Deep Survival Analysis from Whole Slide Images: A Multiple Instance Learning Approach

Minyoung Hwang
Department of Artificial Intelligence
Korea University
Seoul, Korea
minyoung58@korea.ac.kr

Dongjoon Lee
Department of Artificial Intelligence
Chung-Ang University
Seoul, Korea
dongzza97@cau.ac.kr

Changhee Lee
Department of Artificial Intelligence
Korea University
Seoul, Korea
changheelee@korea.ac.kr

***Abstract*—Survival analysis plays a critical role in oncology for patient care, but analyzing Whole Slide Images (WSIs) presents challenges due to their immense size and inherent variability. Traditional approaches often rely on manual Region of Interest (ROI) selection, which introduces subjectivity and limits scalability. In this paper, we propose *Surv-MIL*, a novel deep survival model based on Multiple Instance Learning (MIL) that processes WSIs without the need for ROI selection. Our approach divides WSIs into patches, extracts features using a pre-trained encoder, and then aggregates this information using a gated attention mechanism. This enables the model to focus on salient tumor regions while effectively handling the variability in WSI sizes. Evaluated on a real-world dataset, our method demonstrates superior performance compared to other deep survival models. This scalable framework effectively leverages the rich information contained within WSIs for survival analysis, potentially leading to improved prognosis prediction and treatment planning in oncology.**

***Index Terms*—survival analysis, time-to-event analysis, whole slide image, multiple instance learning**

I. Introduction

Survival analysis (also known as time-to-event analysis) involves analyzing the time until a specific event occurs (e.g., death, disease recurrence). The importance of survival analysis lies in its ability to predict patient prognosis and provide crucial information for developing treatment plans tailored to the corresponding patient. Survival analysis using structured data has been extensively studied over the years. Efforts to utilize deep learning in survival analysis have allowed for the capture of complex relationships that linear models cannot. Following the success of deep learning in survival analysis with structured data [1, 2], research expanded to using images through deep learning models by modifying network architectures [3]. These approaches offer the advantage of handling diverse data in medical domains and demonstrate the potential to effectively utilize visual data for survival predictions.

Whole Slide Images (WSIs), obtained from high-resolution microscopy of tissue specimens, present a rich tapestry of diagnostic information including color, tissue architecture, cell morphology, and complex cellular phenotypes [4]. Analysis of these images unlocks a deeper understanding of cancer characteristics such as tumor heterogeneity, microenvironment, and metastatic potential, thus contributing significantly to our knowledge of cancer progression [5, 6]. However, harnessing the full potential of WSIs necessitates overcoming certain challenges: First, their immense size and inherent variability pose difficulties for traditional neural network architectures (e.g., CNNs) that demand fixed input dimensions. Second, accurate WSI interpretation involves identifying specific patterns across diverse locations and tissues scattered throughout different image segments.

To address the challenges of processing WSIs, traditional methods often rely on either cropping fixed-size regions of interest (ROIs) from the images [7] or having experts analyze each WSI to generate tabular data for input. While focusing on specific ROIs or summary information can simplify tumor analysis, it necessitates highly skilled domain experts to manually examine each WSI and identify the most relevant areas [8]. This manual process is not only labor-intensive and time-consuming but also introduces subjectivity and variability into the results, as individual experts may have differing interpretations or be prone to fatigue, leading to inconsistencies in analysis [9]. Therefore, there is a need for a scalable approach that can effectively handle WSIs of varying sizes, eliminating the reliance on manual expert analysis and its inherent limitations.

We propose a novel deep survival model based on multiple instance learning (MIL) to circumvent the need for ROI selection or expert manual assessment. Our approach addresses the challenge of variable WSI sizes by employing a divide-and-aggregate strategy: WSIs are divided into fixed-size patches, features are extracted using a pre-trained encoder, and patch-level features are aggregated with a gated attention mechanism to focus on salient tumor regions. Through experiments on a real-world histopathology brain tumor dataset, we demonstrate that our method outperforms state-of-the-art deep survival models constrained by fixed input sizes.

II. Related Works

A. Deep Survival Models

Deep learning-based survival models [1, 2, 10, 11, 12, 13] have emerged as a powerful alternative to traditional methods like Cox regression [14] and accelerated failure time models [15], offering greater flexibility and overcoming limitations such as linearity and proportional hazard assumptions [16].

979-8-3503-6463-7/24/$31.00 ©2024 IEEE

These models primarily focus on capturing complex (nonlinear) relationships between input features and time-to-event outcomes, enhancing discriminative and predictive power, and ultimately leading to a better understanding of the underlying disease progression. For example, DeepSurv [1] captures complex relationships among features utilizing neural networks, DeepHit [2] provides a direct estimation of time-to-event distribution to overcome the limitations of the proportional hazard assumption, DRSA [10] utilizes an RNN structure for computing hazard estimation as a function of time, and many others [11, 12, 13]. More recently, deep survival models have been further utilized to solve other clinical problems such as longitudinal analysis [17] and treatment effect estimation [18, 19], but improvements for image-based survival analysis, particularly with WSIs, remain limited [3, 20].

B. Deep learning models for WSI

Although deep learning has made significant strides in WSI analysis, existing models often face limitations due to their reliance on fixed-size inputs and pre-defined ROIs [3][21]. This restricts their ability to process WSIs at their full resolution, potentially hindering the extraction of the rich information contained within WSIs. In survival analysis, deep survival models such as [3] and SCNN [20] also depend on pre-specified ROIs. SCNN attempts to overcome the fixed-size input limitation by randomly cropping patches from the ROI, but this approach can introduce noise from irrelevant regions, leading to inefficient training.

MIL has been explored in tasks like classification to address these challenges. It processes individual instances within a set and aggregates their information for a set-level prediction, reducing the computational burden and eliminating the need for pre-defined ROIs. The core principle of applying the MIL in WSI analysis is to divide WSIs into patches, compute patch-level predictions, and aggregate them effectively. For example, [22] employs an image-level decision fusion model trained on histograms of patch-level predictions, to predict the WSI-level label. Similarly, HIPT [23] leverages a hierarchical transformer-based structure to better capture inter-patch relationships.

Aligning with these established MIL practices, our method partitions WSIs into patches and employs attention-based aggregation to extract relevant information. This approach distinguishes our work from traditional deep survival models that rely on pre-defined ROIs, offering a more flexible and potentially comprehensive analysis of WSIs. Moreover, by utilizing hazard estimation, we circumvent the restrictive proportional hazards assumption and address potential model misspecification, a limitation observed in methods like [3] and SCNN [20]. To further enhance performance, particularly in scenarios with limited training data, we leverage components of the pre-trained HIPT architecture for efficient learning.

III. Deep Survival Analysis using Multiple Instance Learning

A. Survival Data with WSIs

Suppose we are given a discrete-time survival dataset comprising N patients, denoted as $\mathcal{D} = \{(\mathbf{x}_i, \tau_i, \delta_i)\}_{i=1}^N$. Each patient i is represented by the input image $\mathbf{x}_i \in \mathcal{X}$ where $\mathcal{X}$ is the input space. The image $\mathbf{x}_i$ for each patient is divided into non-overlapping patches of size $d_p \times d_p$, where $d_p = 256$. There are a total of K_i patches per patient where each patch is indexed by $k \in \{1, \ldots, K_i\}$. Therefore, $\mathbf{x}_i = (\mathbf{p}_i^1, \ldots, \mathbf{p}_i^{K_i})$, where $\mathbf{p}_i^k$ represents the k-th patch-level image. In our method, we use $\mathbf{p}_i^k$ as input to our image encoder. Additionally, the observed survival outcomes for each patient are denoted by $\tau_i \in \mathcal{T}$ and $\delta_i \in \{0, 1\}$. Here, τ_i represents the time until either the event of interest (e.g., death, cancer relapse) or right-censoring (e.g., loss to follow-up) occurs, and δ_i indicates whether the event is observed ($\delta_i = 1$) or right-censored ($\delta_i = 0$). In our approach, we consider survival time to be discrete and the overall duration to be limited, establishing a set of potential survival times as $\mathcal{T} = \{0, \ldots, T_{\text{max}}\}$. This range is capped by a predefined maximum value, T_{max}.

B. Negative Log-Likelihood Loss

The conditional hazard function, represented as $\lambda : \mathcal{X} \times \mathcal{T} \to [0, 1]$, is the immediate risk of an event at a specific time t given the image $\mathbf{x}$ and is defined as $\lambda(t|\mathbf{x}) = \mathbb{P}(T = t|T \geq t, \mathbf{x})$. Based on the conditional hazard function, we can define the survival function conditioned on the input $\mathbf{x}$, denoted as $S : \mathcal{X} \times \mathcal{T} \to [0, 1]$, as follows:

$$S(t|\mathbf{x}) = \mathbb{P}(T > t|\mathbf{x}) = \prod_{t' \leq t} (1 - \lambda(t'|\mathbf{x})). \tag{1}$$

The survival function is a non-increasing function with respect to t, representing the probability that the event will occur after time t given the input $\mathbf{x}$. Similarly, we can define the risk function, $R : \mathcal{X} \times \mathcal{T} \to [0, 1]$, which represents the probability of the event occurring before or at time t given the input $\mathbf{x}$, i.e., $R(t|\mathbf{x}) = \mathbb{P}(T \leq t|\mathbf{x}) = 1 - S(t|\mathbf{x})$.

Given $\mathcal{D}$, we can estimate the conditional hazard function, $\hat{\lambda}$, by minimizing the negative log-likelihood (NLL) loss:

$$\begin{aligned} \mathcal{L}_{NLL} &= -\sum_{i=1}^{N} \Big[\delta_i \log \hat{p}(\tau_i|\mathbf{x}_i) + (1 - \delta_i) \log \hat{S}(\tau_i|\mathbf{x}_i) \Big] \\ &= -\sum_{i=1}^{N} \Big[\delta_i \log \hat{p}(\tau_i|\mathbf{p}_i^1, \ldots, \mathbf{p}_i^{K_i}) \\ &\qquad + (1 - \delta_i) \log \hat{S}(\tau_i|\mathbf{p}_i^1, \ldots, \mathbf{p}_i^{K_i}) \Big] \end{aligned} \tag{2}$$

where $\hat{p}(t|\mathbf{x}) = \hat{\lambda}(t|\mathbf{x}) \hat{S}(t - 1|\mathbf{x})$ represents the estimated probability of an event occurring at time t, i.e., $\mathbb{P}(T = t|\mathbf{x})$. In this context, (2) leverages two key pieces of information from the survival data: That is, when the event is observed (i.e., $\delta_i = 1$), the event occurred at time τ_i, whereas when the event is not observed (i.e., $\delta_i = 0$), this indicates that the event will occur after time τ_i.

C. Surv-MIL Framework

However, the varying number of patches per sample, due to differing WSI sizes, makes it challenging to predefine a maximum value suitable for all K_i. Setting this value too high may lead to an unnecessarily complex network which is prone to overfitting, while setting it too low may fail to accommodate larger WSIs encountered during inference. This variability makes constructing a deep neural network (DNN) with conventional architectures (e.g., CNNs) less suitable. To address this, we propose a novel MIL framework, ***Surv-MIL***, that combines an image encoder and a hazard estimator. The image encoder, consisting of a patch encoder and an aggregator, processes WSIs into sample-level embeddings. An attention mechanism then prioritizes key diagnostic tumor regions within these embeddings, regardless of the number of patches. Finally, the hazard estimator utilizes these embeddings to predict the patient risk of having an event of interest over time.

1) Image Encoder: The image encoder, $f : \mathcal{X} \rightarrow \mathbb{R}^{d_z}$, transforms a WSI into a unified sample-level representation in a d_z-dimensional space. It consists of two components: a patch encoder, f_p, and an aggregator, f_a. The patch encoder processes each individual patch within a WSI to derive patch-level representations. Subsequently, the aggregator combines these patch-level representations into a single, patient-specific representation. The final output given $\mathbf{x}_i$, represented as $\bar{\mathbf{z}}_i = f(\mathbf{x}_i) = f_a(f_p(\mathbf{p}_i^1), \dots, f_p(\mathbf{p}_i^{K_i}))$, encapsulates the entire WSI as a single vector, enabling us to handle WSIs of varying sizes effectively.

Patch Encoder. The patch encoder, $f_p : \mathbb{R}^{d_p \times d_p} \rightarrow \mathbb{R}^{d_z}$, maps each patch from a given WSI to a corresponding embedding in a d_z-dimensional space. In this work, we employ a pre-trained ViT$_{256-16}$ encoder in HIPT [23], which is trained with the DINO [24] framework. This choice allows us to effectively manage the challenges associated with relatively small survival datasets. This encoder processes patches by partitioning each 256×256 patch into non-overlapping 16×16 tokens, augmented with positional embeddings to capture both local and global information. Hence, utilizing the patch encoder, we convert each WSI image, $\mathbf{x}_i$, comprising K_i patches, i.e., $\mathbf{x}_i = (\mathbf{p}_i^1, \dots, \mathbf{p}_i^{K_i})$, into K_i d_z-dimensional representations, i.e., $(\mathbf{z}_i^1, \dots, \mathbf{z}_i^{K_i})$ where $\mathbf{z}_i^k = f_p(\mathbf{p}_i^k)$.

Aggregator. The aggregator, $f_a : \prod_{k=1}^{K_i} \mathbb{R}^{d_z} \rightarrow \mathbb{R}^{d_z}$, integrates a variable number of patch-level representations, i.e., $(\mathbf{z}_i^1, \dots, \mathbf{z}_i^{K_i})$, into a single sample-level embedding, i.e., $\bar{\mathbf{z}}_i$, using an MIL framework. Formally, this can be expressed as:

$$\bar{\mathbf{z}}_i = f_a(\mathbf{z}_i^1, \dots, \mathbf{z}_i^{K_i}) = f_a(f_p(\mathbf{p}_i^1), \dots, f_p(\mathbf{p}_i^{K_i})).$$

For our approach, we choose the gated attention mechanism (GAM) [25] because it offers two key advantages for WSI analysis: i) it can effectively focus on important patch-level information while filtering out irrelevant ones, and ii) it can adaptively handle the varying number of patches across different samples.

Leveraging the GAM, we treat each patch as a piece of evidence contributing to the overall WSI diagnosis. Hence, the sample-level representation can be computed as a weighted sum of path-level representations, i.e., $\bar{\mathbf{z}}_i = \sum_{k=1}^{K_i} a_k \mathbf{z}_i^k$ where the attention weight for each patch is given as

$$a_k = \frac{\exp\big(\mathbf{w}^\top\big(\tanh\big(\mathbf{V}_1 \mathbf{z}_i^{k\top}\big) \cdot \sigma\big(\mathbf{V}_2 \mathbf{z}_i^{k\top}\big)\big)\big)}{\sum_{j=1}^{K_i} \exp\Big(\mathbf{w}^\top\Big(\tanh\Big(\mathbf{V}_1 \mathbf{z}_i^{j\top}\Big) \cdot \sigma\Big(\mathbf{V}_2 \mathbf{z}_i^{j\top}\Big)\Big)\Big)}. \quad (3)$$

Here, $\mathbf{w} \in \mathbb{R}^L$, $\mathbf{V}_1 \in \mathbb{R}^{L \times d_z}$, and $\mathbf{V}_2 \in \mathbb{R}^{L \times d_z}$ are the learnable parameters, where we set L to 128. Functions $\tanh$ and σ represent the tanh and sigmoid functions, respectively.

2) Hazard Estimator: The hazard estimator, $h : \mathbb{R}^{d_z} \times \mathcal{T} \rightarrow [0, 1]$, predicts the hazard rate at each time point $t \in \mathcal{T}$ given the sample-level embedding $\bar{\mathbf{z}}$. Formally, the hazard function for an input WSI $\mathbf{x}$ is expressed as $\hat{\lambda}(\tau|\mathbf{x}) \triangleq h(f(\mathbf{x}), t) = h(\bar{\mathbf{z}}, t)$, which allows us to dynamically capture how the WSI's influence on the hazard rate changes over time, enabling the model to learn complex relationships between the WSI and the time-to-event outcome. Hence, we can rewrite $\hat{p}$ and $\hat{S}$ as

$$\hat{p}(\tau|\mathbf{x}) = h(f(\mathbf{x}), t) \prod_{t' \leq \tau - 1} \big(1 - h(f(\mathbf{x}), t')\big), \quad (4)$$

$$\hat{S}(\tau|\mathbf{x}) = \prod_{t' \leq \tau} \big(1 - h(f(\mathbf{x}), t')\big). \quad (5)$$

Overall, the image encoder and the hazard estimator are trained using the NLL loss, as defined in (6):

$$\mathcal{L}_{\text{NLL}} = -\sum_{i=1}^{N} \Bigg[\delta_i \big(\log h(f(\mathbf{x}_i), \tau_i) + \sum_{t' \leq \tau_i - 1} \log(1 - h(f(\mathbf{x}_i), t')) \big) + (1 - \delta_i) \sum_{t' \leq \tau_i} \log(1 - h(f(\mathbf{x}_i), t')) \Bigg]. \quad (6)$$

IV. Experiments

A. Experiment Setup

Datasets. We use a real-world WSI dataset from The Cancer Genome Atlas (TCGA) [26]. This dataset integrates the TCGA-LGG cohort with lower-grade gliomas (WHO grades II and III) and the TCGA-GBM cohort with glioblastomas (WHO grade IV). Overall, we utilize clinical follow-up information for 769 glioma patients, of whom 388 patients (50.5%) were followed until death and 381 patients (49.5%) were right-censored. In alignment with the requirements of previous studies that necessitate fixed-size ROIs, our dataset includes pre-defined image segments selected by domain experts as provided in [20]. Overall, the dataset encompasses a total of 1505 image segments, with the number of image segments per patient varying from 1 to 16.

Benchmarks. We compare ***Surv-MIL*** with two commonly used deep survival models: ***DeepSurv*** [1] extends the Cox model by employing a DNN to predict individual hazard rates, based on the proportional hazard assumption. ***DeepHit*** [2] is a DNN model that captures the distribution of event times directly. Unfortunately, we are not able to compare with SCNN [20], which is a state-of-the-art deep survival model using fixed-size WSI patches, assuming that domain experts have pre-identified the most salient ROIs for each sample, rather than utilizing the entire image without prior ROI knowledge. Additionally, the SCNN's code is not publicly available, making it impossible to reproduce their results in our study.

Performance Metrics. We evaluate the risk predictions of ***Surv-MIL*** and those of the benchmarks based on how well the predictions discriminate among individual risks and how accurate the predictions are. We utilize the time-dependent concordance index (CI) [27] as the metric of discriminative performance. In addition, we use time-dependent Brier-Score [28] to evaluate the mean squared error of the risk predictions adjusted for survival analysis. To compare the performance of survival models at various time points, we select the 25%, 50%, and 75%-percentiles of time-to-event to report time-dependent CI and Brier-Score, respectively. Higher CI values indicate better discriminative performance, while lower Brier-Scores reflect better model calibration and accuracy.

Our experimental setup is as follows: We use a batch size of 32, with a learning rate of 10^{-3} for DeepHit and 10^{-4} for the other models, which are set through hyperparameter search. The Adam optimizer is employed for training, with a maximum of 100 epochs. The optimal epoch selection is based on the validation CIs averaged at 25%, 50%, and 75% percentiles of time-to-event. After completing all 100 epochs, the weights from the best-performing epoch according to this criterion are chosen for evaluation. For a fair comparison, all models share the same hidden dimension and depth.

Crucially, we leverage the pretrained ViT_{256-16} encoder from HIPT to extract vital information from WSI patches. This ensures a fair comparison by utilizing the same embeddings (generated by the frozen pre-trained encoder) across our method and all the benchmarks. Since benchmark models are unable to process multiple segments simultaneously and require ROIs, we randomly select one segment from the ROI-defined segments for each sample and use it as the input for benchmark models. During inference, we repeat this process 10 times, averaging the performance across these iterations for the final result. In contrast, our method can utilize multiple segments, so it performs training and inference using the entire segment for each sample. All results are reported by 10 iterations of random 64/16/20 training/validation/testing splits.

B. Experiment Result

Table I shows the discriminative and predictive performance of each model at the 25%, 50%, and 75% time points, along with their average. The best performance metrics are highlighted in bold, while the second-best are underlined for easy identification. The results in Table I demonstrate that our proposed method provides significant performance gain in discriminative performance while providing comparable predictive performance to the best-performing benchmarks, highlighting the effectiveness of our approach.

To show that the performance improvement is due to the effective representation obtained by using our approach and not due to the changes in how we derive the risk function (i.e., our hazard estimation), we conduct an ablation study.

TABLE I
TIME-DEPENDENT CI AND BRIER-SCORE

Methods	**CI**			
	25%	50%	75%	Average
DeepSurv	0.700±0.037	0.724±0.013	0.731±0.017	0.718±0.021
DeepHit	0.681±0.029	0.702±0.039	0.684±0.048	0.689±0.032
Ours (w/o MIL)	0.703±0.039	0.731±0.020	0.737±0.014	0.723±0.021
Ours	**0.717±0.036**	**0.739±0.026**	**0.743±0.021**	**0.733±0.025**

Methods	**Brier-Score**			
	25%	50%	75%	Average
DeepSurv	**0.132±0.023**	**0.191±0.012**	0.181±0.021	**0.168±0.007**
DeepHit	0.154±0.036	0.308±0.041	0.428±0.058	0.297±0.037
Ours (w/o MIL)	0.148±0.017	0.196±0.015	0.177±0.011	0.174±0.010
Ours	0.150±0.011	0.192±0.007	**0.136±0.013**	0.172±0.003

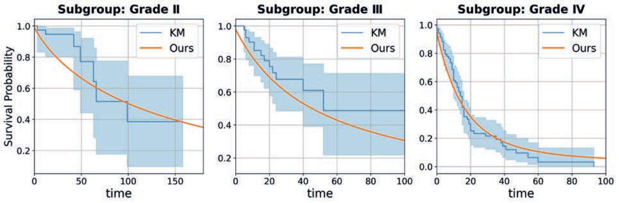

Fig. 1. Survival curves across different WHO Grades

Here, we introduce a variant of our model trained and tested by randomly selecting a single segment, identical to the other benchmarks. The performance results of this setup, including our model without MIL (w/o MIL), are reported in Table I for comparison. Notably, the discriminative performance significantly decreases without the MIL framework.

Fig. 1 presents survival curves for various WHO grades, comparing the Kaplan-Meier (KM) estimates against those predicted by our model. These curves showcase how well our model performs across different tumor grades, providing a visual assessment of its predictive accuracy in relation to the established KM method.

V. CONCLUSION

We introduce a novel deep learning approach for survival analysis using WSIs. Our method eliminates the need for fixed-size inputs and pre-defined ROIs, making it adaptable to WSIs of any size and removing the reliance on expert knowledge for tumor region selection. Our approach divides the WSI into patches and then employs a gated attention mechanism to combine the patch embeddings, creating a comprehensive sample-level representation. Experimental results demonstrate that our method outperforms other models on key metrics, highlighting its effectiveness in survival analysis tasks.

ACKNOWLEDGMENTS

This work was supported by the Institute of Information and communications Technology Planning & Evaluation (IITP) funded by the Korea government (MSIT) under Grant 2021-0-01341 by the AI Graduate School Program (CAU) and

2019-0-00079 by the AI Graduate School Program (Korea University).

References

[1] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network," *BMC medical research methodology*, vol. 18, pp. 1–12, 2018.

[2] C. Lee, W. Zame, J. Yoon, and M. Van Der Schaar, "Deephit: A deep learning approach to survival analysis with competing risks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[3] X. Zhu, J. Yao, and J. Huang, "Deep convolutional neural network for survival analysis with pathological images," in *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2016, pp. 544–547.

[4] R. C. Melo, M. W. Raas, C. Palazzi, V. H. Neves, K. K. Malta, and T. P. Silva, "Whole slide imaging and its applications to histopathological studies of liver disorders," *Frontiers in medicine*, vol. 6, p. 310, 2020.

[5] S.-Y. Yoo, H. E. Park, J. H. Kim, X. Wen, S. Jeong, N.-Y. Cho, H. G. Gwon, K. Kim, H. S. Lee, S.-Y. Jeong *et al.*, "Whole-slide image analysis reveals quantitative landscape of tumor–immune microenvironment in colorectal cancers," *Clinical Cancer Research*, vol. 26, no. 4, pp. 870–881, 2020.

[6] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[7] A. Kiani, B. Uyumazturk, P. Rajpurkar, A. Wang, R. Gao, E. Jones, Y. Yu, C. P. Langlotz, R. L. Ball, T. J. Montine *et al.*, "Impact of a deep learning assistant on the histopathologic classification of liver cancer," *NPJ digital medicine*, vol. 3, no. 1, p. 23, 2020.

[8] M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, and B. Srinivasan, "A generalized deep learning framework for whole-slide image segmentation and analysis," *Scientific reports*, vol. 11, no. 1, p. 11579, 2021.

[9] J. G. Elmore, G. M. Longton, P. A. Carney, B. M. Geller, T. Onega, A. N. Tosteson, H. D. Nelson, M. S. Pepe, K. H. Allison, S. J. Schnitt *et al.*, "Diagnostic concordance among pathologists interpreting breast biopsy specimens," *Jama*, vol. 313, no. 11, pp. 1122–1132, 2015.

[10] K. Ren, J. Qin, L. Zheng, Z. Yang, W. Zhang, L. Qiu, and Y. Yu, "Deep recurrent survival analysis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4798–4805.

[11] M. F. Gensheimer and B. Narasimhan, "A scalable discrete-time survival model for neural networks," *PeerJ*, vol. 7, p. e6257, 2019.

[12] D. Jarrett, J. Yoon, and M. van der Schaar, "Dynamic prediction in clinical survival analysis using temporal convolutional networks," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 424–436, 2019.

[13] Z. Wang and J. Sun, "Survtrace: Transformers for survival analysis with competing events," in *Proceedings of the 13th ACM international conference on bioinformatics, computational biology and health informatics*, 2022, pp. 1–9.

[14] J. Fox and S. Weisberg, "Cox proportional-hazards regression for survival data," *An R and S-PLUS companion to applied regression*, vol. 2002, 2002.

[15] L.-J. Wei, "The accelerated failure time model: a useful alternative to the cox regression model in survival analysis," *Statistics in medicine*, vol. 11, no. 14-15, pp. 1871–1879, 1992.

[16] C. Lee, C. Lee, T. Ha, and S. Cho, "Survey: Strategies for loss-based discrete-time hazard and survival function estimation," in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 2022, pp. 844–846.

[17] C. Lee, J. Yoon, and M. Van Der Schaar, "Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 122–133, 2019.

[18] A. Curth, C. Lee, and M. van der Schaar, "Survite: Learning heterogeneous treatment effects from time-to-event data," *Advances in Neural Information Processing Systems*, vol. 34, pp. 26 740–26 753, 2021.

[19] S. Schrod, A. Schäfer, S. Solbrig, R. Lohmayer, W. Gronwald, P. J. Oefner, T. Beißbarth, R. Spang, H. U. Zacharias, and M. Altenbuchinger, "Bites: balanced individual treatment effect for survival data," *Bioinformatics*, vol. 38, no. Supplement_1, pp. i60–i67, 2022.

[20] A. Parashar, M. Rhu, A. Mukkara, A. Puglielli, R. Venkatesan, B. Khailany, J. Emer, S. W. Keckler, and W. J. Dally, "Scnn: An accelerator for compressed-sparse convolutional neural networks," *ACM SIGARCH computer architecture news*, vol. 45, no. 2, pp. 27–40, 2017.

[21] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature medicine*, vol. 24, no. 10, pp. 1559–1567, 2018.

[22] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2424–2433.

[23] R. J. Chen, C. Chen, Y. Li, T. Y. Chen, A. D. Trister, R. G. Krishnan, and F. Mahmood, "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 144–16 155.

[24] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in

self-supervised vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9650–9660.

[25] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International conference on machine learning*. PMLR, 2018, pp. 2127–2136.

[26] J. Liu, T. Lichtenberg, K. A. Hoadley, L. M. Poisson, A. J. Lazar, A. D. Cherniack, A. J. Kovatich, C. C. Benz, D. A. Levine, A. V. Lee *et al.*, "An integrated tcga pan-cancer clinical data resource to drive high-quality survival outcome analytics," *Cell*, vol. 173, no. 2, pp. 400–416, 2018.

[27] T. A. Gerds, M. W. Kattan, M. Schumacher, and C. Yu, "Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring," *Statistics in medicine*, vol. 32, no. 13, pp. 2173–2184, 2013.

[28] U. B. Mogensen, H. Ishwaran, and T. A. Gerds, "Evaluating random forests for survival analysis using prediction error curves," *Journal of statistical software*, vol. 50, no. 11, p. 1, 2012.