological features using a subtype classification task and obtain genomic features by adaptive grouping. Next, the encoder-decoder structure based on Transport-Guided Attention reconstructs the multimodal embedding. Finally, we use the embedding to predict the survival hazard score. The details are described in the following parts.

2.3 Pathological and Genomic Feature Extraction

WSI Bag Construction. We use the DSMIL [14] framework as the backbone, where the instance branch assigns an attention score for each instance, reflecting the instance's importance to the subtype classification. We select the top-k instances based on the scores to form a new bag B_n, as shown in Fig. 2. To supervise the selection of patches related to the tumor microenvironment, we use the cross-entropy $H(y, \hat{y}) = -\sum_i y_i \cdot log(\hat{y}_i)$ as loss function for bag classification and instance classification: $\mathcal{L}_{DSMIL} = H(Y_i, \hat{Y}_{bag}) + H(y_i, \hat{y}_{instance})$, where $\hat{y}_{instance}$ is the prediction type of the critical instance, y_i is equivalent to Y_i. Then, we can locate relevant regions of the tumor microenvironment of WSI by the subtype classification task.

Gene Bag Construction. Inspired by recent work [4,22,25], we aim for our model to adaptively group and extract gene representations. Specifically, we employ Multi-head Self-attention (MSA) [23] functions in parallel to generate feature representation, which allow the model to consider information from different representation subspaces simultaneously. Then, they are concatenated and once again projected. We obtain the new presentation of genes $B \in \mathbb{R}^{1\times d}$. Here, we replicate the gene embedding k times, resulting in a representation $B \in \mathbb{R}^{k\times d}$ with the same dimensions as the pathological features, where k critical patches have been selected.

2.4 Multi-Task Encoder-Decoder for Feature Reconstruction

In the above steps, we derive the representation of the WSI $B\left(P\right) = \left\{p_1, ..., p_k\right\}$ abbreviated as P and genes $B\left(G\right) = \left\{g_1, ..., g_k\right\}$ abbreviated as G. Then, we concatenate B and G to obtain a multimodal representation $X \in \mathbb{R}^{2k\times d}$. Afterward, we employ a Transport-Guided Attention (TGA) based encoder-decoder for feature reconstruction. It effectively facilitates information interaction to capture task-complementary information and enhance feature representation related to survival analysis.

Transport-Guide Attention (TGA). TGA utilizes optimal transport (OT) to explore the effective knowledge transfer between tasks, yielding cross-task representations. The representations collect complementary information that may not be available within a single task. TGA also employs a self-attention structure to preserve intra-task information within the source distribution, ensuring minimal information loss when it flows from the objective to the source. The computation of the TGA module can be formulated as follows:

$$TGAB\left(S, O\right) = \left(Q^T S\right)\left(K^T S\right) F_n^T \left(V^T S\right), \tag{1}$$

where S is the source data, O is the objective data. F_n is calculated from $W\left(S, O\right) = min \left\langle F_n, C_n \right\rangle_{\mathcal{F}}$, where $< \cdot >_{\mathcal{F}}$ refers to the Frobenius dot product. The best matching flow F_n based on local pairwise similarity and cost matrix C_n measures the distance of embeddings using optimal transport theory [5,7].

TGA-based Encoder-Decoder Blocks. We hope the encoder can learn the shared inter-task features and the decoder can learn the specific intra-task features layer by layer rather than transferring knowledge between two tasks simultaneously. Thus, we firstly define three learnable feature tokens $x_{cls} \in \mathbb{R}^{k\times d}$, $x_{surv} \in \mathbb{R}^{k\times d}$, and $x_{share} \in \mathbb{R}^{k\times d}$ to generate the inputs of encoder:

$$E_{share}^0 = concat(X, x_{share}), E_{cls}^0 = concat(X, x_{cls}), E_{surv}^0 = concat(X, x_{surv}), \tag{2}$$

where E_{share}^0 is inter-task embedding, E_{cls}^0 is the embedding of subtype classification task and E_{surv}^0 is the embedding of survival analysis.

The computation of the encoder block is formulated as follows:

$$E_{cls}^{i+1} = TGA\left(E_{cls}^{i}, E_{share}^{i}\right), E_{surv}^{i+1} = TGA\left(E_{surv}^{i}, E_{share}^{i}\right), \tag{3}$$

$$E_{share}^{i+1} = conv\left(concat\left(E_{cls}^{i+1}, E_{surv}^{i+1}\right)\right). \tag{4}$$

The computation of the decoder block is formulated as follows:

$$D_{cls}^{i+1} = TGA((D_{share}^{i}+E_{share}^{n-i+1}), D_{cls}^{i}), D_{surv}^{i+1} = TGA((D_{share}^{i}+E_{share}^{n-i+1}), D_{surv}^{i}), \tag{5}$$

$$D_{share}^{i+1} = conv\left(concat\left(D_{cls}^{i+1}, D_{surv}^{i+1}\right)\right). \tag{6}$$

During the decoding process, a skip connection is adopted, and the output of the intermediate step of the encoder is connected to the decoder as part of the input. Our model stacks four layers in the encoder-decoder to obtain more task complementary information.

2.5 Loss Function

MCTI uses negative log-likelihood survival loss [9] as the survival analysis loss $\mathcal{L}_{surv}$ and the cross-entropy loss as the loss function of the subtype classification $\mathcal{L}_{cls}$. Combining with $\mathcal{L}_{DSMIL}$ (refer to Section 2.3), the total loss $\mathcal{L}_{total}$ can be formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{DSMIL} + \alpha\mathcal{L}_{surv}, \tag{7}$$

where α is a hyper-parameter for balancing the influence of the loss function, the value of α is 1.

3 Experiment

3.1 Datasets & Experimental Settings & Evaluation Metrics

We conduct extensive experiments on four public datasets from The Cancer Genome Atlas (TCGA). Specifically, we used Breast Invasive Carcinoma (BRCA), Esophageal Carcinoma (ESCA), Kidney Renal Papillary Cell Carcinoma (KIRP), and Non-small Cell Lung Cancer (NSCLC). We conduct 4-fold cross-validation for evaluation and then randomly split the data as the ratio of training: validation: testing = 60: 15: 25. Our MCTI is implemented in PyTorch 1.12.1 using an NVIDIA RTX 3090 GPU. During training, we use Adam optimization with a learning rate of 0.00005. The number of critical instances selected $k = 256$. Following CLAM [16], we segment tissue regions for each WSI and crop 256 × 256 patches over 20× magnification, then use ImageNet pretrained ResNet-50 to extract the embedding ($d = 1024$) for each patch. We use the concordance index (C-Index) to evaluate the performance of survival analysis models. Kaplan-Meier analysis is utilized to measure the statistical significance between low risk group and high risk group.

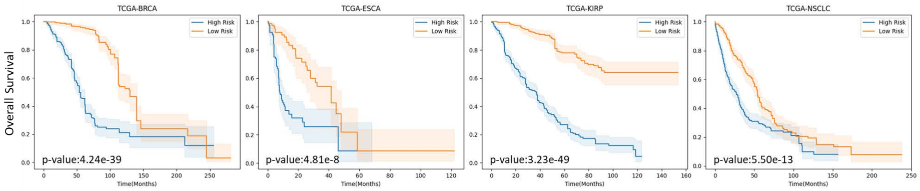

Fig. 3: Kaplan-Meier Analysis on four cancer datasets according to predicted risk scores. P-value < 0.05 means significant statistical difference between low-risk (blue) and high-risk (red).

Table 1: C-Index (mean ± std) performance over four cancer datasets. The best results are shown in **bold**, and the second best ones are underlined.

Model	**TCGA-BRCA**	**TCGA-ESCA**	**TCGA-NSCLC**	**TCGA-KICA**
SNN	0.584 ± 0.035	0.540 ± 0.056	0.542 ± 0.026	0.651 ± 0.011
AvgPool	0.566 ± 0.083	0.567 ± 0.024	0.568 ± 0.028	0.658 ± 0.049
AttnMIL [11]	0.556 ± 0.100	0.583 ± 0.036	0.571 ± 0.030	0.657 ± 0.073
CLAM-SB [16]	0.489 ± 0.083	0.551 ± 0.051	0.552 ± 0.037	0.692 ± 0.058
CLAM-MB [16]	0.563 ± 0.066	0.591 ± 0.041	0.581 ± 0.025	0.695 ± 0.040
DSMIL [14]	0.543 ± 0.095	$\underline{0.594} \pm 0.007$	0.581 ± 0.027	0.645 ± 0.033
MCAT [4]	0.544 ± 0.034	0.553 ± 0.085	0.638 ± 0.013	0.693 ± 0.010
CMTA [27]	0.578 ± 0.029	0.565 ± 0.033	$\mathbf{0.655} \pm 0.028$	$\underline{0.698} \pm 0.015$
PORPOISE [6]	0.587 ± 0.043	0.527 ± 0.034	0.542 ± 0.013	0.677 ± 0.040
M3IF [15]	$\underline{0.579} \pm 0.047$	0.557 ± 0.031	0.502 ± 0.027	0.664 ± 0.018
Ours	$\mathbf{0.656} \pm 0.018$	$\mathbf{0.621} \pm 0.044$	$\underline{0.639} \pm 0.028$	$\mathbf{0.723} \pm 0.046$

3.2 Comparison with State-of-the-Art Methods

Table 1 shows the quantitative results of unimodal and multimodal methods. For unimodal methods, we adopt SNN [13] as the genomic features extractor for survival analysis. Additionally, AvgPool, ABMIL [11], CLAM [16], and DSMIL [14] use pathological images for survival analysis. Table 1 shows that their performances are worse than multimodal methods.

We also compare our method with multimodal survival analysis frameworks, including Porpoise [6], MCAT [4], M3IF [15], CMAT [27].

Unlike them, exploring modality alignment to transfer potential complementary information, our method introduces the subtype classification task to assist survival analysis. Our model outperforms other state-of-the-art multimodal methods by a large margin in the BRCA, ESCA, and KIRP datasets. Especially on the BRCA dataset, our model surpasses the second best by 7.7%. We also visualize the Kaplan-Meier survival curves in Fig. 3 to demonstrate a statistical distinction in patient stratification performance.

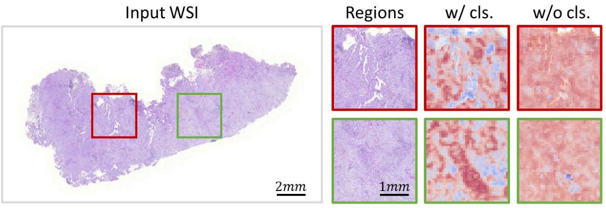

Fig. 4: Visualization of attention map. “cls.” is the subtype classification task.

Table 2: Ablation study assessing C-Index (mean ± std) performance. The best results are shown in **bold**. “recon.” refers to reconstruction.

Settings	**TCGA-BRCA**	**TCGA-ESCA**	**TCGA-KICA**	**TCGA-NSCLC**
MCTI w/o WSI Recon.	0.489 ± 0.083	0.551 ± 0.051	0.692 ± 0.058	0.552 ± 0.037
MCTI w/o Genes Group.	0.563 ± 0.066	0.591 ± 0.041	0.695 ± 0.040	0.581 ± 0.025
MCTI w/o TGA	0.545 ± 0.021	0.592 ± 0.093	0.582 ± 0.148	0.537 ± 0.020
MCTI	$\mathbf{0.656} \pm 0.018$	$\mathbf{0.621} \pm 0.044$	$\mathbf{0.723} \pm 0.046$	$\mathbf{0.639} \pm 0.028$

3.3 Ablation Studies

In this section, we perform ablation experiments to validate the impact of our modules. Quantitative results are presented in Table 2.

Effectiveness of Critical Patch Selection. Table 2 shows that when patches are randomly selected, C-Index drops by 16.7%, 7.0%, 3.1%, and 8.7% in the four datasets. In contrast, choosing critical patches based on subtype classification aims to explore the tumor microenvironment and provide a better WSI bag representation for survival analysis tasks.

Effectiveness of Multi-Head Attention on Genomic Embedding. As shown in Table 2, the BRCA dataset is significantly influenced by multi-head attention. Furthermore, the figure indicates that adaptive genes grouping and feature extraction are beneficial for survival analysis.

Effectiveness of Subtype Classification in TGA. When we remove the loss function of the arbitrary task in TGA, for the sake of fairness of the model, we also replace the branch of the classification task with survival analysis, i.e., $\mathcal{L}_{total} = \mathcal{L}_{surv} + \alpha\mathcal{L}_{surv}$. From Table 2, we can find that the performance decreases by 11.1%, 2.9%, 14.1%, and 10.2% on the four datasets respectively. Fig. 4 shows the attention map generated by DSMIL, from which we can observe that our model excels in capturing tumor-related regions.

4 Conclusion

In this study, we propose a novel Multimodal Cross-Task Interaction (MCTI) framework that leverages subtype classification as an auxiliary task to enhance survival analysis. Based on attention-based multiple instance learning, MCTI performs subtype classification to precisely identify tumor regions within WSIs, enhancing the representation of TME-related features. Furthermore, a Transport-Guided Attention (TGA) module is designed to consider the correlation between tasks and effectively transfer the knowledge from the subtype classification task to survival analysis. Our experiments demonstrate the effectiveness of MCTI, outperforming state-of-the-art frameworks across three public benchmarks. This study provides fresh insight into survival analysis. Future work would focus on explaining the relations between subtype classification and survival analysis and validating MCTI's performance on multi-center datasets.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China (62031023&62331011), in part by the Shenzhen Science and Technology Project (GXWD20220818170353009), and in part by the Fundamental Research Funds for the Central Universities (Grant No. HIT.OCEF.2023050).

Disclosure of Interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Aalen, O., Borgan, O., Gjessing, H.: Survival and event history analysis: a process point of view. Springer Science & Business Media (2008)
2. Brodsky, A.S., Khurana, J., Guo, K.S., Wu, E.Y., Yang, D., Siddique, A.S., Wong, I.Y., Gamsiz Uzun, E.D., Resnick, M.B.: Somatic mutations in collagens are associated with a distinct tumor environment and overall survival in gastric cancer. BMC cancer **22**(1), 139 (2022)
3. Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F.K., 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 (2022). https://doi.org/10.1109/TMI.2020.3021387
4. 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 (ICCV). pp. 4015–4025 (October 2021)
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. Chen, R.J., Lu, M.Y., Williamson, D.F., Chen, T.Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., et al.: Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell **40**(8), 865–878 (2022)

7. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: Scaling algorithms for unbalanced optimal transport problems. Mathematics of Computation **87**(314), 2563–2609 (2018)
8. Collins, F.S., Varmus, H.: A new initiative on precision medicine. New England journal of medicine **372**(9), 793–795 (2015)
9. Dey, R., Zhou, W., Kiiskinen, T., Havulinna, A., Elliott, A., Karjalainen, J., Kurki, M., Qin, A., FinnGen, Lee, S., et al.: Efficient and accurate frailty model approach for genome-wide survival association analysis in large-scale biobanks. Nature communications **13**(1), 5437 (2022)
10. Dey, T., Lipsitz, S.R., Cooper, Z., Trinh, Q.D., Krzywinski, M., Altman, N.: Survival analysis—time-to-event data and censoring (2022)
11. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
12. Jackson, H.W., Fischer, J.R., Zanotelli, V.R., Ali, H.R., Mechera, R., Soysal, S.D., Moch, H., Muenst, S., Varga, Z., Weber, W.P., et al.: The single-cell pathology landscape of breast cancer. Nature **578**(7796), 615–620 (2020)
13. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. Advances in neural information processing systems **30** (2017)
14. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
15. Li, H., Yang, F., Xing, X., Zhao, Y., Zhang, J., Liu, Y., Han, M., Huang, J., Wang, L., Yao, J.: Multi-modal multi-instance learning using weakly correlated histopathological images and tabular clinical information. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 529–539. Springer (2021)
16. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
17. Medema, J.P., Vermeulen, L.: Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. Nature **474**(7351), 318–326 (2011)
18. Nagy, Á., Munkácsy, G., Győrffy, B.: Pancancer survival analysis of cancer hallmark genes. Scientific reports **11**(1), 6047 (2021)
19. Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P., et al.: Multiclass cancer diagnosis using tumor gene expression signatures. Proceedings of the National Academy of Sciences **98**(26), 15149–15154 (2001)
20. Shmatko, A., Ghaffari Laleh, N., Gerstung, M., Kather, J.N.: Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nature cancer **3**(9), 1026–1038 (2022)
21. Tron, L., Belot, A., Fauvernier, M., Remontet, L., Bossard, N., Launay, L., Bryere, J., Monnereau, A., Dejardin, O., Launoy, G., et al.: Socioeconomic environment and disparities in cancer survival for 19 solid tumor sites: An analysis of the french network of cancer registries (francim) data. International journal of cancer **144**(6), 1262–1274 (2019)
22. Välk, K., Vooder, T., Kolde, R., Reintam, M.A., Petzold, C., Vilo, J., Metspalu, A.: Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers. Oncology **79**(3-4), 283–292 (2011)

23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
24. Wang, W., Lu, Z., Wang, M., Liu, Z., Wu, B., Yang, C., Huan, H., Gong, P.: The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma. Frontiers in immunology **13**, 998236 (2022)
25. Yao, J., Wang, S., Zhu, X., Huang, J.: Imaging biomarker discovery for lung cancer survival prediction. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 649–657. Springer (2016)
26. Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Medical Image Analysis **65**, 101789 (2020)
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. Zhu, X., Yao, J., Zhu, F., Huang, J.: Wsisa: Making survival prediction from whole slide histopathological images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6855–6863 (2017). https://doi.org/10.1109/CVPR.2017.725