Modeling Dense Multimodal Interactions Between Biological Pathways and Histology for Survival Prediction

Supplementary Material

1. Survival prediction

Following the notation introduced previously, we aim to predict patient survival from the multimodal embedding $\bar{\mathbf{x}}_{\text{Att}} \in \mathbb{R}^{2d}$. Consistently with previous work [14], we define the patient's survival state by: (1) censorship status c, where $c = 0$ represents an observed patient death and $c = 1$ corresponds to the patient's last known follow-up, and (2) a time-to-event t_i, which corresponds to the time between the patient's diagnostic and observed death if $c = 0$, or the last follow-up if $c = 1$. Instead of directly predicting the observed time of event t, we approximate it by defining non-overlapping time intervals $(t_{j-1}, t_j),\ j \in [1, ..., n]$ based on the quartiles of survival time values, and denoted as y_j. The problem simplifies to classification with censorship information, where each patient is now defined by $(\bar{\mathbf{x}}_{\text{Att}}, y_j, c)$. We build a classifier such that each output logit predicted by the network $\hat{y}_j$ correspond to a time interval. From there, we define the discrete hazard function $f_{\text{hazard}}(y_j|\bar{\mathbf{x}}_{\text{Att}}) = S(\hat{y}_j)$ where S is the sigmoid activation. Intuitively, $f_{\text{hazard}}(y_j|\bar{\mathbf{x}}_{\text{Att}})$ represents the probability that the patient dies during time interval (t_{j-1}, t_j). Additionally, we define the discrete survival function $f_{\text{surv}}(y_j|\bar{\mathbf{x}}_{\text{Att}}) = \prod_{k=1}^{j}\left(1 - f_{\text{hazard}}(y_k|\bar{\mathbf{x}}_{\text{Att}})\right)$ that represents the probability that the patient survives up to time interval (t_{j-1}, t_j). These enable us to define the negative log-likelihood (NLL) survival loss [14], which generalizes NLL to data with censorship. Formally, we express it as:

$$\mathcal{L}\Big(\{\bar{\mathbf{x}}_{\text{Att}}^{(i)}, y_j^{(i)}, c^{(i)}\}_{i=1}^{N_{\mathcal{D}}}\Big) = \tag{1}$$

$$\sum_{i=1}^{N_{\mathcal{D}}} -c^{(i)} \log(f_{\text{surv}}(y_j^{(i)}|\bar{\mathbf{x}}_{\text{Att}}^{(i)})) \tag{2}$$

$$+ (1 - c^{(i)}) \log(f_{\text{surv}}(y_j^{(i)} - 1|\bar{\mathbf{x}}_{\text{Att}}^{(i)})) \tag{3}$$

$$+ (1 - c^{(i)}) \log(f_{\text{hazard}}(y_j^{(i)}|\bar{\mathbf{x}}_{\text{Att}}^{(i)})) \tag{4}$$

where $N_{\mathcal{D}}$ is the number of samples in the dataset. Intuitively, Eq. 2 enforces a high survival probability for patients who remain alive after the final follow-up, Eq. 3 enforces that patients that died have high survival up to the time stamp where death was observed, and Eq. 4 ensures that the correct timestamp is predicted for patients for whom death is observed. A thorough mathematical description can be found in [14].

Finally, by taking the negative of the sum of all logits, we can define a patient-level risk used to identify different risk groups and stratify patients.

2. Implementation

2.1. Model training

The code was implemented using Python 3.9, models were implemented in PyTorch and the interpretability was based on Captum [7]. SURVPATH, baselines and ablations were optimized using the RAdam optimizer [8], a batch size of 1, a learning rate of 5×10^{-4}, and 10^{-3} weight decay. The patch encoder yields 768-dimensional embeddings (CTransPath output) that are projected to $d = 256$, the token dimension. The transcriptomics encoder is composed of 2-layer feed-forward networks with alpha dropout [6] to yield pathway tokens. The Transformer is implemented with a single head and layer, without class (CLS) token. The transformer is followed by a layernorm, a feed-forward layer, and a 2-layer classification head. All model training was done using a single NVIDIA RTX 3090Ti.

2.2. Metrics

The models are evaluated using (1) the concordance index (c-index, higher is better), which measures the proportion of all possible pairs of observations where the model's predicted values correctly predict the ordering of the actual survival (ranges from 0.5 (random prediction) to 1.0 (perfect prediction)), and (2) Kaplan-Meier (KM) curves that enable visualizing the probability of survival of patients of different risk groups over a certain period of time. We apply the logrank statistical significance test to determine if the separation between low and high-risk groups is statistically significant (p-value < 0.05).

3. Additional interpretability

A high-level depiction of the proposed multi-level interpretability framework is shown in Fig. 1.

To complement the interpretability analysis presented in the main paper, we further analyze a low and high risk case from **BLCA** (see Fig 2). The histology interpretation indicates that the presence of healthy bladder muscle reduces risk, and pleomorphic tumor cells with foamy cytoplasm contribute to augmenting risk. The majority of important pathways relate to cell cycle control (e.g., G2M checkpoint, SCF β TrCP degradation of em1), metabolism (e.g., fatty acid metabolism), and immune-related function (allograft rejection and IL2 STAT5 signaling). The contributions of pathways to overall risk are also in line with previous literature. For example, previous pathway expression analyses have found G2M checkpoint and immune-related path-

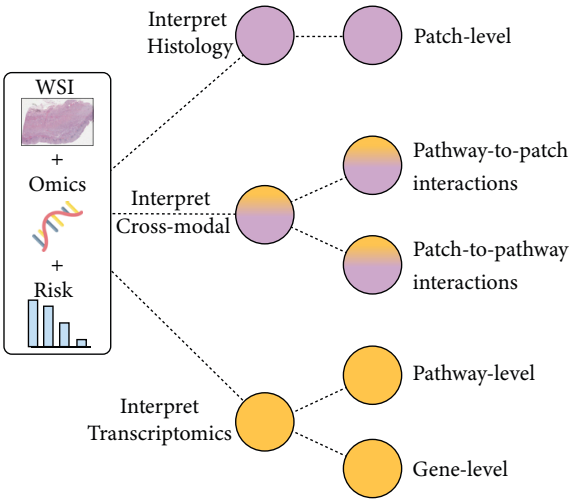

Figure 1. **Multi-level interpretability framework.** From the multimodal input consisting of a WSI and transcriptomic measurements, and the predicted risk, we can attribute risk at slide-, gene- and biological pathway-level. The framework also enables studying pathway-to-patch interactions and patch-to-pathway interactions for unravelling correspondences between the two modalities.

ways to be significant in predicting bladder cancer prognosis [5]. Qualitative assessments of the cross-modal interactions found by SURVPATH are scientifically plausible. For example, the allograft rejection pathway consists of multiple genes that are activated in immune response to allografts and cancer. In the low-risk case, allograft rejection highly attends to tumor-infiltrating lymphocytes and collections of lymphocytes within and near the muscular wall of the bladder. In the higher-risk case, this pathway again attends to collections of inflammatory cells that are interspersed within the muscular wall. The SCF β TrCP degradation of em1 pathway is important in controlling cell division by mitosis. In the low-risk case, this pathway attends to uninvolved bladder muscle, whereas in the high-risk case, the same pathway attends to tumor cells invading the bladder muscle. While there is an overlap between pathways for low and high-risk cases, SURVPATH also identifies pathways present in only one case. For example, in the low-risk case, SURVPATH finds the protein secretion pathway to be highly attending to tumor cells and not the healthy bladder muscle cells. In both cases, the G2M checkpoint pathway (critical for the healthy progress of the cell cycle) is found to be important. In the high-risk case, we see this pathway contributing largely to increasing risk. Interestingly, we also find that this pathway attends to large areas of necrosis, which is reasonable given that aberrations in cell cycle regulation lead to cell death.

4. Additional results

10× results: We also present an analysis of SURVPATH and baselines (Tab. 1), and ablations (Tab. 2) at 10× magnification. Trends from the 20× analysis remain in that (1) SURVPATH achieves overall best performance, (2) transcriptomic baselines remain strong competitors, and (3) multimodal models provide better overall performance. Interestingly, SURVPATH at 10× and 20× provide the same performance (62.9% over the five cohorts).

Kaplan Meier analysis: Fig. 3 shows Kaplan-Meier survival curves of predicted high-risk and low-risk groups at 20×. All patients with a risk higher than the median of the entire cohort are assigned as high risk (red), and patients with a risk lower than the median are assigned low risk (blue). For all five diseases, SURVPATH highlights statistically better discrimination of the two risk groups compared to the best histology baseline (TransMIL), transcriptomics baseline (MLP), and multimodal baseline (MCAT). We believe that SURVPATH can better discriminate between risk groups because a simplified early fusion mechanism allows it to find better correlations between transcriptomics and histology with respect to the patient's risk.

Comparisons with clinical covariates: Clinically, prognostication can be based on patient information such as age, and cancer progression assessment, such as cancer grade. We use a Cox proportional hazards model to predict survival from clinical covariates (Age, Sex, Grade) individually and in combination. We find the SURVPATH outperforms survival prediction from all clinical covariates (Table 3).

Modality attributions: By summing Integrated Gradient (IG) values pre-co-attention per modality, we can derive modality attribution scores (Table 4). We find that histology contributed 77.2% across cohorts, highlighting the need for multi-modality in prognostication.

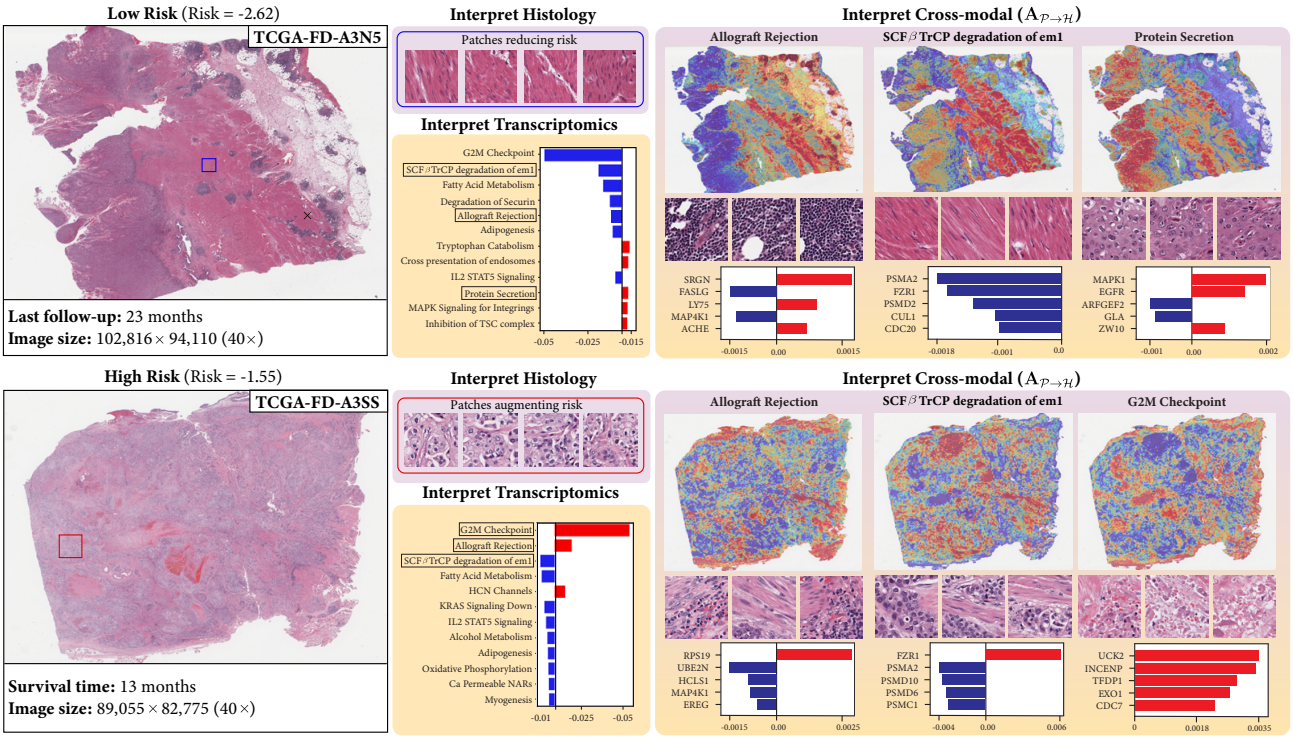

Figure 2. **Multi-level interpretability visualization in a bladder cancer patient. Top:** Low-risk patient. **Bottom:** High-risk patient. Genes and pathways in red increase risk, and those in blue decrease risk. Heatmap colors indicate importance, with red indicating high importance and blue indicating low importance. The pathways and morphologies identified as important in these cases generally correspond well with patterns that have been previously described in bladder urothelial carcinoma (e.g., the G2M checkpoint).

Table 1. Results of SURVPATH and baselines in predicting disease-specific patient survival measured with c-Index (at 10x). Best performance in **bold**, second best underlined. Cat refers to concatenation, KP refers to Kronecker product. All omics and multimodal baselines were trained with the Reactome and Hallmark pathway sets. The omics baselines are carried forward from the 20x experiments.

	Model/Study	BRCA (↑)	BLCA (↑)	COADREAD (↑)	HNSC (↑)	STAD (↑)	Overall (↑)
WSI	ABMIL [4]	0.604±0.110	0.518±0.078	0.652±0.192	0.572±0.070	0.522±0.136	0.574
	AMISL [13]	0.500±0.000	0.500±0.000	0.506±0.012	0.498±0.050	0.500±0.000	0.501
	TransMIL [10]	0.527±0.157	0.541±0.043	0.628±0.193	0.557±0.056	0.516±0.080	0.554
Omics	MLP	0.611±0.080	0.627±0.062	0.625±0.060	0.548±0.045	0.586±0.098	0.599
	SNN [6]	0.528±0.094	0.584±0.113	0.521±0.109	0.550±0.065	0.565±0.080	0.550
	S-MLP [3]	0.512±0.028	0.595±0.114	0.581±0.066	0.542±0.052	0.515±0.081	0.549
Multimodal	ABMIL (Cat) [9]	0.623±0.066	0.619±0.094	0.622±0.165	0.549±0.063	0.547±0.111	0.592
	ABMIL (KP) [2]	0.529±0.099	0.592±0.086	0.640±0.183	0.596±0.039	0.526±0.107	0.577
	AMISL (Cat) [13]	0.508±0.131	0.543±0.069	0.620±0.110	0.539±0.051	0.583±0.104	0.559
	AMISL (KP) [13]	0.551±0.122	0.500±0.068	0.518±0.151	0.523±0.063	0.565±0.062	0.531
	TransMIL (Cat) [10]	0.539±0.072	0.598±0.043	0.632±0.200	0.537±0.065	0.547±0.094	0.571
	TransMIL (KP) [10]	0.538±0.054	0.603±0.043	**0.686**±0.195	0.521±0.111	0.459±0.170	0.561
	MOTCat [12]	0.612±0.156	0.614±0.079	0.569±0.191	0.592±0.080	0.586±0.056	0.595
	MCAT [1]	0.473±0.123	0.545±0.070	0.480±0.243	0.494±0.072	0.433±0.064	0.485
	SURVPATH (Ours)	**0.640**±0.093	**0.628**±0.073	0.675±0.175	**0.605**±0.068	**0.598**±0.081	**0.629**

Table 2. Studying design choices for tokenization (top) and fusion (bottom) in SURVPATHat 10× magnification. **Top:** *Single* refers to no tokenization, using tabular transcriptomics features as a single token. *Families* refers to the set of six gene families in MutSigDB, as used in [1]. *React.+Hallmarks* refers to the main SURVPATH model reported in Table 1. **Bottom:** $A_{\mathcal{P}\to\mathcal{P}}$ and $A_{\mathcal{P}\leftrightarrow\mathcal{H}}$ refers to pathway-to-pathway, pathway-to-patch, and patch-to-pathway interactions, which is the main SURVPATH model reported in Table 1. $\tilde{\mathbf{A}}$ refers to using Nyström attention to approximate $\mathbf{A}$.

	Model/Study	BRCA (↑)	BLCA (↑)	COADREAD (↑)	HNSC (↑)	STAD (↑)	Overall (↑)
Tokenizer	Single	0.617±0.147	0.599±0.077	0.533±0.07	0.544±0.077	0.524±0.117	0.563
	Families	0.534±0.156	0.588±0.060	**0.686**±0.156	0.543±0.077	0.457±0.077	0.562
	Hallmarks	0.609±0.087	0.633±0.090	0.659±0.117	0.601±0.031	0.580±0.052	0.616
	Reactome	**0.665**±0.086	**0.634**±0.077	0.626±0.157	**0.611**±0.067	**0.603**±0.033	0.628
	React.+Hallmarks	0.640±0.093	0.628±0.073	0.675±0.175	0.605±0.068	0.598±0.081	**0.629**
Fusion	$\mathbf{A}_{\mathcal{P}\to\mathcal{P}}$, $\mathbf{A}_{\mathcal{P}\to\mathcal{H}}$	0.589±0.077	0.570±0.099	0.594±0.124	0.568±0.067	0.546±0.135	0.573
	$\mathbf{A}_{\mathcal{P}\to\mathcal{P}}$, $\mathbf{A}_{\mathcal{H}\to\mathcal{P}}$	0.573±0.085	0.577±0.118	0.531±0.221	0.566±0.064	0.521±0.056	0.554
	$\mathbf{A}_{\mathcal{P}\to\mathcal{P}}$, $\mathbf{A}_{\mathcal{H}\to\mathcal{P}}$, $\mathbf{A}_{\mathcal{P}\to\mathcal{H}}$	0.640±0.093	0.628±0.073	0.675±0.175	0.605±0.068	0.598±0.081	**0.629**
	$\tilde{\mathbf{A}}$ [11]	0.495±0.177	0.591±0.068	0.600±0.190	0.508±0.066	0.605±0.075	0.560

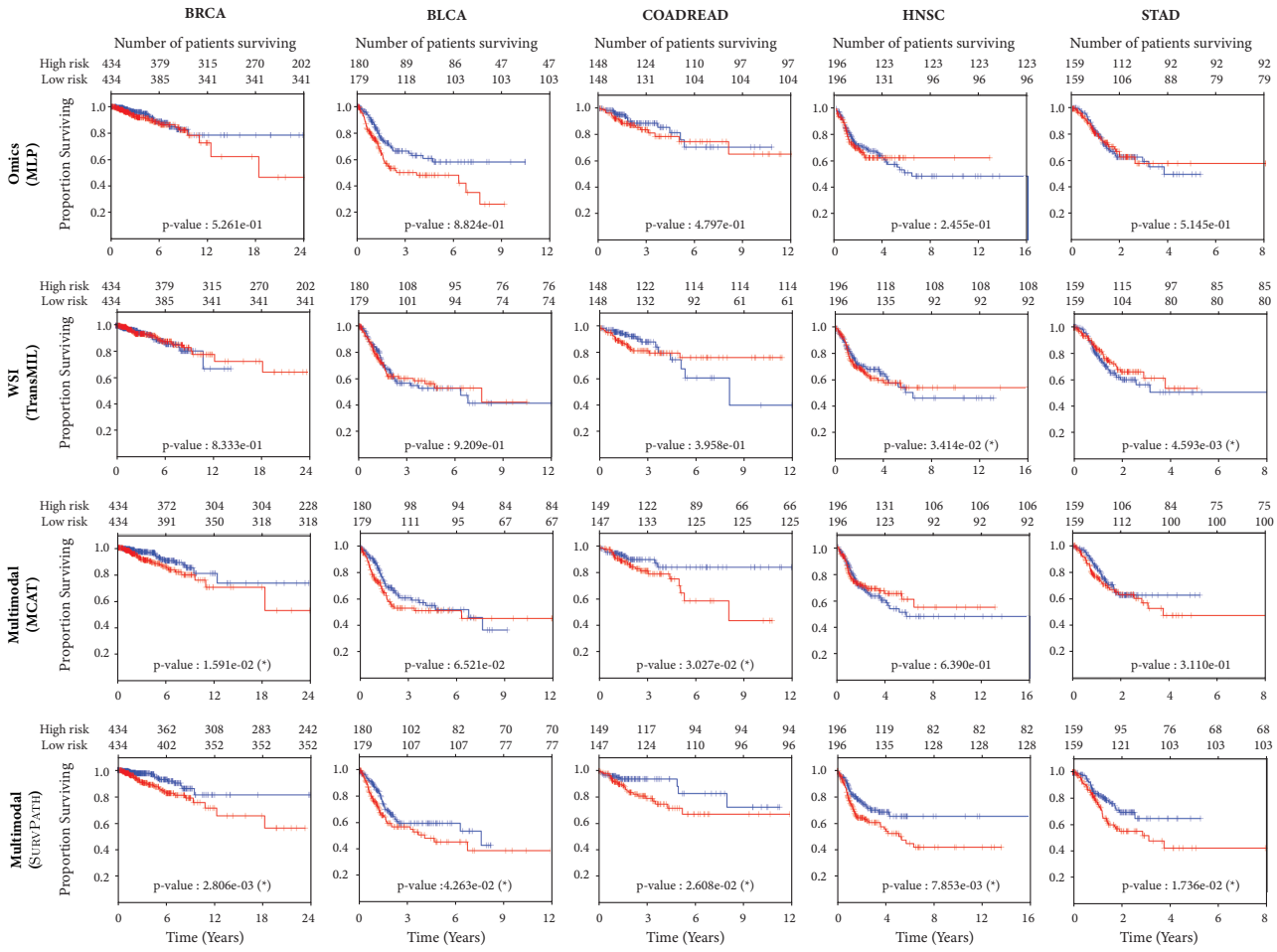

Figure 3. Kaplan Meier curves of SURVPATH, compared against histology, transcriptomics, and multimodal baselines. High (red) and low-risk (blue) groups are identified by using the median predicted risk as cut-off. Logrank test was used to determine statistical significance ($\alpha = 0.05$).

Table 3. Survival prediction results of SURVPATH compared with clinical covariates. Best performance in **bold**, second best underlined. Predictions using clinical covariates are done using Cox proportional hazards model on the same 5-fold cross-validation splits as SURVPATH.

Model/Study	BRCA (↑)	BLCA (↑)	COADREAD (↑)	HNSC (↑)	STAD (↑)	Overall (↑)
Age	0.496±0.086	0.578±0.056	0.357±0.161	0.517±0.073	0.499±0.055	0.489
Sex	0.490±0.011	0.489±0.028	0.542±0.070	0.486±0.035	0.529±0.069	0.507
Grade	0.597±0.078	0.515±0.018	N/A	0.547±0.035	0.552±0.055	0.553
Age + Sex + Grade	0.563±0.055	0.570±0.033	0.655±0.119	0.512±0.093	0.592±0.044	0.578
SURVPATH (Ours)	**0.655**±0.089	**0.625±0.056**	**0.673**±0.170	**0.600**±0.061	**0.592**±0.047	**0.629**

Table 4. Modality attribution percentages. The sum of Integrated Gradients attribution over all modality-specific tokens before co-attention. Scores reported on validation fold with the highest c-index.

Modality/Study	BRCA	BLCA	COADREAD	HNSC	STAD
WSI	0.621±0.251	0.511±0.222	0.971±0.067	0.921±0.060	0.849±0.221
Omics	0.379±0.251	0.489±0.222	0.029±0.067	0.079±0.060	0.151±0.221

References

[1] Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4025, 2021. 3, 4

[2] Richard J. Chen, Ming Y. Lu, Drew F.K. Williamson, Tiffany Y. Chen, Jana Lipkova, Zahra Noor, Muhammad Shaban, Maha Shady, Mane Williams, Bumjin Joo, and Faisal Mahmood. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*, 40(8): 865–878, 2022. 3

[3] Haitham A Elmarakeby, Justin Hwang, Rand Arafeh, Jett Crowdis, Sydney Gang, David Liu, Saud H AlDubayan, Keyan Salari, Steven Kregel, Camden Richter, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598(7880):348–352, 2021. 3

[4] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018. 3

[5] Xuewen Jiang, Yangyang Xia, Hui Meng, Yaxiao Liu, Jianfeng Cui, Huangwei Huang, Gang Yin, and Benkang Shi. Identification of a nuclear mitochondrial-related multi-genes signature to predict the prognosis of bladder cancer. *Frontiers in Oncology*, 11:746029, 2021. 2

[6] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 972–981, Red Hook, NY, USA, 2017. Curran Associates Inc. 1, 3

[7] Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, and Orion Reblitz-Richardson. Captum: A unified and generic model interpretability library for pytorch, 2020. 1

[8] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019. 1

[9] Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A Gutman, Jill S Barnholtz-Sloan, José E Velázquez Vega, Daniel J Brat, and Lee AD Cooper. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, 115(13):E2970–E2979, 2018. 3

[10] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in Neural Information Processing Systems*, 34:2136–2147, 2021. 3

[11] Y. Xiong, Z. Zeng, R. Chakraborty, M. Tan, G. Fung, Y. Li, and V. Singh. Nyströmformer: A nyström-based algorithm for approximating self-attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. 4

[12] Yingxue Xu and Hao Chen. Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 3

[13] Jiawen Yao, Xinliang Zhu, Jitendra Jonnagaddala, Nicholas Hawkins, and Junzhou Huang. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Medical Image Analysis*, 65: 101789, 2020. 3

[14] Shekoufeh Gorgi Zadeh and Matthias Schmid. Bias in cross-entropy-based training of deep survival networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(9):3126–3137, 2021. 1