

# Bipartite Patient-Modality Graph Learning with Event-Conditional Modelling of Censoring for Cancer Survival Prediction

Hailin Yue<sup>1</sup>, Hulin Kuang<sup>1\*</sup>, Jin Liu<sup>1,2</sup>, Junjian Li<sup>1</sup>, Lanlan Wang<sup>1</sup>, Mengshen He<sup>1</sup>, and Jianxin Wang<sup>1,2</sup>

<sup>1</sup>Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha, China

<sup>2</sup>Xinjiang Engineering Research Center of Big Data and Intelligent Software, School of Software, Xinjiang University, Urumqi, China  
hulinkuang@csu.edu.cn

**Abstract.** Accurately predicting the survival of cancer patients is crucial for personalized treatment. However, existing studies focus solely on the relationships between samples with known survival risks, without fully leveraging the value of censored samples. Furthermore, these studies may suffer performance degradation in modality-missing scenarios and even struggle during the inference process. In this study, we propose a bipartite patient-modality graph learning with event-conditional modelling of censoring for cancer survival prediction (CenSurv). Specifically, we first use graph structure to model multimodal data and obtain representation. Then, to alleviate performance degradation in modality-missing scenarios, we design a bipartite graph to simulate the patient-modality relationship in various modality-missing scenarios and leverage a complete-incomplete alignment strategy to explore modality-agnostic features. Finally, we design a plug-and-play event-conditional modelling of censoring (ECMC) that selects reliable censored data using dynamic momentum accumulation confidences, assigns more accurate survival times to these censored data, and incorporates them as uncensored data into training. Comprehensive evaluations on 5 publicly cancer datasets showcase the superiority of CenSurv over the best state-of-the-art by 3.1% in terms of the mean C-index, while also exhibiting excellent robustness under various modality-missing scenarios. In addition, using the plug-and-play ECMC module, the mean C-index of 8 baselines increased by 1.3% across 5 datasets. Code of CenSurv is available at <https://github.com/yuehailin/CenSurv>.

**Keywords:** Survival prediction · Censored data · Plug-and-play.

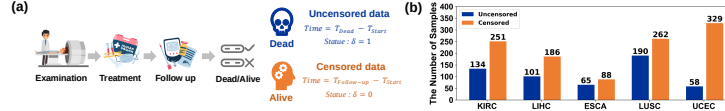
## 1 Introduction

Globally, cancer is a huge public health challenge. According to data from the World Health Organization (WHO), millions of people are diagnosed with cancer

---

\* Corresponding author

each year, leading to millions of deaths [24]. Personalized treatment for patients is recommended in clinical practice [14,11]. To guide personalized treatment for patients, it is necessary to accurately predict the survival of cancer patients.

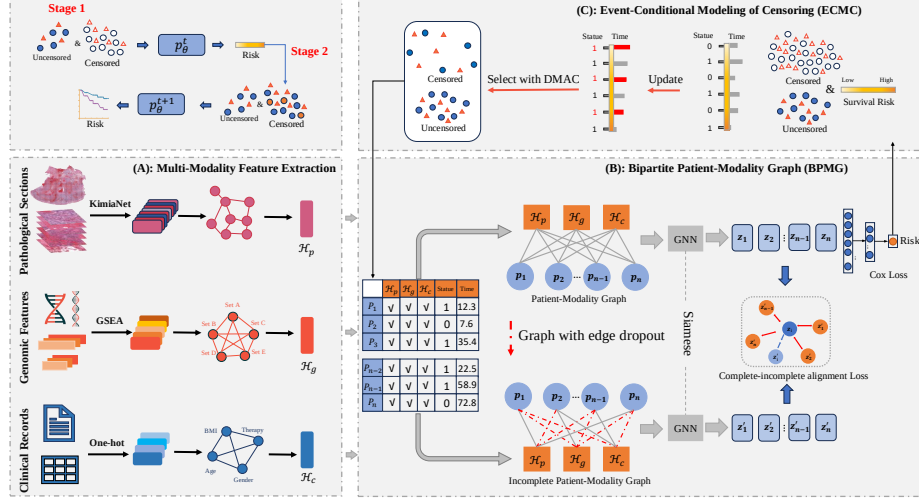


**Fig. 1.** (a) The distinction of censored and uncensored data. For uncensored data, the survival status is dead ( $\delta = 0$ ), and the survival time is accurate. For censored data, the survival status is alive ( $\delta=1$ ), and the survival time calculated with the follow-up time as the event endpoint is inaccurate and smaller than the actual survival time. (b) The number of censored and uncensored data on 5 datasets in our study.

With advances in medical imaging and genomics, many studies have identified potential markers for cancer survival prediction [20,5]. For example, Shao et al. proposed a hierarchical vision transformer for survival prediction using whole slide images [20]. Increasingly, studies combined imaging, genetic, and clinical data for more accurate survival prediction [8,12]. For example, Chen et al. used both pathological images and genetic data for survival prediction [2]. Other methods like MultiSurv [26], SurvMamba [4] and SurvPath [9] also used these multimodal data for survival prediction. However, these methods cannot adapt to the modality-missing scenarios. The existing methods for missing modalities overlook the differences between complete and incomplete modalities [8,15]. Thus, a method that explores differences between complete and incomplete modalities while remaining robust in modality-missing scenarios is essential.

In clinical practice, the number of uncensored samples with accurate survival times is limited, with the majority consisting of censored data lacking precise survival times (the specific details are shown in Fig. 1). Nowadays, most studies on survival prediction only consider the relationship between uncensored samples and other samples with known survival risks, while ignoring the potential value of censored samples with uncertain survival risks [9,18]. Existing studies using censored data do not consider the impact of the true survival time of censored samples on model performance [17]. Therefore, effectively mining the actual survival time of unavoidable censored data, updating the corresponding survival status, and incorporating these data into training as uncensored to increase the proportion of uncensored samples may enhance the performance of models in survival prediction.

Based on these considerations, we propose a novel CenSurv for cancer survival prediction. CenSurv proposes a bipartite graph to simulate the patient-modality relationship and explore modality-agnostic features. In addition, CenSurv introduces event-conditional modelling of censoring (ECMC) to fully use of the censored data. The main contributions of CenSurv can be summarized as follows: 1) We propose a plug-and-play ECMC to uncover the value of censored



**Fig. 2.** Overview of the bipartite patient-modality graph learning with event-conditional modelling of censoring (CenSurv).

data, improving the mean C-index by 1.3% across 8 baselines evaluated on 5 cancer datasets. 2) We propose a bipartite patient-modality graph that fully exploits the flexibility of graph learning and enables the model to not only work with different modality-missing scenarios but also achieve better performance than existing methods for handling modality-missing.

## 2 Methods

The goal of this study is to use pathological sections, genomic features, and clinical records for survival prediction. Survival prediction, also known as time-to-event analysis, uses both survival time and survival status to assess the probability of experiencing a specified event, such as mortality in the clinical setting before a time. An overview of CenSurv is shown in Fig. 2. CenSurv uses the bipartite patient-modality graph with event-conditional modelling of censoring to improve the performance of the model in survival prediction and its robustness under different modality-missing scenarios. CenSurv consists of multimodal feature extraction (Fig. 2 (A)), bipartite patient-modality graph (Fig. 2 (B)), and event-conditional modelling of censoring (Fig. 2 (C)).

### 2.1 Multimodal Feature Extraction

For the pathological section (p), we crop the original images at  $10\times$  magnification into non-overlapping  $512\times 512$  patches. We then use the pre-trained KimiaNet [16] to extract features for each patch. Since survival prediction requires

both local and global tumor features [8], we use a graph to represent both the local and global information of pathological sections. Specifically, each patch is connected to its 8 spatially adjacent patches, forming a graph of the entire pathological section. For the genomic profile (g), we generate 5 genomic embeddings using Gene Set Enrichment Analysis (GSEA) [23]. These embeddings are then connected to form a graph representation of the genomic profile. For clinical records (c), we one-hot encode each clinical feature to obtain a vector representation, which we fully concatenate to create a graph representation of the clinical data. To ensure consistency across modalities, we align all node features to 1024 dimensions via zero padding. Subsequently, we use three independent GraphSAGE [6] models to facilitate information transfer across each modality graph.

To unify the dimensions of different modalities, we introduce hyperedge. Specifically, we first link all nodes belonging to the same modality by a hyperedge. Then, we use the attention pooling operation to extract the high-order representation of each modality. The representation of modality  $m$  with  $N$  nodes can be implemented as:

$$\mathcal{H}_m = \sum_{n=1}^N \text{Softmax}(\text{MLP}_1(\mathbf{V}_m^n)) \odot \mathbf{V}_m^n \quad (1)$$

where  $\mathbf{V}_m^n$  is feature node of  $\tilde{\mathcal{G}}_m$ . MLP is a Multilayer Perceptron that assigns attention scores to  $\mathbf{V}_m$ .  $\mathcal{H}_m$  is the feature representation of modality  $m$ .

## 2.2 Bipartite Patient-Modality Graph

Bipartite patient-modality graph can be denoted as  $\mathcal{G}_P^M = \{\mathbf{V}_P, \mathbf{V}_M, \mathbf{A}\}$ , the types of nodes consists of patient nodes  $\mathcal{V}_P$  and modality nodes  $\mathcal{V}_M$ .  $\mathcal{V}_P = \{u_1, \dots, u_P\}$ ,  $\mathcal{V}_M = \{v_1, \dots, v_M\}$ .  $P$  and  $M$  are the numbers of patients and modalities, respectively. In our study, there are 3 modalities, so  $M = 3$ . The patient-modality relationship is presented as a patient-modality matrix  $\mathbf{A}$ . Specifically, if patient  $p$  contains modality  $m$ , then the edge  $e_{u_p v_m}$  is available. Otherwise, the edge is unavailable. For available edges, we initialize the edge embeddings using the extracted features of the corresponding available modalities. On the basis of the complete graph  $\mathcal{G}_P^M$ , an incomplete graph  $\mathcal{G}_P^{M'}$  is obtained through random edge dropout to simulate different modality-missing scenarios. Then, we use a siamese Graph Neural Network (GNN) to encode the complete and incomplete graphs and integrate information from different modalities to obtain consolidated patient representations, as follows:

$$\mathbf{Z} = \text{GNN}(\mathcal{G}_P^M), \mathbf{Z}' = \text{GNN}(\mathcal{G}_P^{M'}) \quad (2)$$

where  $\mathbf{Z}$  and  $\mathbf{Z}'$  are the representations of the same patient but with complete and incomplete modalities. To effectively predict survival risk, even when the modality is missing, the model should focus on features that are shared across modalities. Therefore, we propose a complete-incomplete alignment strategy ( $\mathcal{L}_{\text{Cia}}$ ) to

encourage the alignment of  $\mathbf{Z}$  and  $\mathbf{Z}'$  for uncovering modality-agnostic features. The specific implementation process is shown in Eq. 3.

$$\mathcal{L}_{Cia}(\mathbf{Z}, \mathbf{Z}') = - \sum_{p=1}^N \log \frac{e^{s(\mathbf{z}_p, \mathbf{z}'_p)/\varphi}}{\sum_{q=1}^N e^{s(\mathbf{z}_p, \mathbf{z}'_q)/\varphi}} \quad (3)$$

where  $\mathbf{z}_p$  and  $\mathbf{z}'_p$  are the features of same patient.  $s(\cdot, \cdot)$  is the cosine similarity function,  $\varphi$  is the temperature hyperparameter. It should be noted that this complete-incomplete alignment loss is not only applicable to uncensored data with accurate survival time but also to censored data.

After this, the representation of patient nodes  $\mathbf{Z}$  is used to predict survival risk. After obtaining the survival risk, we use Cox loss ( $\mathcal{L}_{\text{Cox}}$ ) as one of the optimization objectives. The definition of Cox loss is as follows:

$$\mathcal{L}_{\text{Cox}} = \sum_i^B \delta_i \left( -\mathbf{O}(i) + \log \sum_{j: t_j \geq t_i} e^{\mathbf{O}(j)} \right) \quad (4)$$

where  $\mathbf{O}(i)$  is equal to  $\theta x_i$  and  $\mathbf{O}(j)$  is equal to  $\theta x_j$ , which represents the survival risks of the  $i$ -th sample and the  $j$ -th sample respectively.

### 2.3 Event-Conditional Modelling of Censoring

**Dynamic Momentum Accumulation Confidences.** To assess model confidence of each sample in survival prediction, we introduce a method based on the ranking difference between epochs. The difference serves as a stability criterion: larger differences indicate lower confidence, while smaller differences suggest higher confidence. Since confidence can fluctuate, especially in early epochs, we use the dynamic momentum accumulation confidence (DMAC) of each sample calculated at all previous epochs as the confidence of the model as follows:

$$\tau(t) = \lambda \tau(t-1) + (1-\lambda)p(t), \tau(0) = 0 \quad (5)$$

where  $p(t)$  is the reciprocal of the absolute value of the difference between the current epoch and the previous epoch, and  $\lambda$  is the balance parameter that balances the confidence of the current epoch and all previous epochs.

**Selection and Update for Censored Data.** To enhance model performance, we select censored samples with high confidence for training. Specifically, all samples are first sorted by the predicted survival risk from CenSurv. Then, for each censored sample, with the goal of obtaining the best C-index within its  $K$  immediate range, the survival time and survival status of the sample are updated. Finally, traverse and update the survival time and status of all selected censored samples. This process ensures the updated survival time exceeds the original censored time, and the status is updated to death from alive ( $\delta: 0 \rightarrow 1$ ). By iterating over all censored samples, we increase the proportion of censored samples in the training set, which ultimately improves the model performance.

## 2.4 Optimization Objective

In our study, the optimization objective includes Cox loss and Complete-incomplete alignment loss. The total loss of CenSurv is as follows:

$$\mathcal{L}_{All} = \alpha \mathcal{L}_{Cox} + \beta \mathcal{L}_{Cia}. \quad (6)$$

where  $\alpha$  and  $\beta$  are the weights used to balance each loss. It is important to note that the optimization process consists of a preheating stage ( $0 \leq \text{epoch} < 60$ ) and an iterative update stage ( $60 \leq \text{epoch} < 120$ ). In the preheating stage, the status of censored and uncensored data is determined by the original clinical features. Once the model stabilizes, the process enters the iterative update stage, during which the survival events and survival statuses of the censored data are updated and treated as uncensored samples for inclusion in the training process.

## 3 Experiments

### 3.1 Experimental Setting

**Datasets.** We validate the effectiveness of CenSurv on 5 TCGA cancer datasets: Kidney Clear Cell Carcinoma (KIRC), Liver Hepatocellular Carcinoma (LIHC), Esophageal Carcinoma (ESCA), Lung Squamous Cell Carcinoma (LUSC), and Uterine Corpus Endometrial Carcinoma (UCEC).

**Implementation.** We use Adam as the optimizer, and its learning rate is 0.00003. After extensive experiments, the weights for the optimization objectives in this study are  $\alpha=5$  and  $\beta=1$ .  $\lambda$  in DMAC is set to 0.4. We use five-fold cross-validation to evaluate the performance of the proposed model on all 5 datasets, where the model achieving the best performance on the validation set in each fold is selected for testing. The results of all comparative experiments are reproduced based on the original settings and presented in terms of mean and standard deviation.

**Evaluation Metric.** The main evaluation metric used in our study is the Concordance index (C-index) [7,22,20]. We also calculate P-Value using the Logrank test to verify the model’s ability to distinguish between high and low risk groups.

### 3.2 Results with Complete Modalities

**Unimodal v.s. CenSurv.** We implement Cox proportional hazards models in pathological data, genetic data, and clinical records. For pathological data, we compare the SOAT methods TranMIL [19] and HvtSurv [20]. As shown in Table 1, the mean C-index of CenSurv on 5 datasets is 5.7% higher than the best comparative unimodal method. These results show that CenSurv can effectively fuse multi-modal features, thereby achieving more accurate survival prediction.

**Multimodal v.s. CenSurv.** We compare CenSurv with GSCNN [13], MultiSurv [26], OuterP [2], MetricL [1], MCAT [3], HGCN [8], SurvMamba [4] and SurvPath [9]. Compared with SOAT multimodal methods, CenSurv can also

**Table 1.** The C-index of multiple unimodal and multimodal methods on 5 datasets. The best results and the second-best results are highlighted in bold and in underline.

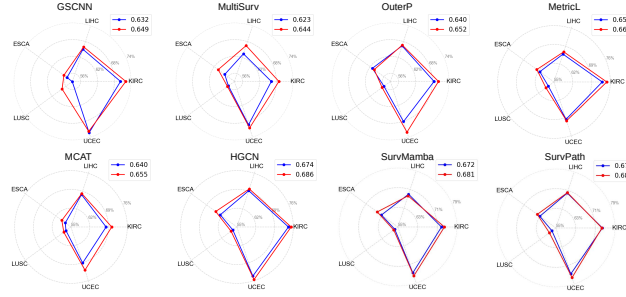
|            | Method         | KIRC         |       | LIHC         |       | ESCA         |       | LUSC         |       | UCEC         |       | Mean         |
|------------|----------------|--------------|-------|--------------|-------|--------------|-------|--------------|-------|--------------|-------|--------------|
| Unimodal   | Cox(c)[10]     | 0.687        | 0.012 | 0.513        | 0.031 | 0.585        | 0.030 | 0.559        | 0.009 | 0.675        | 0.017 | 0.605        |
|            | Cox(g)[10]     | 0.665        | 0.008 | 0.649        | 0.013 | 0.533        | 0.020 | 0.519        | 0.017 | 0.686        | 0.027 | 0.610        |
|            | Cox(p)[10]     | 0.647        | 0.009 | 0.635        | 0.012 | 0.620        | 0.017 | 0.551        | 0.007 | 0.662        | 0.015 | 0.603        |
|            | TranMIL(p)[19] | 0.674        | 0.012 | 0.657        | 0.016 | 0.568        | 0.016 | 0.551        | 0.009 | 0.686        | 0.020 | 0.627        |
|            | HvtSurv(p)[20] | 0.698        | 0.009 | 0.664        | 0.015 | 0.622        | 0.014 | 0.564        | 0.013 | 0.709        | 0.019 | 0.651        |
| Multimodal | GSCNN[13]      | 0.702        | 0.011 | 0.651        | 0.009 | 0.554        | 0.012 | 0.531        | 0.010 | 0.723        | 0.012 | 0.632        |
|            | MultiSurv[26]  | 0.662        | 0.015 | 0.634        | 0.009 | 0.573        | 0.019 | 0.557        | 0.016 | 0.693        | 0.025 | 0.623        |
|            | OuterP[2]      | 0.686        | 0.013 | 0.664        | 0.021 | 0.610        | 0.019 | 0.559        | 0.012 | 0.681        | 0.026 | 0.640        |
|            | MetricL[1]     | 0.730        | 0.022 | 0.659        | 0.019 | 0.615        | 0.022 | <u>0.576</u> | 0.015 | 0.696        | 0.010 | 0.655        |
|            | MCAT[3]        | 0.681        | 0.011 | 0.677        | 0.012 | 0.576        | 0.016 | 0.573        | 0.014 | 0.691        | 0.026 | 0.640        |
|            | HGCN [8]       | <u>0.742</u> | 0.008 | 0.692        | 0.011 | 0.627        | 0.017 | 0.569        | 0.014 | 0.742        | 0.015 | 0.674        |
|            | SurvMamba[4]   | 0.727        | 0.009 | 0.690        | 0.013 | <u>0.633</u> | 0.015 | 0.566        | 0.016 | 0.746        | 0.016 | 0.672        |
|            | SurvPath[9]    | 0.739        | 0.011 | <u>0.699</u> | 0.010 | 0.631        | 0.013 | 0.570        | 0.013 | <u>0.746</u> | 0.013 | <u>0.677</u> |
|            | <b>CenSurv</b> | <b>0.751</b> | 0.009 | <b>0.708</b> | 0.010 | <b>0.691</b> | 0.015 | <b>0.607</b> | 0.015 | <b>0.781</b> | 0.013 | <b>0.708</b> |

achieve the best performance. Specifically, the best-performing multimodal comparative method achieves a mean C-index of 0.677 on 5 datasets, while CenSurv achieves a mean C-index of 0.708 on the same 5 datasets, exceeding 3.1%. In addition, the P-Values of CenSurv for distinguishing high and low risks on the 5 datasets are  $6.92 \times 10^{-15}$ ,  $3.50 \times 10^{-6}$ ,  $5.31 \times 10^{-4}$ ,  $1.57 \times 10^{-3}$  and  $7.41 \times 10^{-8}$ , respectively, which are better than the other comparison methods. These results show that CenSurv outperforms other multimodal methods. In addition, CenSurv contains 2,308.91K parameters and requires 2.26G FLOPs. While not the most lightweight in terms of model complexity, it achieves a good balance between computational cost and predictive performance.

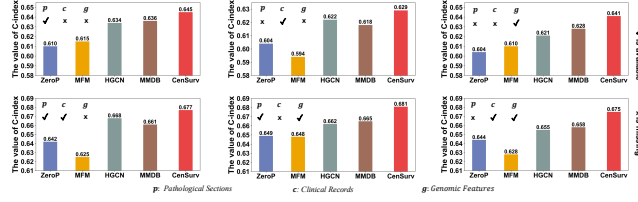
**Robustness Analysis of ECMC.** To verify the robustness of the plug-and-play module ECMC, we conduct experimental comparisons with and without ECMC in 8 multimodal survival prediction methods. As seen in Fig. 3, as a plug-and-play module, after adding the ECMC module, the mean C-index across 5 datasets improved for all methods. For example, compared with the traditional MultiSurv, MultiSurv with the ECMC module improves the mean C-index by 2.1% on 5 datasets. All analyses show the superiority and robustness of ECMC.

### 3.3 Results with Incomplete Modalities

We compare CenSurv with existing methods for handling missing modalities, including ZeroP [21], MFM [25], and HGCN [8] and MMDb [15]. As shown in Fig. 4, CenSurv shows excellent performance in different missing-modality scenarios, and its mean C-index on 5 datasets is ahead of other methods.



**Fig. 3.** The C-index of compared multimodal methods with and without ECMC. Red represents the methods with ECMC, and blue represents the methods without ECMC.



**Fig. 4.** The mean C-index of methods under various modality-missing scenarios.

**Table 2.** The ablation study on 5 datasets, the best results are highlighted in bold.

|                | KIRC         |       | LIHC         |       | ESCA         |       | LUSC         |       | UCEC         |       | Mean         |
|----------------|--------------|-------|--------------|-------|--------------|-------|--------------|-------|--------------|-------|--------------|
| w/o ECMC       | 0.745        | 0.008 | 0.691        | 0.011 | 0.643        | 0.014 | 0.567        | 0.013 | 0.747        | 0.010 | 0.678        |
| w/o BPMG       | 0.747        | 0.011 | 0.699        | 0.009 | <b>0.694</b> | 0.012 | 0.585        | 0.013 | 0.765        | 0.014 | 0.698        |
| w/o DMAC       | 0.733        | 0.010 | 0.662        | 0.015 | 0.637        | 0.013 | 0.557        | 0.014 | 0.745        | 0.015 | 0.667        |
| <b>CenSurv</b> | <b>0.751</b> | 0.009 | <b>0.708</b> | 0.010 | 0.691        | 0.015 | <b>0.607</b> | 0.015 | <b>0.781</b> | 0.013 | <b>0.708</b> |

### 3.4 Ablation Studies

To validate the effectiveness of each module in our proposed method, we sequentially remove each component (ECMC, BPMG, DMAC) while keeping the other components unchanged. As shown in Table 2, as a plug-and-play module, adding the ECMC can improve the performance of CenSurv on all 5 datasets. We designed the BPMG module, which not only enables the model to adapt to various modality-missing scenarios but also enhances the performance across most datasets. Using the DMAC module to select samples with higher comprehensive confidence can avoid adding noisy samples and thus improve model performance. The above analyses also show the effectiveness of each module.

We conduct a quality analysis of updated survival times with censored data by randomly pruning the true survival time between 0 and the true value to simulate censored data. The initial MAE for both uncensored and censored data is



14.5 months, decreasing to 5.7 months after applying ECMC, demonstrating its ability to provide more accurate survival times. In addition, some hyperparameters, such as  $\alpha$ ,  $\beta$ , and  $\lambda$ , are also determined through additional experiments.

## 4 Conclusion

In our study, we propose a bipartite patient-modality graph learning with event-conditional censoring modeling for survival prediction. Overall, CenSurv offers a novel approach for cancer survival prediction. In future work, we will enhance interpretability across multiple modalities, investigate model complexity, and incorporate additional evaluation metrics.

**Acknowledgements.** This work was supported in part by Xinjiang Uygur Autonomous Region Key R&D program (No. 2024B03039-3), the central government guides local funds for scientific and technological development of China (No. ZYYD2025QY25), the Science and Technology Major Project of Changsha (No. kh2502004), Scientific Research Fund of Hunan Provincial Education Department (No. 23A0020), and the Fundamental Research Funds for the Central Universities of Central South University (No. 2024ZZTS0106). This work was carried out in part using computing resources at the High-Performance Computing Center of Central South University.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Cheerla, A., Gevaert, O.: Deep learning with multimodal representation for pan-cancer prognosis prediction. *Bioinformatics* **35**(14), i446–i454 (2019)
2. 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)
3. 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)
4. Chen, Y., Xie, J., Lin, Y., Song, Y., Yang, W., Yu, R.: Survmamba: State space model with multi-grained multi-modal interaction for survival prediction. *arXiv preprint arXiv:2404.08027* (2024)
5. Elforaici, M.E.A., Montagnon, E., Romero, F.P., Le, W.T., Azzi, F., Trudel, D., Nguyen, B., Turcotte, S., Tang, A., Kadoury, S.: Semi-supervised vit knowledge distillation network with style transfer normalization for colorectal liver metastases survival prediction. *Medical Image Analysis* **99**, 103346 (2025)
6. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. *Advances in neural information processing systems* **30** (2017)
7. Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. *Jama* **247**(18), 2543–2546 (1982)

8. Hou, W., Lin, C., Yu, L., Qin, J., Yu, R., Wang, L.: Hybrid graph convolutional network with online masked autoencoder for robust multimodal cancer survival prediction. *IEEE Transactions on Medical Imaging* **42**(8), 2462–2473 (2023)
9. Jaume, G., Vaidya, A., Chen, R., Williamson, D., Liang, P., Mahmood, F.: Modeling dense multimodal interactions between biological pathways and histology for survival prediction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024)
10. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deep-surv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology* **18**, 1–12 (2018)
11. Lee, K.H., Choi, G.H., Yun, J., Choi, J., Goh, M.J., Sinn, D.H., Jin, Y.J., Kim, M.A., Yu, S.J., Jang, S., et al.: Machine learning-based clinical decision support system for treatment recommendation and overall survival prediction of hepatocellular carcinoma: a multi-center study. *npj Digital Medicine* **7**(1), 2 (2024)
12. Liu, H., Shi, Y., Xu, Y., Li, A., Wang, M.: Agnostic-specific modality learning for cancer survival prediction from multiple data. *IEEE Journal of Biomedical and Health Informatics* (2024)
13. Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Velázquez Vega, J.E., Brat, D.J., Cooper, L.A.: Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences* **115**(13), E2970–E2979 (2018)
14. Psutka, S.P., Gulati, R., Jewett, M.A., Fadaak, K., Finelli, A., Legere, L., Morgan, T.M., Pierorazio, P.M., Allaf, M.E., Herrin, J., et al.: A clinical decision aid to support personalized treatment selection for patients with clinical t1 renal masses: Results from a multi-institutional competing-risks analysis. *European urology* **81**(6), 576–585 (2022)
15. Qu, L., Huang, D., Zhang, S., Wang, X.: Multi-modal data binding for survival analysis modeling with incomplete data and annotations. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 501–510. Springer (2024)
16. Riasatian, A., Babaie, M., Maleki, D., Kalra, S., Valipour, M., Hemati, S., Zaveri, M., Safarpour, A., Shafiei, S., Afshari, M., et al.: Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides. *Medical image analysis* **70**, 102032 (2021)
17. Shahin, A.H., Zhao, A., Whitehead, A.C., Alexander, D.C., Jacob, J., Barber, D.: Centime: Event-conditional modelling of censoring in survival analysis. *Medical Image Analysis* **91**, 103016 (2024)
18. Shao, W., Liu, J., Zuo, Y., Qi, S., Hong, H., Sheng, J., Zhu, Q., Zhang, D.: Fam3l: Feature-aware multi-modal metric learning for integrative survival analysis of human cancers. *IEEE Transactions on Medical Imaging* **42**(9), 2552–2565 (2023)
19. 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)
20. Shao, Z., Chen, Y., Bian, H., Zhang, J., Liu, G., Zhang, Y.: Hvturv: Hierarchical vision transformer for patient-level survival prediction from whole slide image. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 37, pp. 2209–2217 (2023)
21. Shen, G., Wang, X., Duan, X., Li, H., Zhu, W.: Memor: A dataset for multimodal emotion reasoning in videos. In: *Proceedings of the 28th ACM international conference on multimedia*. pp. 493–502 (2020)

22. Steck, H., Krishnapuram, B., Dehing-Oberije, C., Lambin, P., Raykar, V.C.: On ranking in survival analysis: Bounds on the concordance index. *Advances in neural information processing systems* **20** (2007)
23. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**(43), 15545–15550 (2005)
24. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **71**(3), 209–249 (2021)
25. Tsai, Y.H.H., Liang, P.P., Zadeh, A., Morency, L.P., Salakhutdinov, R.: Learning factorized multimodal representations. *arXiv preprint arXiv:1806.06176* (2018)
26. Vale-Silva, L.A., Rohr, K.: Long-term cancer survival prediction using multimodal deep learning. *Scientific Reports* **11**(1), 13505 (2021)