Multimodal Hypergraph Guide Learning for Non-Invasive ccRCC Survival Prediction

Jielong Yan[1], Xiangmin Han[1], Jieyi Zhao[2], and Yue Gao[1(✉)]

[1] Tsinghua University, 100084, China
yanjl.jason@gmail.com, {hanxiangmin, gaoyue}@tsinghua.edu.cn
[2] The University of Hong Kong, 999077, Hong Kong, China
zjy800@connect.hku.hk

Abstract. Multimodal medical imaging provides critical data for the early diagnosis and clinical management of clear cell renal cell carcinoma (ccRCC). However, early prediction primarily relies on computed tomography (CT), while whole-slide images (WSI) are often unavailable. Consequently, developing a model that can be trained on multimodal data and make predictions using single-modality data is essential. In this paper, we propose a multimodal hypergraph guide learning framework for non-invasive ccRCC survival prediction. First, we propose a patch-aware global hypergraph computation (PAGHC) module, including a hypergraph diffusion step for capturing correlational structure information and a control step to generate stable WSI semantic embeddings. These WSI semantic embeddings are then used to guide a cross-view fusion method, forming the hypergraph WSI-guided cross-view fusion (HWCVF) to generate CT semantic embeddings, improving single-modality performance in inference. We validate our proposed method on three ccRCC datasets, and quantitative results demonstrate a significant improvement in C-index, outperforming state-of-the-art methods. The source code is available in https://github.com/iMoonLab/PAGHC.

Keywords: Survival prediction · Hypergraph · WSI · CT scan.

1 Introduction

Survival prediction has gained significant attention in medical imaging, aiming to model survival duration from imaging data [6, 23, 27]. WSIs, the gold standard in diagnosis, offer high-resolution visualization of tumor morphology and microenvironment, but require tissue biopsy and expert pathologists [4, 7, 24]. In contrast, CT is a non-invasive 3D imaging modality providing valuable tumor information, though it lacks the detailed pathological data needed for accurate subtype classification [20, 26]. This underscores the need for methods combining the rich pathological insights of WSIs with the non-invasive nature of CT. Early ccRCC prediction often relies on CT scans, as WSIs are commonly unavailable. Thus, we propose a method to obtain accurate WSI semantic embeddings and guide CT training, enabling inference for patients with either WSI or CT data.

Significant efforts have been dedicated to managing these. For WSI, images are segmented into patches, and correlations among patches are captured using multiple instance learning (MIL) [10, 16, 19, 25], graph neural networks (GNN) [1, 15], or hypergraph neural networks (HGNN) [3]. These methods encapsulate tumor morphology and tissue structure information into embeddings for survival prediction. For CT, radiomics and deep learning models extract morphological, textural, and computer vision features to create CT embeddings. Despite these advancements, two primary challenges remain. First, how to accurately construct and utilize patch-level correlations globally within WSIs to achieve stable and accurate embeddings for survival prediction. Second, how to leverage rich pathological information from WSI embeddings during training to enhance CT-based models, ensuring accurate cross-view relationships during inference.

To tackle these challenges, we propose a multimodal hypergraph-guided learning framework for non-invasive ccRCC survival prediction. We propose a patch-aware global hypergraph computation method to generate stable WSI semantic embeddings for accurate survival prediction. It consists of two key steps, namely a diffusion step to capture global hypergraph correlations through information diffusion among patches, and a control step to ensure embedding stability. Using the stable WSI embedding, we apply a cross-view fusion method to achieve precise CT survival predictions, guided by WSI data. Notably, our method requires only WSI data during training, with CT data used only in inference. Experiments on three datasets demonstrate consistent and significant improvements over existing methods. The main contributions of this paper are as follows:

- For stable WSI embedding and effective survival prediction, we propose a patch-aware global hypergraph computation module, including a diffusion step to capture high-order correlations and a control step for external effects.
- Powered by the stable WSI embedding, we propose a cross-view fusion method, guided by WSI during training, to achieve accurate survival prediction based on CT data effectively, with WSI not required during inference.
- The proposed method is validated on three ccRCC datasets. The proposed method consistently outperforms the state-of-the-art methods by a large margin for both WSI-based and CT-based survival prediction tasks.

2 Method

The framework is illustrated in Fig. 1. We first construct a hypergraph from WSI patch features and input them into the PAGHC module. Multiple PAGHC layers capture high-order patch correlations, producing a stable WSI embedding for risk prediction. Next, axial, sagittal, and coronal CT features are extracted, and the HWCVF method generates a CT embedding for risk prediction. During training, the CT embedding is supervised by the WSI embedding, integrating hypergraph correlations and WSI modal information. In the inference phase, no WSI data is required, allowing direct risk prediction using only CT features.

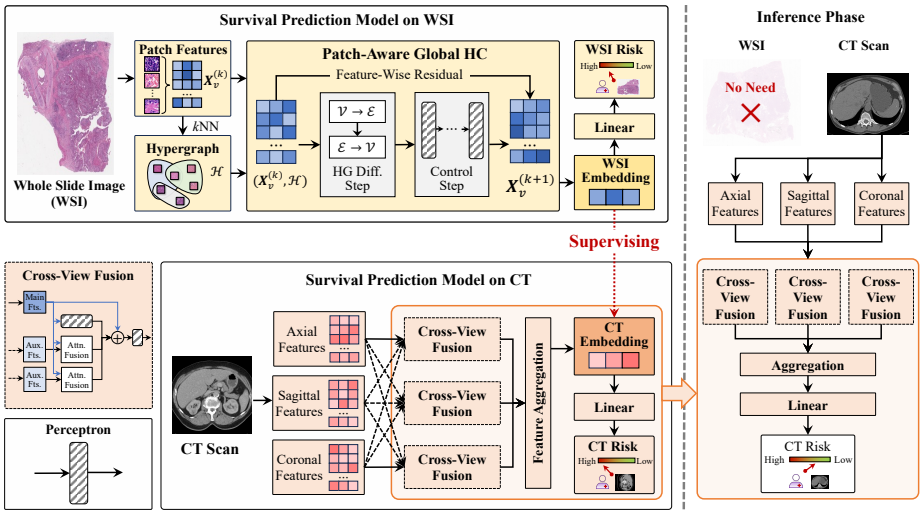

Fig. 1. Illustration of multimodal hypergraph guide learning. 1) The WSI-based survival prediction model utilizes patch features and a constructed hypergraph for patch-aware global hypergraph computation, producing the WSI embedding for risk prediction. 2) The CT-based model employs a cross-view fusion method, integrating three features to obtain a WSI-supervised CT embedding for risk prediction. 3) During inference, no WSI data is required, and the CT model directly predicts the risk.

2.1 Survival Prediction on WSI

Given a patient i with the corresponding WSI, CT scan, survival time t_i, and survival status δ_i, we first apply the OTSU [18] to filter out the background of the WSI and randomly select N_{wsi} patches. Each patch is represented using a neural network pre-trained on a computer vision dataset. As a result, the patient's WSI patch-level visual semantic features are represented as $\mathbf{X}^{(1)} = [\mathbf{x}_1^\top, \mathbf{x}_2^\top, \cdots, \mathbf{x}_{N_{wsi}}^\top] \in \mathbb{R}^{N_{wsi} \times d_{wsi}}$, where d_{wsi} represents the feature dimension. The survival prediction model consists of three main components, namely hypergraph construction, PAGHC for WSI embedding, and risk prediction.

In the hypergraph construction phase, high-order correlations among patches are captured by utilizing their visual semantic features to form hyperedges. Each patch is treated as a vertex v_i, and for each vertex, a hyperedge is constructed using the k-nearest neighbors (kNN) algorithm, with the Euclidean distance between patch features as the metric $dis(v_i, v_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2$. The hypergraph is represented as $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the sets of vertices and hyperedges, and the incidence matrix $\mathbf{H} \in \{0,1\}^{\|\mathcal{V}\| \times \|\mathcal{E}\|}$ has entries $H_{v,e} = \mathbb{I}(v \in e)$. The degree of a vertex v is defined as $d(v) = \sum_{e \in \mathcal{E}} H_{v,e}$, and the degree of a hyperedge e is $\gamma(e) = \sum_{v \in \mathcal{V}} H_{v,e}$. The diagonal degree matrices for vertices and hyperedges are denoted as $\mathbf{D}_v = \text{diag}(d)$ and $\mathbf{D}_e = \text{diag}(\gamma)$, respectively.

Following hypergraph construction, we introduce the PAGHC layer to capture high-order correlations among WSI patches and generate a stable WSI

embedding. The PAGHC layer consists of two steps, namely a hypergraph diffusion step and a control step. In this way, each patch can capture long-range information while avoiding the over-smoothing problem. In the diffusion step, we perform a two-stage process, first diffusing from vertices to hyperedges, and then from hyperedges back to vertices to capture patch correlations. Specifically, the input feature $\mathbf{X}^{(k)}$ is diffused to the hyperedge feature $\mathbf{X}_e^{(k)}$ by multiplication with $\mathbf{H}^\top$, and then the hyperedge feature is diffused back to the vertex feature $\mathbf{X}_v^{(k)}$ by multiplying with $\mathbf{H}$. These steps are expressed as:

$$\mathbf{X}_e^{(k)} = \mathbf{D}_e^{-1}\mathbf{H}^\top\mathbf{X}^{(k)} \quad \text{and} \quad \mathbf{X}_v^{(k)} = \mathbf{D}_v^{-1}\mathbf{H}\mathbf{X}_e^{(k)}. \tag{1}$$

To prevent over-smoothing in dense models, the control step operates directly on each vertex, managing external influences to maintain stability and robustness of the hypergraph computation. The output of the PAGHC layer is the weighted feature-wise residual between the input and the computed results, expressed as:

$$\hat{\mathbf{X}}_v^{(k)} = \mathbf{X}_v^{(k)} + l(\mathbf{X}_v^{(k)}) \quad \text{and} \quad \mathbf{X}^{(k+1)} = \alpha\mathbf{X}^{(k)} + (1-\alpha)\hat{\mathbf{X}}_v^{(k)}, \tag{2}$$

where $l(\cdot)$ represents the control function computed by an MLP, and α represents the keep-rate, controlling the influence of the input feature in each layer.

After K layers of PAGHC, the final output is denoted as $\mathbf{X}^{(K+1)}$. The WSI embedding $\mathbf{x}_{wsi}$ is computed by self-attention aggregation of $\mathbf{X}^{(K+1)}$, given by $\mathbf{x}_{wsi} = \mathrm{Norm}((\mathbf{X}^{(K+1)}\mathbf{W}_{wsi}^q)(\mathbf{X}^{(K+1)}\mathbf{W}_{wsi}^k)^\top\mathbf{W}_{wsi}^v)^\top\mathbf{X}^{(K+1)}$, where $\mathbf{W}_{wsi}^q$, $\mathbf{W}_{wsi}^k \in \mathbb{R}^{d_{wsi}\times d_{attn}}$, and $\mathbf{W}_{wsi}^v \in \mathbb{R}^{N_{wsi}\times 1}$ are learnable weight matrices correlated with the self-attention method, respectively, and $\mathrm{Norm}(\cdot)$ represents normalization. The resulting WSI embedding is passed through a linear layer to predict the WSI risk $p_{wsi}^{(i)}$ for patient i. The training loss function is the negative Cox log partial likelihood (NLL) loss [14], defined as:

$$\mathcal{L}_{wsi} = \mathcal{L}_{nll}^{(wsi)} = \sum_{i=1}^{M} \delta_i(-p_{wsi}^{(i)} + \log \sum_{j\in\{j:t_j\le t_i\}} \exp(p_{wsi}^{(j)})), \tag{3}$$

where M is the batch size, and δ_i indicates whether the i-th sample is censored.

2.2 Survival Prediction on CT

Given the CT scan of patient i, we first extract multi-view features from the axial, sagittal, and coronal planes using a pre-trained neural network, resulting in $\mathbf{Y}_{axial}, \mathbf{Y}_{sagittal}, \mathbf{Y}_{coronal} \in \mathbb{R}^{N_{ct}\times d_{ct}}$, respectively, where N_{ct}, d_{ct} represent the number of slices and feature dimension, respectively. The survival prediction model, HWCVF, consists of two main steps, namely cross-view fusion and aggregation to generate the CT embedding, followed by risk prediction. The details of this process are outlined as follows.

To obtain an accurate CT embedding, we fuse multi-view information through three cross-view fusion modules. For each module, we set one view's features as the main features and the other two as auxiliary features. For instance, in the

first module, the axial feature $\mathbf{Y}_{axial}$ serves as the main feature $\mathbf{F}^{(1)}_{main}$, while the sagittal and coronal features $\mathbf{Y}_{sagittal}$ and $\mathbf{Y}_{coronal}$ are auxiliary features $\mathbf{F}^{(1)}_{aux_1}$ and $\mathbf{F}^{(1)}_{aux_2}$, respectively. The fusion process uses two attention modules to combine the main and auxiliary features. A linear perceptron is then applied to the main feature to capture relevant patterns. The fused features from all three directions are computed as follows for the i-th fusion module:

$$\hat{\mathbf{F}}^{(i)} = l^{(i)}_{CV}(\mathbf{F}^{(i)}_{main}) + \mathrm{AttnF}^{(i)}_1(\mathbf{F}^{(i)}_{main}, \mathbf{F}^{(i)}_{aux_1}) + \mathrm{AttnF}^{(i)}_2(\mathbf{F}^{(i)}_{main}, \mathbf{F}^{(i)}_{aux_2}), \quad (4)$$

where $l^{(i)}_{CV}$ is a self-perceptron function computed by a linear layer, and $\mathrm{AttnF}(\cdot,\cdot)$ is the attention fusion function defined as $\mathrm{AttnF}(\mathbf{F}_{main}, \mathbf{F}_{aux}) = \mathrm{Norm}((\mathbf{F}_{main}\mathbf{W}^q)(\mathbf{F}_{aux}\mathbf{W}^k)^\top)(\mathbf{F}_{aux}\mathbf{W}^v)$. Here, $\mathbf{W}^q, \mathbf{W}^k \in \mathbb{R}^{d_{ct}\times d_{attnF}}$, and $\mathbf{W}^v \in \mathbb{R}^{d_{ct}\times d_{ct}}$ are learnable parameters, respectively. After obtaining $\hat{\mathbf{F}}^{(i)}$, we add a feature-wise residual and apply a perceptron layer to the mixed features, producing the final output $\mathbf{F}^{(i)}$ of the cross-view fusion. The outputs of the three modules are averaged at the slice level, and summed to form the overall CT embedding $\mathbf{y}_{ct}$.

Given the CT embedding, we aim to predict the patient's risk powered by the stable WSI embedding. To begin, we compute the credibility of each CT in the batch by introducing Gaussian noise to the axial, sagittal, and coronal features. For patient i, the change in CT embedding Δ_i is calculated under Gaussian noise $\mathcal{N}(\sigma, \mathbf{\Sigma})$, and the credibility ω_i is given by $\omega_i = \Delta_i/(\sum_{i=1}^{M}\Delta_i)$, where $\Delta_i = \mathbb{E}_{\epsilon\sim\mathcal{N}(\sigma,\mathbf{\Sigma})}\|\mathbf{y}^{(i)}_{ct} - \hat{\mathbf{y}}^{(i)}_{ct}\|_2$. In prediction, the CT embedding is passed through a linear layer to predict the CT risk $p^{(i)}_{wsi}$ for patient i. WSI-guided supervision during training ensures that the WSI and CT embeddings are aligned in both angle and distance, as reflected in the following loss function:

$$\mathcal{L}_{ct} = \alpha_{nll}\mathcal{L}^{(ct)}_{nll} + (1-\alpha_{nll}) \sum_{i\sim \mathrm{p}(s_i|\omega_i)} \left(\alpha_{\angle}(1 - \frac{\mathbf{x}^{(i)}_{wsi}\cdot\mathbf{y}^{(i)}_{ct}}{\|\mathbf{x}^{(i)}_{wsi}\|\|\mathbf{y}^{(i)}_{ct}\|}) + \alpha_d\|\mathbf{x}^{(i)}_{wsi} - \mathbf{y}^{(i)}_{ct}\|^2_2 \right),$$

where $\mathcal{L}^{(ct)}_{nll}$ represents the NLL loss similar in Eq.(3), $\mathrm{p}(s_i|\omega_i)$ is a Bernoulli distribution on ω_i, and $\alpha_{nll}, \alpha_{\angle}, \alpha_d$ are hyperparameters controlling the contributions of NLL loss, angular similarity, and distance, respectively. Notably, since WSI-guided CT only impacts the loss function, during inference, CT risk can be predicted directly from CT data without requiring WSI input.

3 Experiments

3.1 Datasets

The proposed method is evaluated on KIRC, a public cancer dataset from The Cancer Genome Atlas (TCGA) [11], and two ccRCC datasets from cooperative hospitals, the Guizhou Provincial People's Hospital (H1) and the Affiliated Hospital of Guizhou Medical University (H2). Detailed statistics of datasets are presented in Table 1.

Table 1. Dataset statistics.

Datasets	KIRC	H1	H2
#Patients	505	344	288
#WSIs	505	344	288
SST(Days)	11	65	251
LST(Days)	4,537	3,516	3,900
C-rate(%)	66.14	88.66	87.85

3.2 Compared Methods

We compare the proposed method with 7 WSI-based and 5 CT-based survival prediction methods. The experimental code for these methods is either sourced from published implementations or reproduced following detailed descriptions from the respective papers. The specifics of these methods are outlined below.

The WSI survival prediction methods are categorized into three groups. The first group comprises MIL-based methods, including four popular architectures, namely MIL-Attention [10], DTFD [25], TransMIL [19], and CLAM [16]. MIL-Attention employs MIL by treating each patch as an instance. DTFD introduces a two-layer MIL framework for improved task-specific feature extraction. TransMIL integrates a Transformer with MIL to incorporate both morphological and spatial information, while CLAM uses attention to identify important sub-regions and employs instance-level aggregation for feature refinement. The second group consists of two GNN-based models, namely DeepGraphSurv [15] and Patch-GCN [1]. DeepGraphSurv models global graph topology to extract WSI features, while Patch-GCN aggregates instance-level histological features to model both local and global graph structures. The final group includes the HGNN-based model, HGSurvNet [3], which captures high-order correlations between patches and constructs a global WSI representation using a hypergraph.

The CT survival prediction methods are divided into four categories, namely the multi-layer perceptron method (MLP), a GNN-based method (GCN [13]), an HGNN-based method (HGNN [5]), and two transfer learning methods (DDC [22] and KLGDA [17]).

3.3 Implementation

From each WSI, we randomly extract $N = 2,000$ patches, and their semantic features are extracted using the EfficientNet [21] model pre-trained on ImageNet [2]. The number of patches affects WSI sampling coverage. Generally, more patches improve performance at the cost of increased computation. For each CT scan, we use the ResNet [8] model pre-trained on ImageNet to extract axial, sagittal, and coronal features, respectively. All models are trained for 200 epochs using stochastic gradient descent with a batch size of 16, a momentum of 0.9, a learning rate selected from $\{10^{-1}, 10^{-2}, 10^{-3}\}$, and weight decay values from $\{5 \times 10^{-4}, 10^{-4}, 5 \times 10^{-5}\}$. The dataset is randomly split into five-folds, and both the proposed and compared methods employ five-fold cross-validation. The mean and standard errors are reported.

3.4 Results and Discussions

The survival prediction C-index results [9] for WSI and CT data are presented in Tables 2 and 3, respectively. Our method consistently outperforms the compared methods across multiple datasets. For WSI-based inference, the C-index values on the KIRC, H1, and H2 datasets are 0.74, 0.83, and 0.82, respectively. Guided

Table 2. C-index results for different WSI-based methods on KIRC, H1, and H2.

Methods	KIRC	H1	H2	Average
MIL-Attention [10]	$0.7110_{\pm 0.0168}$	$0.7949_{\pm 0.0436}$	$0.7766_{\pm 0.1190}$	$0.7608_{\pm 0.0360}$
DTFD [25]	$0.6981_{\pm 0.0135}$	$0.7524_{\pm 0.0863}$	$0.7884_{\pm 0.1249}$	$0.7463_{\pm 0.0371}$
TransMIL [19]	$0.6814_{\pm 0.0234}$	$0.7533_{\pm 0.0696}$	$0.7967_{\pm 0.0701}$	$0.7438_{\pm 0.0476}$
CLAM [16]	$0.6267_{\pm 0.0132}$	$0.7154_{\pm 0.0471}$	$0.7010_{\pm 0.0469}$	$0.6810_{\pm 0.0389}$
DeepGraphSurv [15]	$0.6227_{\pm 0.0114}$	$0.6845_{\pm 0.0563}$	$0.7172_{\pm 0.0584}$	$0.6748_{\pm 0.0392}$
Patch-GCN [1]	$0.6359_{\pm 0.0324}$	$0.7575_{\pm 0.0594}$	$0.7344_{\pm 0.0911}$	$0.7093_{\pm 0.0527}$
HGSurvNet [3]	$0.7144_{\pm 0.0316}$	$0.8004_{\pm 0.0592}$	$0.7881_{\pm 0.0811}$	$0.7676_{\pm 0.0380}$
PAGHC(Ours)	$\mathbf{0.7434_{\pm 0.0167}}$	$\mathbf{0.8320_{\pm 0.0227}}$	$\mathbf{0.8229_{\pm 0.0676}}$	$\mathbf{0.7994_{\pm 0.0398}}$

Table 3. C-index results for different CT-based methods on KIRC, H1, and H2.

Methods	KIRC	H1	H2	Average
MLP	$0.5698_{\pm 0.0271}$	$0.6230_{\pm 0.0503}$	$0.6300_{\pm 0.0651}$	$0.6076_{\pm 0.0269}$
GCN [15]	$0.5787_{\pm 0.0298}$	$0.6308_{\pm 0.0505}$	$0.6431_{\pm 0.0304}$	$0.6175_{\pm 0.0279}$
HGNN [5]	$0.5963_{\pm 0.0429}$	$0.6493_{\pm 0.0329}$	$0.6504_{\pm 0.0501}$	$0.6320_{\pm 0.0253}$
DDC [22]	$0.6491_{\pm 0.0613}$	$0.6863_{\pm 0.0592}$	$0.6730_{\pm 0.0799}$	$0.6695_{\pm 0.0154}$
KLGDA [17]	$0.6396_{\pm 0.0481}$	$0.6945_{\pm 0.0975}$	$0.6958_{\pm 0.0464}$	$0.6766_{\pm 0.0262}$
HWCVF(Ours)	$\mathbf{0.6682_{\pm 0.0537}}$	$\mathbf{0.7135_{\pm 0.0598}}$	$\mathbf{0.7118_{\pm 0.0463}}$	$\mathbf{0.6978_{\pm 0.0210}}$

from WSI by hypergraph computation, our CT-based method achieves C-index values of 0.66, 0.71, and 0.71 on the same datasets, respectively.

In survival prediction using WSI, our method outperforms the MIL-based methods. Most MIL-based methods (MIL-Attention, DTFD, and CLAM) neglect patch correlations, limiting the transfer of correlational information. In contrast, the Transformer-based MIL method (TransMIL) holds redundant information by modeling the correlation between fully connected patches. Unlike graph edges, hyperedges can connect two or more vertices as a generalized form of edges, allowing our model to capture high-order correlations more accurately. PAGHC and HGSurvNet are hypergraph-based methods, whereas DeepGraphSurv and Patch-GCN are graph-based approaches. The results show that hypergraph methods with high-order connections outperform graph methods. Compared to HGSurvNet, our PAGHC module integrates both diffusion and control steps, increasing layers and facilitating more stable global information perception, as opposed to HGSurvNet relying only on local patch features.

Our method HWCVF outperforms MLP, GCN, and HGNN by the WSI-guided PAGHC during training, capturing global patch-aware WSI information by the hypergraph and overcoming the limitations of single-modality CT models. Unlike transfer learning methods such as DDC and KLGDA, which rely on selecting an appropriate kernel or similar distribution difference between the source and target domains, our method performs better when there are radiological differences in multimodal data such as WSI and CT.

We select one method each from the MIL-based, GNN-based, and HGNN-based categories to compare the KM estimation curves [12], as shown in Fig. 2. The results demonstrate that our method effectively distinguishes between low-

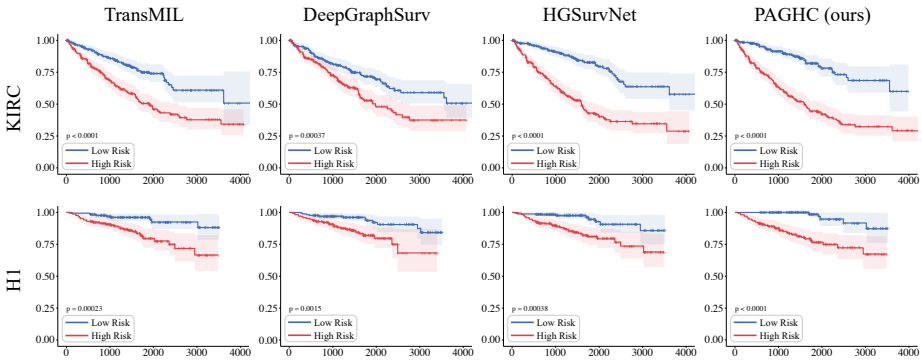

Fig. 2. The KM estimation curves of four methods. The X-axis is survival time/day.

Table 4. Experimental comparison of our method on different settings.

Guided Model			Guided Loss		Dataset		
Patch-GCN	HGSurvNet	PAGHC	Angle	Distance	KIRC	H1	H2
✓				✓	$0.6276_{\pm 0.0710}$	$0.6520_{\pm 0.0778}$	$0.6608_{\pm 0.0653}$
	✓			✓	$0.6397_{\pm 0.0631}$	$0.6812_{\pm 0.0527}$	$0.7001_{\pm 0.0792}$
		✓	✓		$0.6442_{\pm 0.0607}$	$0.6798_{\pm 0.0625}$	$0.6801_{\pm 0.0693}$
		✓		✓	$0.6638_{\pm 0.0415}$	$0.7032_{\pm 0.0580}$	$0.7042_{\pm 0.0521}$
		✓		✓	$\mathbf{0.6682_{\pm 0.0537}}$	$\mathbf{0.7135_{\pm 0.0598}}$	$\mathbf{0.7118_{\pm 0.0463}}$

risk patients and high-risk patients. In addition to the best C-index, our method also has the best binary discrimination ability.

3.5 Ablation Studies

We perform ablation experiments on three datasets to evaluate the impact of different WSI guidance models and loss functions, with results shown in Table 4. Our proposed PAGHC outperforms other methods, especially Patch-GCN, demonstrating that high-order correlations, when used to guide CT model training, produce better performance compared to graph correlational models. Furthermore, PAGHC provides global patch-level perception, beating HGSurvNet, which only captures local patch correlations. Finally, we analyze the individual effects of angle loss and distance loss, showing that their combination generates the best performance, highlighting their complementary roles in WSI guidance.

4 Conclusion

We propose a multimodal hypergraph guide learning framework for non-invasive ccRCC survival prediction. We propose the PAGHC, which consists of a diffusion step for information transfer between vertices and hyperedges, and a control step that ensures stable capture of long-distance correlations globally. By leveraging

patch-aware global hypergraph computation, it effectively captures high-order correlations between patches, leading to improved WSI semantic embeddings for accurate survival prediction. Additionally, we propose a WSI-guided survival prediction model based on CT data, where stable WSI embedding guides the model during training to indirectly capture high-order correlations and modal information, but WSI is not required during inference for accurate prediction. Experimental results on three ccRCC datasets show that our method outperforms state-of-the-art methods.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 62401330), the China Postdoctoral Science Foundation (No. 2024M761727), and the Beijing Natural Science Foundation (No. L242167).

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

References

1. Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F., Mahmood, F.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In: Medical Image Computing and Computer Assisted Intervention. pp. 339–349. Springer (2021)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255. IEEE (2009)
3. Di, D., Zou, C., Feng, Y., Zhou, H., Ji, R., Dai, Q., Gao, Y.: Generating hypergraph-based high-order representations of whole-slide histopathological images for survival prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(5), 5800–5815 (2022)
4. Farahani, N., Parwani, A.V., Pantanowitz, L.: Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathology and Laboratory Medicine International pp. 23–33 (2015)
5. Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y.: Hypergraph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 3558–3565 (2019)
6. Han, X., Zhou, H., Tian, Z., Du, S., Gao, Y.: Inter-intra hypergraph computation for survival prediction on whole slide images. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)
7. Hanna, M.G., Parwani, A., Sirintrapun, S.J.: Whole slide imaging: Technology and applications. Advances in Anatomic Pathology **27**(4), 251–259 (2020)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
9. Heagerty, P.J., Zheng, Y.: Survival model predictive accuracy and ROC curves. Biometrics **61**(1), 92–105 (2005)
10. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning. pp. 2127–2136. PMLR (2018)

11. Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al.: Mutational landscape and significance across 12 major cancer types (2013)
12. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American Statistical Association **53**(282), 457–481 (1958)
13. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2022)
14. Kvamme, H., Borgan, Ø., Scheel, I.: Time-to-event prediction with neural networks and Cox regression. Journal of Machine Learning Research **20**(129), 1–30 (2019)
15. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Medical Image Computing and Computer-Assisted Intervention. pp. 174–182. Springer (2018)
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. Nguyen, A.T., Tran, T., Gal, Y., Torr, P., Baydin, A.G.: KL guided domain adaptation. In: International Conference on Learning Representations (2022)
18. Otsu, N., et al.: A threshold selection method from gray-level histograms. Automatica **11**(285-296), 23–27 (1975)
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. Sluimer, I., Schilham, A., Prokop, M., Van Ginneken, B.: Computer analysis of computed tomography scans of the lung: A survey. IEEE Transactions on Medical Imaging **25**(4), 385–405 (2006)
21. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114. PMLR (2019)
22. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. arXiv preprint arXiv:1412.3474 (2014)
23. Yao, J., Zhu, X., Huang, J.: Deep multi-instance learning for survival prediction from whole slide images. In: Medical Image Computing and Computer Assisted Intervention. pp. 496–504. Springer (2019)
24. 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)
25. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)
26. Zhu, W., Jin, Y., Ma, G., Chen, G., Egger, J., Zhang, S., Metaxas, D.N.: Classification of lung cancer subtypes on CT images with synthetic pathological priors. Medical Image Analysis **95**, 103199 (2024)
27. Zhu, X., Yao, J., Zhu, F., Huang, J.: Wsisa: Making survival prediction from whole slide histopathological images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7234–7242 (2017)