# IHCSurv: Effective Immunohistochemistry Priors for Cancer Survival Analysis in Gigapixel Multi-stain Whole Slide Images

Yejia Zhang*[1,3], Hanqing Chao*[1,4], Zhongwei Qiu[1,4], Wenbin Liu[2], Yixuan Shen[2], Nishchal Sapkota[3], Pengfei Gu[3], Danny Z. Chen[3], Le Lu[1], Ke Yan[1,4], Dakai Jin[1], Yun Bian[2], and Hui Jiang[2]

[1] DAMO Academy, Alibaba Group
{hanqing.chq, qiuzhongwei.qzw}@alibaba-inc.com
[2] Departments of Radiology & Pathology,
Changhai Hospital, Shanghai 200433, China
bianyun2012@foxmail.com, jianghui5131@163.com
[3] University of Notre Dame, Notre Dame IN 46556, USA
chazhang0310@gmail.com
[4] Hupan Lab, 310023, Hangzhou, China

**Abstract.** Recent cancer survival prediction approaches have made great strides in analyzing H&E-stained gigapixel whole-slide images. However, methods targeting the immunohistochemistry (IHC) modality remain largely unexplored. We remedy this methodological gap and propose IHCSurv, a new framework that leverages IHC-specific priors to improve downstream survival prediction. We use these priors to guide our model to the most prognostic tissue regions and simultaneously enrich local features. To address drawbacks in recent approaches related to limited spatial context and cross-regional relation modeling, we propose a spatially-constrained spectral clustering algorithm that preserves spatial context alongside an efficient tissue region encoder that facilitates information transfer across tissue regions both within and between images. We evaluate our framework on a multi-stain IHC dataset of pancreatic cancer patients, where IHCSurv markedly outperforms existing state-of-the-art survival prediction methods. Our code is available on Github.

**Keywords:** Computational Pathology · Cancer Survival Analysis · Multi-stain Immunohistochemistry · Transformers · Clustering

## 1 Introduction

Survival prediction in cancer patients is a central task in computational pathology that facilitates effective treatment planning and clinical decision-making. This task aims to accurately predict a patient's overall survival probability given one or multiple whole-slide images (WSIs) containing stained tumor tissues.

* Contributed equally to this work.

Recent studies have demonstrated promising results by adopting deep learning techniques [19,10,15,9] to mine prognostic features from publicly-available H&E-stained WSI datasets. However, clinical pathology is increasingly recognizing the advantages of multi-stain immunohistochemistry (IHC) images. Like H&E staining, IHC staining also employs the hematoxylin stain to delineate cell, tissue, and tumor structures from dyed cell nuclei. However, IHC can uniquely uncover specific biomarkers that enable cell and cancer sub-type identification which H&E cannot. For instance, IHC can reveal the infiltration rate of CD8-positive T-cells [8] which is correlated with improved outcomes in many solid tumors [8,12]. Despite these promising prospects, modern frameworks for IHC-based survival prediction remain unexplored. Our work seeks to fill this gap by presenting the first general survival prediction framework for multi-stain IHC images.

Our framework aims to maximally leverage prognostic information and demonstrates its effectiveness on pancreatic cancer patients from IHC-stained WSIs targeting the CD4 and CD8 immune receptors. The design tackles two primary research questions. First, we address the challenge of effectively extracting discriminative and spatially-preserved information from multiple gigapixel WSIs using weakly supervised patient-level labels, where each WSI commonly exceeds $100{,}000 \times 100{,}000$ pixels. Second, we investigate how to effectively extract and integrate prognostic priors inherent in IHC-stained images to enhance survival prediction accuracy.

Recent survival prediction works [1] address the challenge of large images and sparse learning signals by employing a multi-instance learning (MIL) formulation where a patient-level bag is populated with image patches that are either randomly sampled [9,6], selected from clusters [19,13], or manually chosen by clinicians [21]. This sub-sampling approach, however, discards the contextual and spatial information around sampled patches. To better preserve these relations, some studies adopted Graph Neural Networks (GNNs) where vertices are vector representations of image patches and edges sparsely connect vertices based on spatial proximity [5] or semantic similarity [10]. Nevertheless, GNNs remain limited to local information passing, failing to aggregate information from distant patches both within and across WSIs. Other methods adopted Transformers [9,16,18] which enable global relation modeling and can preserve spatial information via modified positional embeddings. However, due to the quadratic compute scaling with respect to input lengths, these methods are forced to adopt random patch sub-sampling which incurs the drawbacks mentioned above. Moreover, effective modeling of prognostic information across patches and WSIs remains an open question.

In this work, we introduce IHCSurv, a new survival prediction framework that integrates multi-stain features via a tissue-level Transformer encoder and a patient-level aggregator. Our approach begins with the extraction of vector representations for image patches, followed by the application of a new spectral clustering algorithm designed to organize patches into coherent tissue regions (i.e., clusters) while preserving spatial context. Distinct from H&E-stained images, IHC enables precise detection and categorization of cells as immune, can-

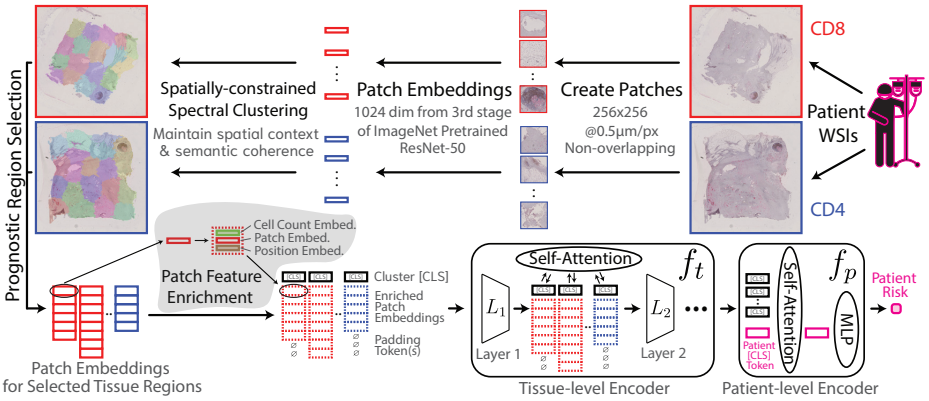


**Fig. 1.** An overview of our proposed method, IHCSurv. The top portion illustrates the pre-processing steps where we obtain patch embeddings and cluster them with our spatially-preserving spectral clustering algorithm. The bottom portion describes how prognostic clusters are selected and processed via the layers, $L_i$, of the tissue-level encoder $f_t$. Patch features are augmented with additional priors such as the cell and position embeddings via a feature enrichment step. Cross-region self-attention modules are also employed to facilitate global information passing. Finally the patient-level encoder $f_p$ regresses risk by aggregating the summary $[cls]$ tokens from each tissue region. The figure is best viewed in color and in digital form.

cerous, or unspecified types without additional annotation. We leverage these priors to select prognostically significant tissue regions and enrich the vector representation of patches. The tissue-level Transformer batches these prognostic regions to compute descriptive tissue region embeddings in parallel and adopts a cross-region attention module for information integration across regions and images. Finally, a patient-level encoder aggregates region embeddings and regresses the final survival prediction. Due to a lack of publicly available multi-stain data, we evaluate IHCSurv on a challenging in-house IHC-stained dataset with 564 pancreatic cancer patients, which notably outperforms recent survival analysis approaches. Our main **contributions** are as follows.

1. To our best knowledge, we are the first to propose a general survival prediction framework for multi-stain IHC analysis. We judiciously design a hierarchical architecture with cross-region attention that effectively captures prognostic features from intra-regional and inter-regional scopes.
2. We leverage accessible IHC-focused priors via cell categorization without incurring large computational burdens and without additional annotation.
3. A spectral clustering algorithm is introduced that preserves local spatial context and synergistically enables information passing between regions.
4. IHCSurv and its components are thoroughly evaluated and they outperform state-of-the-art methods in both single-stain and multi-stain settings.

## 2 Methodology

The task of predicting the overall survival involves analyzing a dataset $\mathcal{D}$ consisting of $N$ patients $\boldsymbol{P}_i = (\boldsymbol{I}_i, t_i, \delta_i)$, $i = 1, \ldots, N$, where $\boldsymbol{I}_i = \{I_i^j\}$ for $j = 1, \ldots, M_i$ denotes a set of $M_i$ whole-slide images (WSIs) for patient $\boldsymbol{P}_i$ ($M_i$ may vary across patients), $t_i$ denotes the observation or right-censored time, and $\delta_i$ is a binary indicator of censorship status. In line with previous studies [5,16], we adopt the discrete-time survival model—detailed in [20]—and stratify uncensored patients into $n$ time intervals written as $[t_0, t_1), \ldots, [t_{n-1}, t_n)$, where $t_1, \ldots, t_{n-1}$ describe interval bounds and $t_0 = 0, t_n = \infty$. Our model $f$ regresses logits $\tau \in \mathbb{R}^n$ for each patient, $\tau_i = f(\boldsymbol{I}_i)$, and maximizes the conditional hazard probability $h(k|\boldsymbol{I}_i)$, where $k$ indexes the time interval $T_i$ in which the target event occurs, $T_i = k$ iff $t_i \in [t_k, t_k + 1)$. From the logits, this can be computed by $h(k|\boldsymbol{I}_i) = \sigma(\tau_i[k])$, where $\sigma()$ is the sigmoid function and $k$ indexes $\tau_i$.

Given the intractability of processing multiple whole-slide images directly, our approach first extracts patch embeddings from each WSI and groups these embeddings into tissue regions (i.e., clusters) using the proposed spacially-constrained spectral clustering method (see §2.1). Enabled by IHC-specific priors, we localize and categorize cells from each patch to later select highly prognostic tissue regions and enrich the extracted patch embeddings. A tissue-level encoder $f_t$ processes the enriched patch embeddings from the selected tissue regions in parallel, employing cross-region attention (see §2.2) to facilitate global information passing. These embeddings are then aggregated via a patient-level encoder $f_p$ to generate the final patient survival prediction, where $\tau_i = f_p(f_t(\boldsymbol{I}_i))$ (see §2.3).

### 2.1 Patch Extraction and Clustering

For each WSI, $I_i^j$, we apply color normalization [14] with a pre-selected template and segment the foreground using [3]. Foreground patches of size 256×256 each are then extracted at 20x magnification and 0.5 microns per pixel (see the top of Figure 1). Similar to CLAM [11], we obtain 1024-dimensional patch embeddings by global average pooling the third stage feature outputs of an ImageNet-pretrained ResNet50, followed by Z-normalization. Features pre-trained on pathology datasets (e.g., HIPT [4]) were also evaluated but ImageNet-based features consistently outperformed them. This is likely due to the lack of IHC stain colors in the H&E pretraining data while ImageNet includes these color variations.

After extracting patch embeddings, clustering is a common step (see [19,1]) to reduce a WSI into manageable segments for tractable feature encoding. Traditional methods like K-means [19,13] and SLIC [17] are typical in survival analysis but introduce notable drawbacks. K-means focuses on semantic features but overlooks spatial continuity, leading to clusters containing scattered patches that lack surrounding spatial context. SLIC, on the other hand, maintains spatial contiguity but falls short in capturing similar semantics due to its reliance on simple color descriptors, struggling with staining noise and dense tissue regions.

Our clustering aims to organize patches into semantically coherent tissue regions where spatial context is preserved. This strategy not only retains more

patch information, avoiding excessive sub-sampling, but also mirrors the holistic manner in which pathologists evaluate WSIs which often focuses at the tissue region level with adequate contiguous sections to gauge broader tissue organizations. To this end, we adopt a spatially-constrained spectral clustering (SCSC) algorithm that modifies the affinity matrix to prioritize patches that are close both semantically and spatially. For patch embeddings $\boldsymbol{x}_i \in \mathbb{R}^{N_p \times 1024}$ and their corresponding spatial coordinates $\boldsymbol{c}_i \in \mathbb{R}^{N_p \times 2}$, where $N_p$ is the number of foreground patches, we define the affinity matrix $A_i = w_s\phi(\boldsymbol{x}_i) + (1 - w_s)\psi(\boldsymbol{c}_i)$. Here, $w_s$ balances semantic versus spatial weighting, with $\phi(\boldsymbol{x})$ quantifying visual similarity and $\psi(\boldsymbol{c})$ quantifying spatial proximity. The semantic affinity $\phi(\boldsymbol{x}) = 1 - \langle \boldsymbol{x}, \boldsymbol{x}^T \rangle / (||\boldsymbol{x}||^2)$ uses the negative cosine distance between pairs of extracted vectors, where $\boldsymbol{x}^T$ indicates the transpose operation and $||\boldsymbol{x}||$ is the $L_2$-norm. The spatial affinity can be described as $\psi(\boldsymbol{c})_{ij} = e^{-\frac{D(c_i, c_j)}{std(\psi(\boldsymbol{c}))}}$, where the value at row $i$ and column $j$ is the exponential of the negative Euclidean distance $D()$ between points $c_i$ and $c_j$, normalized by the overall standard deviation. From $A_i$, the Laplacian is computed and $K$ clusters are obtained via eigen-decomposition. $K$ is dynamically determined for each WSI based on the image foreground size and a hyperparameter for the number of patches per cluster $N_k$. From clinical motivation, we set the target region size to be approximately 2.5x2.5mm (or $N_k$=400 patches per cluster) which adequately captures both regional topology and broader cell community interactions. We'd also like to highlight that although other works have utilized coordinates to inform clustering [7], our novel formulation enables more fine-grained and flexible control over semantic and distance weighting.

### 2.2 Tissue Region Encoding

Before processing image regions with the tissue-level encoder $f_t$, we first leverage the priors available in IHC images to extract useful patch-level cell counts. We utilize the scripting environment in QuPath [2], an open-source pathology image analyzer, to segment cell nuclei using their Watershed implementation via local optical densities and categorize cells by thresholding the average nucleic color value in the DAB color channel. To separate immune cells (dyed brown) with cancer cells (dyed red), we classify a cell as cancer if the average red to green ratio is greater than 1.5.

To select prognostic tissue regions for a patient, we choose the top $N_t$ tissue regions across all patient images $\boldsymbol{I}_i$ sorted in decreasing order by cancer cell count. When there are fewer than $N_t$ tissue regions with cancer cells, we then prioritize the selection of cancer-adjacent regions. Although simple, we found this policy to outperform uniform sampling across all three types, policies that bias toward immune cells, and uniform sampling between stains.

To further enrich input patch features for the prognostic regions (see the bottom-left of Figure 1), we augment each patch embedding with positional, cell count, and stain information. Given patch embeddings $\boldsymbol{e}_k \in \mathbb{R}^{n_k \times 1024}$ from region $k$, where $n_k$ is the number of patches in the region, we obtain the cell

counts embedding $\boldsymbol{e}_k^c \in \mathbb{R}^{n_k \times d_c}$ and spatial embedding $\boldsymbol{e}_k^s \in \mathbb{R}^{n_k \times d_s}$ as a linear projection of their original values. Concretely, $\boldsymbol{e}_k^c = MLP_c((x_c - \mu_c)/std_c)$, where MLP is a linear layer followed by layer normalization, $x_c \in \mathbb{R}^{n_k \times 2}$ is the raw cell count vector, $\mu_c$ is the average across all patches in the patient, and $std_c$ is the patient-wise standard deviation of cell counts. Similarly, the spatial embedding $\boldsymbol{e}_k^s = MLP_s((x_s - min_s)/(max_s - min_s))$, where $x_s \in \mathbb{R}^{n_s \times 2}$ is the raw center coordinate for the patch, and $min_s$ and $max_s$ are the minimum and maximum $x$ and $y$ coordinates, respectively, in the foreground image where the patch originates. The enriched patch embedding concatenates the original, cell, and spatial embeddings $\dot{\boldsymbol{e}}_k = concat(\boldsymbol{e}_k, \boldsymbol{e}_k^c, \boldsymbol{e}_k^s)$.

To process the prognostic regions, we adopt a shared ViT backbone $f_t = \{L_i | i \in [1, n_t]\}$ with $n_t$ Transformer layers, $d_t$ token dimensionality, and $h_t$ heads. Each layer $L_i$ contains a self-attention model $SA_i$ and feed-forward module $FF_i$. Processing patches in contiguous regions ensures spatial context of that semantically-coherent region and also reduces Transformer computational complexity from $O((N_k \cdot n_k)^2)$ to $O(N_k \cdot n_k^2)$. The encoder inputs patch embeddings from all selected regions $\{\dot{\boldsymbol{e}}_k | k \in N_t\}$ where embeddings of different regions are appended in the batch dimension. To address the different input lengths across regions, we first pad shorter sequences with a learnable empty token. Additionally, we drop the vanilla ViT positional embeddings due to already present spatial information and add a learnable stain embedding to inform the model whether the patch is from a CD4 or CD8 image. The tissue-level Transformer outputs the embedding associated with the input $[cls]$ token that is concatenated with the enriched embeddings. Thus, the input of $f_t$ is $concat([cls], \dot{\boldsymbol{e}}_k)$.

Finally, after each self-attention layer $SA_i$, the class tokens are separated and processed through a light self-attention module $SA_i^{cls}$ and replaced with the output [cls] token before $FF_i$. This is visualized in Figure 1 with the self-attention component in $f_t$.

### 2.3 Patient Survival Prediction

The tissue-level encoder outputs $N_t$ tokens representing the embedding for the region. We implement the patient-level aggregation module as a light self-attention module $f_p = MLP_p(SA_p)$ that takes as input the region embeddings (see the bottom-right of Figure 1). In practice, this is similar to appending another $SA_i^{cls}$ module to the model outputs, but with one clear distinction: there is a summarizing patient-level $[cls]$ token. A final linear layer $MLP_p$ maps the patient-level prognostic features to the hazard logits $\tau_i$.

From the definition of $h(k|\boldsymbol{I})$ above, we train $f$ using the negative partial log-likelihood loss [20].

$$\mathcal{L} = -\sum_{(\boldsymbol{I},t,\delta)\in\mathcal{D}} \delta \cdot \log(S(t|\boldsymbol{I})) - (1-\delta)(\log S(k-1|\boldsymbol{I} + \log h(t|\boldsymbol{I})), \quad (1)$$

$$S(k|\boldsymbol{I}) = \prod_{a=1}^{k}(1 - h(a|\boldsymbol{I})). \quad (2)$$

# 3 Experiments and Results

## 3.1 Data, Evaluation, and Implementation

To evaluate our proposed framework, we collect a multi-stain pancreatic cancer dataset comprising of 564 patients and 1185 WSIs. Each patient includes at least one CD4 and one CD8 IHC-stained slide, prepared with DAB on paraffin-fixed assays obtained through serial sectioning. Although less common, a patient's CD4 and CD8 WSIs may originate from different tissue samples, presenting high tumor appearance heterogeneity. The dataset is divided into training, validation, and testing sets with a ratio of 7:1:2, respectively.

Our evaluation employs the concordance index (CI) for rank-based prediction in our main results (presented in §3.2) and the log-ranked test's p-value for risk stratification analysis (see §3.3). We report the test metrics corresponding to the epoch with the lowest validation loss and each reported value is averaged across multiple experimental runs with different random seeds. For baselines, we selected four recent methods covering the prominent survival analysis approaches including multi-instance learning, graph neural networks, and Transformers. We do not incorporate public datasets since there exists no open IHC survival data and the use of popular H&E datasets (e.g., TCGA) may prevent fair direct comparisons since our work targets the IHC modality.

In our clustering approach, we segment each WSI into $N_k = 400$ patches for each cluster (see the end of §2.1 for the clinical justification of $N_k = 400$), applying a weight of $w_s = 0.8$ to balance between semantic and spatial distances. For patch selection, we prioritize $N_t = 24$ regions with the most cancer cells. The framework's architecture integrates a tissue-level encoder $f_t$ ($n_t = 2, h_t = 6, d_t = 96$) with a cross-region self-attention layer with a single head and hidden dimensionality $d_t$. Similarly, the patient-level aggregation module uses a single head and has hidden dimensionality $d_p = 64$. Spatial embeddings are $e^s \in \mathbb{R}^{60}$ and cell embeddings are $e^c \in \mathbb{R}^{48}$. For the survival loss, we binned patients into $n = 4$ intervals. All our experiments run with a batch size of 1 and use the AdamW optimizer with a $4e-5$ learning rate and 0.1 weight decay.

**Table 1. Main comparisons against state-of-the-art survival analysis works**. The CI column reports concordance index scores.

| Multi-Stain (CD4+CD8) | | | | Single-Stain (CD8) | | | |
|---|---|---|---|---|---|---|---|
| Method | Params (M) | FLOPs (G) | CI | Method | Params (M) | FLOPs (G) | CI |
| *Recent Approaches* | | | | | | | |
| DeepAttnMISL$_{20}$ [19] | **0.07** | **1.1** | 0.5310±0.0054 | DeepAttnMISL$_{20}$ [19] | **0.07** | **0.6** | 0.5456±0.0056 |
| PatchGCN$_{21}$ [5] | 1.19 | <u>18.3</u> | 0.5461±0.0150 | PatchGCN$_{21}$ [5] | 1.19 | <u>9.4</u> | 0.5279±0.0193 |
| TransMIL$_{21}$ [15] | 2.67 | 53.2 | 0.5300±0.0247 | TransMIL$_{21}$ [15] | 2.67 | 48.4 | <u>0.5575</u>±0.0090 |
| HVTSurv$_{23}$ [16] | 3.25 | 65.5 | <u>0.5645</u>±0.0082 | HVTSurv$_{23}$ [16] | 3.25 | 33.9 | 0.5363±0.0242 |
| *Ours* | | | | | | | |
| IHCSurv (*ours*) | <u>0.68</u> | 23.3 | **0.6373**±0.0142 | IHCSurv (*ours*) | <u>0.68</u> | 23.3 | **0.6005**±0.0117 |

**Fig. 2.** Kaplan–Meier curve comparisons.

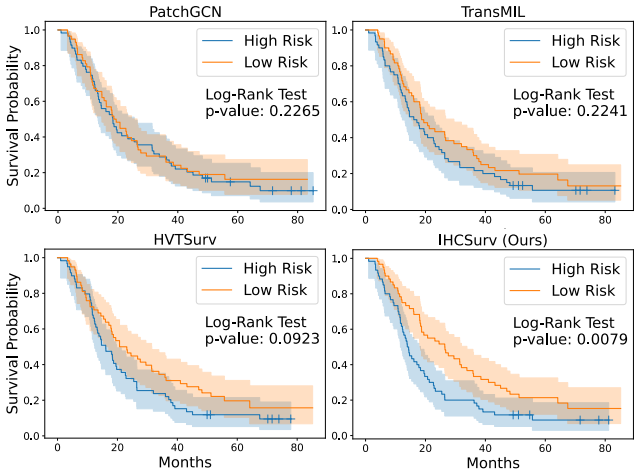


**Fig. 3.** Ablation studies.

| | Method | CI |
|---|---|---|
| | *Clustering* | |
| 1 | K-means ($N_k$=100) | 0.5913 |
| 2 | K-means ($N_k$=400) | 0.6149 |
| 3 | SCSC ($N_k$=100) | 0.6025 |
| 4 | SCSC ($N_k$=400) | **0.6373** |
| | *Region Selection* | |
| 5 | Random ($N_t = 24$) | 0.5945 |
| 6 | Immune Count ($N_t = 24$) | 0.5887 |
| 7 | Cancer Count ($N_t = 12$) | 0.5764 |
| 8 | Cancer Count ($N_t = 24$) | **0.6373** |
| 9 | Cancer Count ($N_t = 30$) | 0.6187 |
| | *Components* | |
| 10 | No Enrichment | 0.6016 |
| 11 | + Stain Embedding | 0.6025 |
| 12 | + Spatial Embedding | 0.6260 |
| 13 | + Cell Embedding | 0.6273 |
| 14 | + Cross Attention | **0.6373** |

### 3.2 Study 1: Overall Performance

Our main performance comparisons are summarized in Table 1, where we study the results for single-stain (CD8 only) and multi-stain (CD4 and CD8) settings. We select CD8 for our single stain experiments for two reasons: 1) CD8 is empirically a stronger baseline compared to CD4, and 2) CD8 has more published evidence to be a potent prognostic predictor while CD4's prognostic value is weaker and often originates from its interaction with other immune-targeted stains.

In the single-stain scenario, our approach surpasses the leading competitor by 0.04 CI, showcasing the efficacy of our clustering and patch enrichment strategies even without leveraging cross-stain context. When incorporating both stains, our model's performance exceeds the next best by nearly 0.08 CI, highlighting the critical role of integrating IHC-based priors.

Compared to DeepAttnMISL [19] and PatchGCN [5], our approach has the additional advantage of modeling global relations with spatial awareness. Moreover, our success against other Transformer-based models [15,16], known for their substantial data requirements, underscores the value of incorporating meaningful priors for improved data efficiency.

### 3.3 Study 2: Survival Stratification

In Fig. 2, we visualize the Kaplan-Meier curves after stratifying patients to low and high risks via their median predicted risk scores. Our method incurs much better separation between risk groups and results in the only statistically significant predictions across all competitors.

### 3.4 Study 3: Ablations

Fig. 3 presents ablation studies of our model's main components. The first four rows compare various tissue region sizes between our clustering method and K-means, demonstrating the advantages of contiguous context from our clustering method and larger cluster sizes. Rows 5 to 9 validate the intuition that prioritizing cancerous regions enhances model performance. Lastly in rows 10 to 14, we incrementally study the contribution of each model component.

## 4 Conclusions

In this work, we proposed a new cancer survival prediction framework that leverages priors in IHC whole-slide images to significantly improve survival prediction accuracy, and designed new approaches to address the drawbacks of patch subsampling. Specifically, we introduced a new spatially-constrained spectral clustering algorithm that improves on K-means clustering to preserve spatial context and patch semantics. By extracting cell counts enabled by IHC staining—a process that incurs no additional annotation and minimal overhead costs—we improved model performance by selecting the most prognostic regions and enhancing the descriptiveness of individual and aggregated patch features. Our results significantly surpassed recent state-of-the-art survival analysis methods, highlighting the benefits of incorporating accessible IHC-based priors. With the promising advantages of the IHC imaging modality, we hope to motivate more studies that explore this avenue and ultimately lead to improved patient care.

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

## References

1. Atabansi, C.C., Nie, J., Liu, H., Song, Q., Yan, L., Zhou, X.: A survey of Transformer applications for histopathological image analysis: New developments and future directions. BioMedical Engineering OnLine **22**(1), 96 (2023)
2. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al.: QuPath: Open source software for digital pathology image analysis. Scientific Reports **7**(1), 1–7 (2017)
3. Bug, D., Feuerhake, F., Merhof, D.: Foreground extraction for histopathological whole slide imaging. In: Bildverarbeitung für die Medizin 2015: Algorithmen-Systeme-Anwendungen. pp. 419–424. Springer (2015)
4. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision Transformers to gigapixel images via hierarchical self-supervised learning. In: CVPR. pp. 16144–16155 (2022)
5. 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: MICCAI. pp. 339–349. Springer (2021)

6. Di, D., Zhang, J., Lei, F., Tian, Q., Gao, Y.: Big-hypergraph factorization neural network for survival prediction from whole slide image. IEEE Transactions on Image Processing **31**, 1149–1160 (2022)
7. Dwivedi, C., Nofallah, S., Pouryahya, M., Iyer, J., Leidal, K., et al.: Multi stain graph fusion for multimodal integration in pathology. In: CVPR. vol. 2021, pp. 1835–1845 (2021)
8. Foersch, S., Glasner, C., Woerl, A.C., Eckstein, M., Wagner, D.C., Schulz, S., Kellers, F., Fernandez, A., Tserea, K., Kloth, M., et al.: Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. Nature Medicine **29**(2), 430–439 (2023)
9. Huang, Z., Chai, H., Wang, R., Wang, H., Yang, Y., Wu, H.: Integration of patch features through self-supervised learning and Transformer for survival analysis on whole slide images. In: MICCAI. pp. 561–570. Springer (2021)
10. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: MICCAI. pp. 174–182. Springer (2018)
11. 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)
12. Mi, H., Sivagnanam, S., Betts, C.B., Liudahl, S.M., Jaffee, E.M., Coussens, L.M., Popel, A.S.: Quantitative spatial profiling of immune populations in pancreatic ductal adenocarcinoma reveals tumor microenvironment heterogeneity and prognostic biomarkers. Cancer Research **82**(23), 4359–4372 (2022)
13. Muhammad, H., Xie, C., Sigel, C.S., Doukas, M., Alpert, L., Simpson, A.L., Fuchs, T.J.: EPIC-survival: End-to-end part inferred clustering for survival analysis, with prognostic stratification boosting. In: Medical Imaging with Deep Learning (2021)
14. Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. Computer Graphics and Applications **21**(5), 34–41 (2001)
15. 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. NeurIPS **34**, 2136–2147 (2021)
16. Shao, Z., Chen, Y., Bian, H., Zhang, J., Liu, G., Zhang, Y.: HVTSurv: Hierarchical vision Transformer for patient-level survival prediction from whole slide image. In: AAAI. vol. 37, pp. 2209–2217 (2023)
17. Wood, R., Domingo, E., Sirinukunwattana, K., Lafarge, M.W., Koelzer, V.H., Maughan, T.S., Rittscher, J.: Joint prediction of response to therapy, molecular traits, and spatial organisation in colorectal cancer biopsies. In: MICCAI. pp. 758–767. Springer (2023)
18. Yan, R., Lv, Z., Yang, Z., Lin, S., Zheng, C., Zhang, F.: Sparse and hierarchical Transformer for survival analysis on whole slide images. IEEE Journal of Biomedical and Health Informatics (2023)
19. 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)
20. Zadeh, S.G., Schmid, M.: Bias in cross-entropy-based training of deep survival networks. TPAMI **43**(9), 3126–3137 (2020)
21. Zhu, X., Yao, J., Huang, J.: Deep convolutional neural network for survival analysis with pathological images. In: BIBM. pp. 544–547. IEEE (2016)

# Supplementary Materials for Paper #495. IHCSurv: Effective Immunohistochemistry Priors for Cancer Survival Analysis in Gigapixel Multi-stain Whole Slide Images

Yejia Zhang*[1,3], Hanqing Chao*[1,4], Zhongwei Qiu[1,4], Wenbin Liu[2], Yixuan Shen[2], Nishchal Sapkota[3], Pengfei Gu[3], Danny Z. Chen[3], Le Lu[1], Ke Yan[1,4], Dakai Jin[1], Yun Bian[2], and Hui Jiang[2]

[1] DAMO Academy, Alibaba Group
[2] Departments of Radiology & Pathology, Changhai Hospital, Shanghai 200433, China
[3] University of Notre Dame, Notre Dame IN 46556, USA
[4] Hupan Lab, 310023, Hangzhou, China

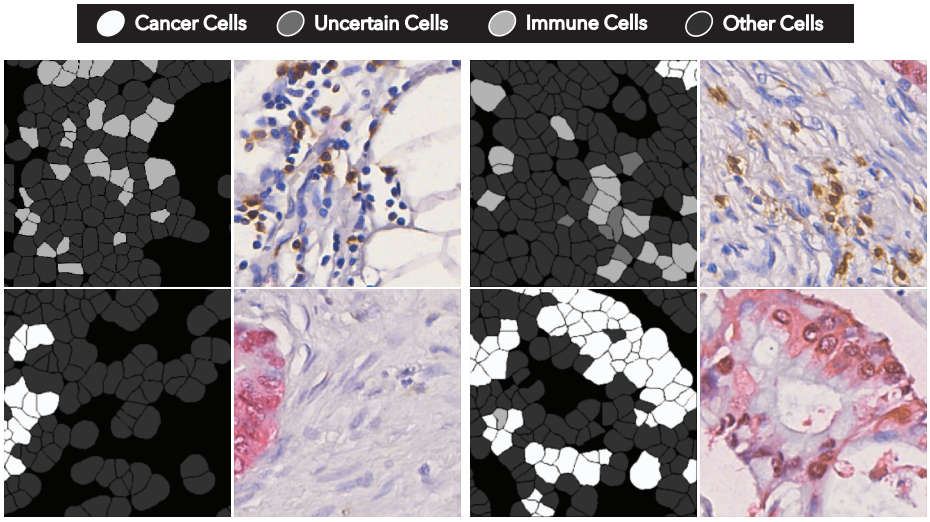


**Fig. 1.** Visualization of extracted cell masks and types. To retrieve patch-wise cell counts, cells are first detected and segmented using QuPath's Watershed algorithm with the computed optical densities from the image's Hematoxylin channel. Next, RGB statistics are extracted from cell nuclei pixels and categorized by thresholding their mean color values. These categories include cancer cells (dyed red), immune cells (dyed brown), uncertain cells (ambiguous cancer or immune cells), and other cells (stromal and other cells without immunohistochemistry stains).

* Contributed equally to this work.

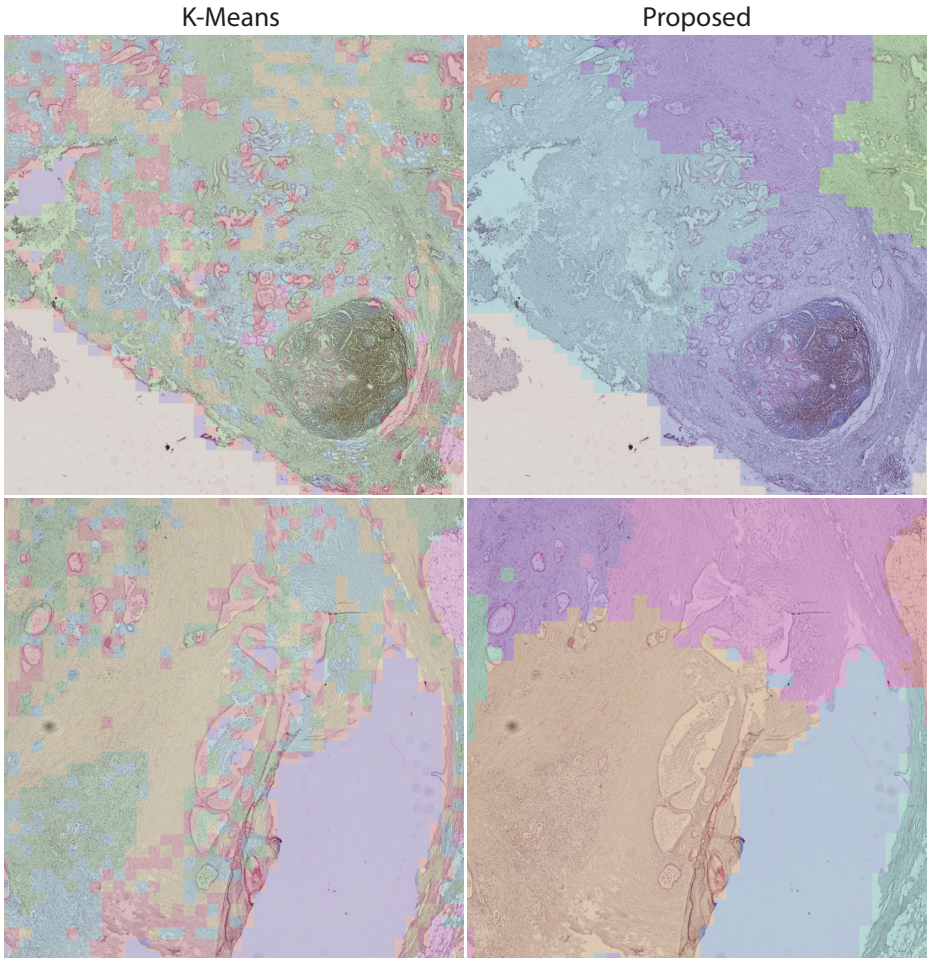


**Fig. 2.** Comparisons between results from k-means clustering (left column) and from our proposed spatially-constrained spectral clustering method (right column). Both approaches are clustered on 1024-dimensional features from an ImageNet-pretrained ResNet-50 with settings of $K = 6$ for k-means (motivated by previous work such as DeepAttnMISL) and $K = 19$ for our proposed clustering method (using $N_k = 400$ patches per cluster). In the top row, we observe segregated intra-cluster patches around an important tertiary lymphoid structure for k-means while ours not only captures the spatial context around the main structures but also preserves contiguous patches around cancerous ducts. The bottom row demonstrates similar patchiness from k-means while ours not only preserves spatial coherence across cancer cells and stroma but also remains faithful to tissue boundaries.