Ensemble of Prior-guided Expert Graph Models for Survival Prediction in Digital Pathology

Vishwesh Ramanathan[1,2*], Pushpak Pati[3*], Matthew McNeil[1,2], and Anne L. Martel[1,2]

[1] Department of Medical Biophysics, University of Toronto, Toronto, ON, CA
[2] Sunnybrook Research Institute, Toronto, ON, CA
[3] Independent Researcher
vishwesh.ramanathan@mail.utoronto.ca

Abstract. Survival prediction in pathology is a dynamic research field focused on identifying predictive biomarkers to enhance cancer survival models, providing valuable guidance for clinicians in treatment decisions. Graph-based methods, especially Graph Neural Networks (GNNs) leveraging rich interactions among different biological entities, have recently successfully predicted survival. However, the inherent heterogeneity among the entities within tissue slides significantly challenges the learning of GNNs. GNNs, operating with the homophily assumption, diffuse the intricate interactions among heterogeneous tissue entities in a tissue microenvironment. Further, the convoluted downstream task relevant information is not effectively exploited by graph-based methods when working with large slide-graphs. We propose a novel prior-guided, edge-attributed tissue-graph construction to address these challenges, followed by an ensemble of expert graph-attention survival models. Our method exploits diverse prognostic factors within numerous targeted tissue subgraphs of heterogeneous large slide-graphs. Our method achieves state-of-the-art results on four cancer types, improving overall survival prediction by 4.33% compared to the competing methods. Our code is publically available on https://github.com/Vishwesh4/DGNN

Keywords: Survival Prediction · Graph Representation Learning · Digital Pathology · Segmentation Prior

1 Introduction

Machine learning in computational pathology is advancing survival prediction by analyzing patient data and extracting biomarkers from whole slide images (WSIs) [20, 21, 27], and thereby offering crucial guidance to clinicians by estimating survival likelihood. Traditionally, survival prediction methods relied on handcrafted features and the Cox proportional hazards model [3, 8]. However, recent advancements, such as Multiple Instance Learning (MIL) and graph-based approaches, have gained traction. MIL [11, 20, 21] treats WSIs as bags

* Equal contributions

of patches, encodes the patches via pre-trained feature extractors, and aggregates patch features using pooling operations, *e.g.*, attention pooling [10, 11], Nystromformer [20, 26], and cluster-based pooling [27]. Conversely, graph-based methods [6, 13, 25] directly operate on WSIs by leveraging both local and global morphological and topological relationships within WSIs. By constructing WSI-graphs, *e.g.*, patch-graph [6, 14, 29, 30] and hierarchical graph [25], these methods apply graph convolutions and various pooling strategies to derive slide-level representations. Graph-based methods are gaining traction for capturing contextual and hierarchical information with noted prognostic significance [6, 12].

However, the predictive performance of graph-based methods is hindered by two major challenges. First, relevant prognostic factors are often obscured within large WSI-graphs, making them cumbersome to identify [31]. For example, pathologists focus on spatial distributions of tumors and tumor-infiltrating lymphocytes (TILs) in the stroma, while disregarding lymphocytic activities outside tumor beds for risk analysis [19]. Emphasizing such factors among numerous interactions is difficult when working with large WSI-graphs. Second, the learnability of graph-based methods is impacted by the misalignment between the homophilic assumption of graph convolutions and the heterogeneous composition of tissue environments. The former dictates connectivity among semantically similar entities [9]. In contrast, the latter includes the spatial organization of diverse entities, *e.g.*, the coexistence of epithelial cells and TILs in ductal carcinoma *in situ*. Applying graph convolutions to these heterogeneous spatial graphs leads to erroneous message passing among dissimilar entities and smoothens local entity attributes, degrading the global slide-representation.

In this paper, we address the aforementioned challenges by proposing a novel prior-guided tissue-graph construction and an ensemble of expert graph-models. Firstly, we employ a tissue classification and a segmentation model to hierarchically categorize the tissue types into *tumor*, *tumor-associated stroma*, and *other* classes. Then, we construct edge-attributed directed WSI-graphs using subsets of tissue types. This tissue-graph model serves two main purposes: (1) to deconvolve a complete-graph into focused sub-graphs, enhancing informative prognostic factors, and (2) to differentiate between similar and dissimilar nodes in a heterophilic neighborhood, and thereby improve local node-level representation learning. Secondly, we train individual expert graph models for each set of sub-graphs, specializing in survival prediction for targeted subsets of tissues. A simple linear aggregation of these experts is used for the final prediction. The process is analogous to real-world clinical practice, where distinct prognostic markers are individually estimated and afterward put together for collective evaluation. We empirically demonstrate that the individual expert models can potentially render comparable or better predictive performance than a model trained with complete-graphs. Furthermore, our framework allows better exploration of model-attended prognostic factors. The prior-guided tissue graphs provide a nuanced understanding of the biological relevance of the attended prognostic factors, making them more interpretable. Moreover, our method offers a flexible framework for incorporating additional pathologically relevant biomark-

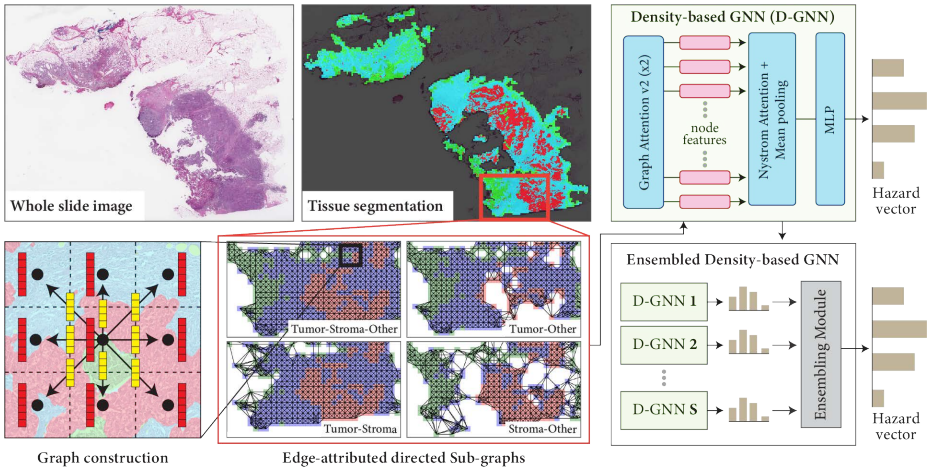

Fig. 1: (a) *Irrelevant* patches (gray) are removed, and *relevant* patches undergo TIL and tissue segmentation (tumor (red), tumor-associated stroma (blue), and others (green)). Edge-attributed directed graphs with node (red) and edge (yellow) features are constructed for tissue subsets. Density-based GNNs expert models are trained for each subgraph and they are ensemble for final prediction.

ers and testing their prognostic significance. We validate our approach on four cancer types from TCGA (BRCA, COAD&READ, STAD, and UCEC) [23], demonstrating consistent improvements in survival prediction over other baselines across all cancer types and achieving a gain of 4.33% on overall C-index.

2 Methods

In this section, we first discuss tissue-prior extraction from WSIs and propose our approach for constructing prior-guided edge-attributed directed subgraphs. Then, we introduce the individual expert Density-based Graph Neural Networks (D-GNNs) and conclude by defining ED-GNN, an ensemble of multiple expert D-GNNs. The overview of our method is illustrated in Fig. 1.

Segmentation prior: Inspired by the pathological process of tumor risk assessment, we adopt a two-staged coarse-to-fine-grained analytical approach. First, a ResNet18 binary classifier is trained on the WSIBulk dataset, a subset of the TIGER [1] dataset, which identifies patches within tumor bulk, *e.g.*, tumor, stroma, necrosis, and inflammations, from fat tissues and blood vessels, etc. Subsequently, tumor bulk patches undergo fine-grained simultaneous tissue and TIL segmentation using a bi-headed MAUnet [5] with a ResNet34 encoder, achieving pixel-wise segmentation at 0.5 $\mu m/pixel$. The first head segments three tissue types: tumor, tumor-associated stroma (TAS), and others, while the second head

segments TILs vs. no-TILs. Trained on the WSIROI dataset from the TIGER dataset, this architecture enables simultaneous tissue and TIL segmentation for faster processing. Model architecture is illustrated in the supplementary.

WSI Graph Construction: We overcome the challenge of convoluted heterogeneous WSI-graph by decomposing a WSI into tissue subsets $s \in S$, where S is the superset of tissue subsets and constructing edge-attributed directed graphs $G^s \; \forall s$, resulting in a set of graphs $\mathcal{G} = \{G^1, ..G^s.., G^{|S|}\}$. Given N tissue types, we can build $|S| = 2^N - 1$ subgraphs with different tissue combinations. However, we select only pathologically relevant tissue subsets based on literature studies, *i.e.*, *tumor-stroma*, *tumor-other*, *stroma-other*, *tumor-stroma-other*, as in Fig. 1. Next, we extract patches of size 224×224 with uniform stride at $1.0\,\mu m/pixel$. We compute the densities of tissue types per patch by utilizing the segmentation prior. Tissue subsets are created by selecting patches with $\geq 25\%$ tissue densities. Thresholding is decided over *argmax* to ensure sufficient per tissue-type representation while avoiding instances lacking certain tissue-types in WSIs.

We construct $G^s = (\mathbf{X}^s, \mathbf{A}^s, \mathbf{E}^s)$ at patient-level with n nodes and m edges, where $\mathbf{A}^s \in \mathbb{R}^{n \times n}$ is an unsymmetric adjacency matrix, and $\mathbf{X}^s \in \mathbb{R}^{n \times d_n}$, $\mathbf{E}^s \in \mathbb{R}^{m \times d_m}$ denote node and edge features, respectively. Similar to [6], $\mathbf{A}^s$ is built using k-NN(k=8) with a distance threshold resulting in a directed graph. Node features $\mathbf{x}_i^s \in \mathbf{X}^s$ is produced by passing patch i through a pre-trained pathology-specific feature extractor, CTranspath [24] that provides $d_n = 768$. We further concatenate Random Walk Positional Embeddings (RWPEs) [7, 18] to $\mathbf{X}^s$ to support transformer-based pooling in subsequent GNN. A RWPE is generated by taking random walks of different lengths and calculating the probability of landing back at the starting nodes. We use RWPE of size 24 based on [7] to ensure node-level spatial distinguishability, making $d_n = 792$. For edge features $\mathbf{e}_{ij}^s \in \mathbf{E}^s$, a density vector for node i, $\mathbf{d}_i^s \in \mathbb{R}^4$ is generated using the segmentation prior. $\mathbf{d}_i^s$ represents the densities of tumor, TAS, other, and TILs within patch i. The edge features $\mathbf{e}_{ij}$ between nodes i and j is defined as $\mathbf{e}_{ij}^s = [\mathbf{d}_i^s || \mathbf{d}_j^s]$, where $||$ denotes concatenation, making $d_m = 8$. Recognizing the significance of relationships between different entities [17], we posit that the proposed approach reinforces the GNN to consider various tissue relation types during message-passing and aggregation, effectively addressing the heterophilic tumor microenvironment.

Density based GNN (D-GNN): We propose to train D-GNNs that are individual expert models M^s for each subset s, resulting a set of models $\mathcal{M} = \{M^1, ..M^s, .., M^{|S|}\}$. M^s consists of L layers of Graph Attention Networks (GATv2) [4]. The message aggregation and update functions of a GATv2 layer are shown in equations 1 and 2. GATv2 learns the weight parameters $\mathbf{a}^s, \mathbf{\Theta}_u^s$, $\mathbf{\Theta}_v^s$, and $\mathbf{\Theta}_e^s$ to output α_{ij}^s (Eq. 2), the attention of i to j within neighborhood $\mathcal{N}(i)$. The anisotropic aggregation (Eq. 1) incorporates both neighbor's morphology x_j^s, and more importantly, the type of tissue relations and TIL densities e_{ij}^s defined from prior knowledge. This facilitates improved contextualized representation learn-

ing in a heterophillic tumor microenvironment. Further, the operation within a subset s enables more focused learning to utilize informative signals. Following, L layers of GNN, global WSI-level representations are produced by reading out the local contextualized node representations via self-attention-based pooling. This step exploits the long-distance interactions among tissue microenvironments to derive informative WSI representation. To this end, we use Nystromformer [26], due to its efficient computation of self-attention over long sequences, followed by *mean*-pooling. Finally, we use an MLP to generate the final hazard vector $p^s \in \mathbb{R}^t$, where the continuous patient survival time was divided into t time intervals [6]. Here, each logit $p^s[j]$ denotes the probability of a patient dying within the time interval t_j and t_{j+1}.

$$\mathbf{x}_i^{s(l+1)} = \alpha_{i,i}^s \mathbf{\Theta}_u^s \mathbf{x}_i^{s(l)} + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j}^s \mathbf{\Theta}_v^s \mathbf{x}_j^{s(l)} \tag{1}$$

$$\alpha_{i,j}^s = \frac{\exp\left(\mathbf{a}^{s\top} \text{ LeakyReLU}\left(\mathbf{\Theta}_u^s \mathbf{x}_i^s + \mathbf{\Theta}_v^s \mathbf{x}_j^s + \mathbf{\Theta}_e^s \mathbf{e}_{i,j}^s\right)\right)}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} \exp\left(\mathbf{a}^{s\top} \text{ LeakyReLU}\left(\mathbf{\Theta}_u^s \mathbf{x}_i^s + \mathbf{\Theta}_v^s \mathbf{x}_k^s + \mathbf{\Theta}_e^s \mathbf{e}_{i,k}^s]\right)\right)} \tag{2}$$

Ensemble of D-GNNs: Training independent M^s on G^s allows to specialize in specific tissue entities and interactions that denote different prognostic factors. We propose Ψ, an Ensemble of D-GNNs (ED-GNNs), which integrates the expertise of $\{M^s\}$, *i.e.*, $P = [p^1||..||p^s||..||p^{|S|}] \in \mathbb{R}^{|S| \times t}$, and predicts the hazard vector using $p = \Psi(P)$. Specifically, we define $\Psi(P) = W^T P$, where $W \in \mathbb{R}^{|S|}$ is a hyperparameter. Ensembling also helps in robustifying the survival prediction that builds on the synergies and complementarities of multiple predictors.

3 Experiments

Datasets: We assessed the efficacy of our method across four datasets from TCGA, namely Breast Invasive Carcinoma (BRCA, n=943, 129 events), Colon and Rectum Adenocarcinoma (COAD&READ, n=566, 118 events), Stomach Adenocarcinoma (STAD, n=332, 135 events), and Uterine Corpus Endometrial Carcinoma (UCEC, n=478, 75 events), encompassing a total of 2,319 patients, 2,655 slides, and 475 adverse events. We adopted the same 5-fold cross-validation splits as [6] for BRCA and UCEC. For STAD and COAD&READ, we stratified the dataset into 5-folds based on adverse events. The mean and standard deviation of the concordance index (C-index) across the 5-validation splits is used to benchmark the model performances.

For segmentation, we used the TIGER dataset [1] that includes breast cancer tissues consisting of WSIBulk (coarse tumor bulk annotation) and WSIROI (tissue and TILs annotations) sourced from multiple centers. To note, we transferred the segmentation prior extraction models trained on breast cancer to other cancer types without fine-tuning, due to the absence of tissue and TIL annotations.

Methods	BRCA	COAD&READ	STAD	UCEC	Overall
Attention MIL [11]	0.610 ± 0.053	0.571 ± 0.081	0.535 ± 0.065	0.629± 0.066	0.586
Patch-GCN [6]	0.611 ± 0.072	0.596 ± 0.066	0.538 ± 0.057	0.604 ± 0.129	0.587
TransMIL [20]	0.659± 0.071	0.533 ± 0.083	0.519 ± 0.061	0.660± 0.095	0.593
GTN [30]	0.651 ± 0.074	0.583 ± 0.032	0.542 ± 0.029	0.630 ± 0.052	0.601
HvTSurv [21]	0.636 ± 0.048	0.558 ± 0.046	0.558 ± 0.072	0.629 ± 0.094	0.595
Deep Attention MISL [27]	0.584 ± 0.045	0.528 ± 0.038	0.513 ± 0.075	0.618 ± 0.071	0.561
Tumor-Stroma D-GNN (Ours)	0.634 ± 0.076	0.560 ± 0.036	0.569 ± 0.060	0.630 ± 0.094	0.598
Tumor-Other D-GNN (Ours)	0.665 ± 0.071	0.558 ± 0.069	0.558 ± 0.051	0.655± 0.103	0.609
Stroma-Other D-GNN (Ours)	0.649 ± 0.038	0.600 ± 0.070	0.555 ± 0.093	0.675 ± 0.085	0.620
Complete-graph D-GNN (Ours)	0.669 ± 0.054	0.610± 0.047	0.551 ± 0.075	0.655 ± 0.074	0.621
ED-GNN (Ours)	0.672 ± 0.059	0.611 ± 0.040	0.562 ± 0.066	0.664± 0.076	0.627
Ablation studies for Complete-graph D-GNN					
Complete-graph D-GNN	**0.669 ± 0.054**	**0.610± 0.047**	0.551 ± 0.075	0.655 ± 0.074	**0.621**
✗ edge features	0.657 ± 0.055	0.565 ± 0.086	**0.565 ± 0.059**	**0.668 ± 0.116**	0.614
✗ edge features, coarse filtration	0.632 ± 0.037	0.572 ± 0.055	0.563 ± 0.050	0.666 ± 0.096	0.608
✗ edge features, coarse filtration, RWPE	0.629 ± 0.084	0.561 ± 0.051	0.543 ± 0.028	0.668± 0.105	0.600

Table 1: Survival prediction C-Index scores across 4 cancer types. Best model in green, second best in blue. Top baseline in yellow. Best ablation models in **bold**.

Implementation details: We used PyTorch Geometric (v2.3.1) and PyTorch (v1.13.0) for all our experiments. Negative log-likelihood survival loss [28] was optimized with identical hyperparameters as in [6], namely, Adam optimizer with 2×10^{-4} learning rate, 20 epochs, 1×10^{-5} weight decay, and batch size of 1 with 32 steps for gradient accumulations. For D-GNNs, we used $L = 2$. In ED-GNN, for subsets *tumor-stroma*, *tumor-other*, *stroma-other*, and *complete-graph*, we used weighted mean, with equal weightage to tissue sub-graphs and more weightage to the complete-graph. Same hyperparameter values were used for all cancer types and models from the last epochs were selected for evaluation.

4 Results and Discussions

Segmentation Prior: On the private TIGER testing dataset, our prior segmentation model achieved a dice score of 0.6927 for tumor and 0.7498 for TAS, compared to the challenge organizer's algorithm [2], that obtained 0.7152 for tumor and 0.6900 for TAS. For TIL identification, our model achieved a FROC score of 0.3203, compared to 0.2367 for the challenge organizer's algorithm. We qualitatively assessed segmentations on non-breast cancer datasets, demonstrating satisfactory results and enabling cross-application of the breast cancer dataset-trained prior, to other cancer types.

Quantitative analysis of D-GNNs & ED-GNN: We first assessed complete-graph D-GNN. Table 1 shows that D-GNN outperforms competing methods by 3.32% in overall metric while resulting in the best performance for BRCA and COAD&READ, and comparable for STAD and UCEC. Among graph-based methods, GTN and PatchGCN, D-GNN consistently showed improvement indicating its effectiveness in handling heterophilic tumor microenvironments. Ablation results in Table 1 affirm the effectiveness of our graph construction choices in terms of improved performance for BRCA, COAD&READ, and STAD over the

vanilla graph. Ablation studies showed that individual components impact differently across cancer types. While edge information was crucial for BRCA and COAD&READ; coarse filtration was important for BRCA, STAD and UCEC; and RWPE for all cancers except UCEC. These can be attributed to the variability in relevant prognostic factors across cancer types; for instance, UCEC prognosis relies more on tumor sizes and depth of tumor invasion than individual interactions [6], thus less dependent on edge information. Nevertheless, incorporating all components positively contributed to the overall performance.

In Table 1, we observe that, across all cancer datasets, ED-GNN outperforms competitors and complete-graph D-GNN by 4.33% and 1.13% in overall metric, respectively. Notably, stroma-other D-GNN results are comparable or better than complete-graph D-GNN, potentially reasoned to the subset graph focusing on prognostically relevant tumor-stroma boundaries [17]. In STAD and UCEC, experts focusing on different tissue subsets obtain the best model predictions, suggesting complementarity of prognostic factors across subsets. This indicates the utility of individual expert models that learn from relatively sparser tissue sub-graphs. It further reinforces our motivation to deconvolve complete graphs, enabling models to specialize in subset-specific prognostic factors and, through ensemble learning, eventually improving overall performance. While our current strategy involves independent expert modeling, they can be learned together in an end-to-end manner. Notably, our framework can aid in identifying biomarkers among exploratory markers by exploiting co-learning and knowledge discovery.

Qualitative analysis setup: To understand D-GNN's prognostic factors, we analyzed model-attended regions in segmentation-informed tissue graphs. We particularly analyzed the complete-graph to explore the prognostic factors across all the sets of segmented tissues and their interactions. We examined highly attended regions with attention weights $\geq 95^{\text{th}}$ percentile in TCGA-BRCA, categorizing nodes based on the tissue type with the highest density. TAS tissues were further subdivided into *high-TILs* and *low-TILs* using a predefined threshold derived from the 90^{th} percentile of TIL densities across all patches in the dataset. Based on model predictions, we stratified patients into high- and low-risk groups selecting 150 most/least risky patients based on model predictions. We tallied edges between various node types to investigate trends in edge connections within highly attended regions. A comparison was made with random graphs [16] having the same node types and degree distribution to mitigate bias in edge counts. For each patient and edge type, we checked if the graph had more or fewer edges than the average of edge connections among 50 randomly generated graphs. We assigned a value of 1 if it had more edges and 0 if it had fewer. The results of both analyses are presented as an average count across all patients in the respective groups.

Qualitative analysis: Fig. 2(a) reveals higher TILs in the low-risk group and more tumor nodes in the high-risk groups, consistent with high immune activity leads to better prognosis [19]. Fig. 2(b) shows a higher frequency of connections

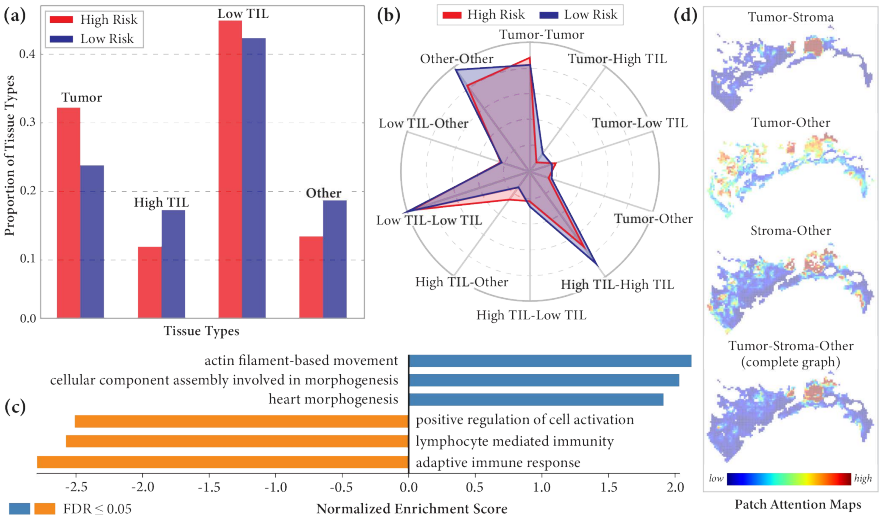

Fig. 2: (a) Proportion of tissue types and (b) tissue connection types (compared to random subgraphs), that complete-graph D-GNN paid high attention across high and low-risk groups. (c) Significant up/down-regulations of Gene Ontology (GO) terms across the groups. Orange bars show up-regulation of GO terms in low-risk groups compared to high-risk groups and vice versa for blue bars. For brevity, top 3 most up/down regulated GO terms were selected. (d) Attention maps of expert D-GNNs for different tissue subgraphs.

between similar entities than cross-entities, as similar entities often form cohesive tissue clusters, resulting in increased connections. Notably, connections between high TILs and tumor nodes were higher in low-risk compared to high-risk groups, suggesting the network emphasizes regions of interaction between high TILs and tumors, validating the importance of tumor-TIL interactions [19]. Additionally, Other and high TIL connections were higher than low-risk cases. As 'Other' includes diverse tissues, including high-risk elements like necrosis, further analysis is needed to determine the tissue type with such characteristics.

We wanted to further explore the differences between the two risk groups from the genomics context. We performed gene set enrichment analysis (GSEA) [22] on the high and low-risk groups, using patient's gene expressions [15], to identify significantly up or downregulated GO terms. Interestingly, GO terms significantly upregulated in the low-risk group primarily related to immune response Fig. 2(c). Additional GO terms are presented in the supplementary material. Moreover, we observe that our model captures significantly more immune-related GO terms than competing models, such as GTN (see supplementary), which also matches Fig. 2(a) and Fig. 2(b), where increased attention in low-risk groups is directed towards high TIL regions and tumor-high TIL connections.

In Fig. 2(d), attention maps for various tissue subset graphs are displayed. D-GNNs in addition to centering attention on the tumor, exhibit distinct focus

within different subset graphs, providing diverse prognostic factors. This allows ED-GNN to integrate diverse perspectives for enhancing survival prediction.

5 Conclusion

In summary, this paper presents an innovative approach to tackle the challenges in survival prediction arising from the heterogeneity of WSIs. By integrating segmentation priors, our method guides tissue graph construction and ensembled graph-model development. Our ensembled model achieves state-of-the-art performance by improving 4.33% overall C-Index over four cancer types, demonstrating our method's effectiveness in analyzing complex tumor microenvironments in WSIs. However, our method requires training individual expert models, which can be limiting when applied to large datasets. We aim to explore end-to-end expert aggregation pipelines for ED-GNN in the future. Future work will also involve more studies on the generalizability of the segmentation model to other cancer types and the effects of erroneous segmentation on survival prediction. We also plan to investigate modeling more prognostic markers across cancer types, which can be seamlessly integrated into our flexible framework.

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

References

1. Data - grand challenge. https://tiger.grand-challenge.org/Data/, (Accessed on 02/29/2024)
2. Diagnijmegen/pathology-tiger-baseline. https://github.com/DIAGNijmegen/pathology-tiger-baseline, (Accessed on 02/29/2024)
3. Beck, A.H., Sangoi, A.R., Leung, S., Marinelli, R.J., Nielsen, T.O., Van De Vijver, M.J., West, R.B., Van De Rijn, M., Koller, D.: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Science translational medicine **3**(108), 108ra113–108ra113 (2011)
4. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? arXiv preprint arXiv:2105.14491 (2021)
5. Cai, Y., Wang, Y.: Ma-unet: An improved version of unet based on multi-scale and attention mechanism for medical image segmentation. In: International Conference on Electronics and Communication; Network and Computer Technology. vol. 12167, pp. 205–211. SPIE (2022)
6. 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)
7. Dwivedi, V.P., Luu, A.T., Laurent, T., Bengio, Y., Bresson, X.: Graph neural networks with learnable structural and positional representations. arXiv preprint arXiv:2110.07875 (2021)
8. Fuchs, T.J., Buhmann, J.M.: Computational pathology: challenges and promises for tissue analysis. Computerized Medical Imaging and Graphics **35**(7-8), 515–530 (2011)
9. Hamilton, W.: Graph representation learning. Synthesis Lectures on Artificial Intelligence and Machine Learning **14**, 1–159 (09 2020)
10. 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)
11. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136 (2018)
12. Lee, Y., Park, J.H., Oh, S., Shin, K., Sun, J., Jung, M., Lee, C., Kim, H., Chung, J.H., Moon, K.C., et al.: Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. Nature Biomedical Engineering pp. 1–15 (2022)
13. 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)
14. Liu, P., Ji, L., Ye, F., Fu, B.: Graphlsurv: A scalable survival prediction network with adaptive and sparse structure learning for histopathological whole-slide images. Computer Methods and Programs in Biomedicine **231**, 107433 (2023)
15. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome biology **15**(12), 1–21 (2014)
16. Newman, M.E.: The structure and function of complex networks. SIAM review **45**(2), 167–256 (2003)
17. Pietras, K., Östman, A.: Hallmarks of cancer: interactions with the tumor stroma. Experimental cell research **316**(8), 1324–1331 (2010)
18. Rampášek, L., Galkin, M., Dwivedi, V.P., Luu, A.T., Wolf, G., Beaini, D.: Recipe for a general, powerful, scalable graph transformer. Advances in Neural Information Processing Systems **35**, 14501–14515 (2022)

19. Salgado, R., Denkert, C., Demaria, S., Sirtaine, N., Klauschen, F., Pruneri, G., Wienert, S., Van den Eynden, G., Baehner, F.L., Pénault-Llorca, F., et al.: The evaluation of tumor-infiltrating lymphocytes (tils) in breast cancer: recommendations by an international tils working group 2014. Annals of oncology **26**(2), 259–271 (2015)
20. 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)
21. 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: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2209–2217 (2023)
22. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences **102**(43), 15545–15550 (2005)
23. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: Review the cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary Oncology/Współczesna Onkologia **2015**(1), 68–77 (2015)
24. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Medical image analysis **81**, 102559 (2022)
25. Wang, Z., Li, J., Pan, Z., Li, W., Sisk, A., Ye, H., Speier, W., Arnold, C.W.: Hierarchical graph pathomic network for progression free survival prediction. In: MICCAI. pp. 227–237 (2021)
26. Xiong, Y., Zeng, Z., Chakraborty, R., Tan, M., Fung, G., Li, Y., Singh, V.: Nyströmformer: A nyström-based algorithm for approximating self-attention. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 14138–14148 (2021)
27. 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)
28. Zadeh, S.G., Schmid, M.: Bias in cross-entropy-based training of deep survival networks. IEEE transactions on pattern analysis and machine intelligence **43**(9), 3126–3137 (2020)
29. Zhao, L., Hou, R., Teng, H., Fu, X., Han, Y., Zhao, J.: Coads: Cross attention based dual-space graph network for survival prediction of lung cancer using whole slide images. Computer Methods and Programs in Biomedicine **236**, 107559 (2023)
30. Zheng, Y., Gindra, R.H., Green, E.J., Burks, E.J., Betke, M., Beane, J.E., Kolachalama, V.B.: A graph-transformer for whole slide image classification. IEEE transactions on medical imaging **41**(11), 3003–3015 (2022)
31. 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)