ALL-IN: A Local GLobal Graph-Based DIstillatioN Model for Representation Learning of Gigapixel Histopathology Images With Application In Cancer Risk Assessment

Puria Azadi[1], Jonathan Suderman[1], Ramin Nakhli[1], Katherine Rich[1], Maryam Asadi[1], Sonia Kung[2], Htoo Oo[2], Mira Keyes[1], Hossein Farahani[1], Calum MacAulay[3], Larry Goldenberg[2], Peter Black[2], and Ali Bashashati[1(✉)]

[1] University of British Columbia, Vancouver, BC, Canada
ali.bashashati@ubc.ca
[2] Vancouver Prostate Centre, Vancouver, BC, Canada
[3] BC Cancer Agency, Vancouver, BC, Canada

Abstract. The utility of machine learning models in histopathology image analysis for disease diagnosis has been extensively studied. However, efforts to stratify patient risk are relatively under-explored. While most current techniques utilize small fields of view (so-called local features) to link histopathology images to patient outcome, in this work we investigate the combination of global (i.e., contextual) and local features in a graph-based neural network for patient risk stratification. The proposed network not only combines both fine and coarse histological patterns but also utilizes their interactions for improved risk stratification. We compared the performance of our proposed model against the state-of-the-art (SOTA) techniques in histopathology risk stratification in two cancer datasets. Our results suggest that the proposed model is capable of stratifying patients into statistically significant risk groups ($p < 0.01$ across the two datasets) with clinical utility while competing models fail to achieve a statistical significance endpoint ($p = 0.148 - 0.494$).

Keywords: Histopathology · Risk Assessment · Graph Processing

1 Introduction

The examination of tissue and cells using microscope (referred to as histology) has been a key component of cancer diagnosis and prognostication since more than a hundred years ago. Histological features allow visual readout of cancer biology as they represent the overall impact of genetic changes on cells [20].

The great rise of deep learning in the past decade and our ability to digitize histopathology slides using high-throughput slide scanners have fueled interests in the applications of deep learning in histopathology image analysis. The

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43987-2_74.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 765–775, 2023.
https://doi.org/10.1007/978-3-031-43987-2_74

majority of the efforts, so far, focus on the deployment of these models for diagnosis and classification [27]. As such, there is a paucity of efforts that embark on utilizing machine learning models for patient prognostication and survival analysis (for example, predicting risk of cancer recurrence or expected patient survival). While prognostication and survival analysis offer invaluable insights for patient management, biological studies and drug development efforts, they require careful tracking of patients for a lengthy period of time; rendering this as a task that requires a significant amount of effort and funding.

In the machine learning domain, patient prognostication can be treated as a weakly supervised problem, which a model would predict the outcome (e.g., time to cancer recurrence) based on the histopathology images. Their majority have utilized Multiple Instance Learning (MIL) [8] that is a two-step learning method. First, representation maps for a set of patches (i.e., small fields of view), called a bag of instances, are extracted. Then, a second pooling model is applied to the feature maps for the final prediction. Different MIL variations have shown superior performances in grading or subtype classification in comparison to outcome prediction [10]. This is perhaps due to the fact that MIL-based technique do not incorporate patch locations and interactions as well as tissue heterogeneity which can potentially have a vital role in defining clinical outcomes [4,26].

To address this issue, graph neural networks (GNN) have recently received more attention in histology. They can model patch relations [17] by utilizing message passing mechanism via edges connecting the nodes (i.e., small patches in our case). However, most GNN-based models suffer from over smoothing [22] which limits nodes' receptive fields [3]. While local contexts mainly capture cell-cell interactions, global patterns such as immune cell infiltration patterns and tumor invasion in normal tissue structures (e.g., depth of invasion through myometrium in endometrial cancer [1]) could capture critical information about outcome [10]. Hence, locally focused methods are unable to benefit from the coarse properties of slides due to their high dimensions which may lead to poor performance.

This paper aims to investigate the potential of extracting fine and coarse features from histopathology slides and integrating them for risk stratification in cancer patients. Therefore, the contributions of this work can be summarized as: 1) a novel graph-based model for predicting survival that extracts both local and global properties by identifying morphological super-nodes; 2) introducing a fine-coarse feature distillation module with 3 various strategies to aggregate interactions at different scales; 3) outperforming SOTA approaches in both risk prediction and patient stratification scenarios on two datasets; 4) publishing two large and rare prostate cancer datasets containing more than 220 graphs for active surveillance and 240 graphs for brachytherapy cases. The code and graph embeddings are publicly available at https://github.com/pazadimo/ALL-IN

2 Related Works

2.1 Weakly Supervised Learning in Histopathology

Utilizing Weakly Supervised Learning for modeling histopathology problems has been getting popular due to the high resolution of slides and substantial time

and financial costs associated with annotating them as well as the development of powerful deep discriminative models in the recent years [24].

Such models are used to perform nuclei segmentation [18], identify novel subtypes [12], or later descendants are even able to pinpoint sub-areas with a high diagnostic value [19].

2.2 Survival Analysis and GNNs in Histopathology

MIL-based models have been utilized for outcome prediction [29,32] which can also be integrated with attention-based variants [14]. GNNs due to their structural preserving capacity [28] have drawn attention in various histology domains by constructing the graph on cells or patches. However, current GNN-based risk assessment variants are only focused on short-range interactions [16,17] or consider local contexts [10]. We hypothesize that graph-based models' performance in survival prediction improves by leveraging both fine and coarse properties.

3 Method

Figure 1 summarizes our proposed end-to-end solution. Below, we have provided details of each module.

3.1 Problem Formulation

For P_n, which is the n-th patient, a set of patches $\{patch_j\}_{j=1}^{M}$ is extracted from the related whole slide images. In addition, a latent vector $z_j \in R^{1\times d}$ is extracted from $patch_j$ using our encoder network (described in Sect. 3.2) that results in feature matrix $Z_n \in R^{M\times d}$ for P_n. Finally, a specific graph (G_n) for the n-th patient (P_n) can be constructed by assuming patches as nodes. Also, edges are connected based on the patches' k-nearest neighbour in the spatial domain resulting in an adjacency matrix A_n. Therefore, for each patient such as P_n, we have a graph defined by adjacency matrix A_n with size $M \times M$ and features matrix Z_n ($G_n = graph(Z_n, A_n)$). We estimate K super-nodes as matrix $S_n \in R^{K\times d}$ representing groups of local nodes with similar properties as coarse features for P_n's slides. The final model (ϵ_θ) with parameters θ utilizes G_n and S_n to predict the risk associated with this patient:

$$risk_n = \epsilon_\theta(G_n, S_n) = \epsilon_\theta(graph(X_n, A_n), S_n) \tag{1}$$

3.2 Self-supervised Encoder

Due to computational limits and large number of patches available for each patient, we utilize a self-supervised approach to train an encoder to reduce the inputs' feature space size. Therefore, We use DINO [9], a knowledge distillation model (KDM), with vision transformer (ViT) [13] as the backbone. It utilizes global and local augmentations of the input $patch_j$ and passes them to the student ($S_{\theta_1,ViT}$) and teacher ($T_{\theta_2,ViT}$) models to find their respective

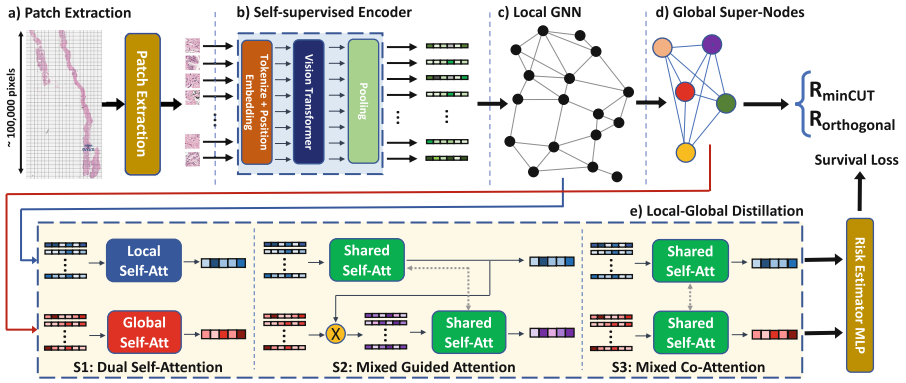

Fig. 1. The overview of our proposed method. a) The input slide is tiled into non-overlapping patches. b) The patches are fed into a self-supervised encoder to extract embeddings. c) A graph is constructed and the new local instance-level embeddings are obtained through the message-passing process. d) The global context representations in the form of super-nodes are extracted utilizing two unsupervised loss functions (R_{minCUT}, $R_{orthognal}$). e) The fine and coarse feature vectors are aggregated in the distillation module to obtain a representation that accounts for both local and global (contextual) histo-morphological features. Three different strategies (S1, S2, S3) are explored in this module. Finally, a Multilayer Perceptron (MLP) is deployed to estimate the risk using final resultant vectors.

representations without any labels. Then, by using distillation loss, it makes the representations' distribution similar to each other. Finally, the fixed weights of the teacher model are utilized in order to encode the input patches.

3.3 Local Graph Neural Network

GNN's objective is to find new nodes' embeddings via integrating local neighbors' interactions with individual properties of patches. By exploiting the message passing mechanism, this module iteratively aggregates features from neighbors of each vertex and generates the new node representations. We employ two graph convolution isomorphism operators (GINconv) [30] with the generalized form as:

$$X'_n = \phi\left(A_n + (1+\epsilon).I\right).X_n\right), \tag{2}$$

where ϵ is a small positive value and I is the identity matrix. Also, ϕ denotes the weights of two MLP layers. $X_n \in R^{M\times d}$ and $X'_n \in R^{M\times d}$ are GINconv's input and output feature matrices for P_n, which X_n equals Z_n for the first layer.

3.4 Super-Nodes Extractor

In order to find the coarse histo-morphological patterns disguised in the local graph, we propose extracting K Super-nodes, which each represents a weighted cluster of further processed local features. Intuitively, the number of super-nodes K should not be very large or small, as the former encourages them to only represent local clusters and the latter leads to larger clusters and loses subtle

details. We exploit the minCUT [5] idea to extract super-nodes in a differentiable process after an auxiliary GINconv to focus more on large-scale interactions and to finally learn the most global correlated super-nodes. Inspired by the relaxation form of the known K-way minCUT problem, we create a continuous cluster matrix $C_n \in R^{M\times K}$ using MLP layers and can finally estimate the super-nodes features ($S_n \in R^{M\times d}$) as:

$$S_n = C_n^T.X'_n, \quad C_n = softmax\left(ReLU(X'_n.W_1).W_2\right), \tag{3}$$

where W_1, W_2 are MLPs' weights. Hence, the extracted nodes are directly dependent on the final survival-specific loss. In addition, two additional unsupervised weighted regularization terms are optimized to improve the process:

MinCut Regularizer. This term is motivated by the original minCUT problem and intends to solve it for the the patients' graph. It is defined as:

$$R_{minCUT} = -\frac{Tr(C_n^T.A_{n,norm}.C_n)}{Tr(C_n^T.D_n.C_n)}, \tag{4}$$

where D_n is the diagonal degree matrix for A_n. Also, $Tr(.)$ represents the trace of matrix and $A_{n,norm}$ is the normalized adjacency matrix. R_{minCUT}'s minimum value happens when $Tr(C_n^T.A_{n,norm}.C_n)$ equals $Tr(C_n^T.D_{g,n}.C_n)$. Therefore, minimizing R_{minCUT} causes assigning strongly similar nodes to a same super-node and prevent their association with others.

Orthogonality Regularizer. R_{minCUT} is non-convex and potent to local minima such as assigning all vertexes to a super-node or having multiple super-nodes with only a single vertex. $R_{orthogonal}$ penalizes such solutions and helps the model to distribute the graph's features between super-nodes. It can be formulated as:

$$R_{orthogonal} = \left|\left| \frac{C_n^T.C_n}{||C_n^T.C_n||_F} - \frac{I}{\sqrt{K}} \right|\right|_F, \tag{5}$$

where $||.||_F$ is the Frobenius norm, and I is the identity matrix. This term pushes the model's parameters to find coarse features that are orthogonal to each other resulting in having the most useful global features.

Overall, utilizing these two terms encourages the model to extract super-nodes by leaning more towards the strongly associated vertexes and keeping them against weakly connected ones [5], while the main survival loss still controls the global extraction process.

3.5 Fine-Coarse Distillation

We propose our fine-coarse morphological feature distillation module to leverage all-scale interactions in the final prediction by finding a local and a global patient-level representations ($\hat{h}_{l,n}$, $\hat{h}_{g,n}$). Assume that $X'_n \in R^{M\times d}$ and $S_n \in R^{K\times d}$ are the feature matrices taken from local GNN (Sect. 3.3) and super-nodes for P_n, respectively. We explore 3 different attention-based feature distillation strategies for this task, including:

– **Dual Attention (DA):** Two gated self-attention modules for local and global properties with separate weights $(W_{\phi,l}, W_{\phi,g}, W_{k,l}, W_{k,g}, W_{q,l}, W_{q,g})$ are utilized to find patches scores $\alpha_l \in R^{1\times M}$ and $\alpha_g \in R^{1\times K}$ and the final features $(\hat{h}_{l,n}, \hat{h}_{g,n})$ as:

$$\hat{h}_{l,n} = \sum_{i=1}^{M} W_{\phi,l}\alpha_{l,i}x'_{n,i}, \quad \alpha_l = softmax\left[W_{v,l}\left(tanh(W_{q,l}X_n^{'T}) \cdot sigm(W_{k,l}X_n^{'T})\right)\right], \tag{6}$$

$$\hat{h}_{g,n} = \sum_{i=1}^{K} W_{\phi,g}\alpha_{g,i}s_{n,i}, \quad \alpha_g = softmax\left[W_{v,g}\left(tanh(W_{q,g}S_n^{T}) \cdot sigm(W_{k,g}S_n^{T})\right)\right], \tag{7}$$

where $x'_{n,i}$ and $s_{n,i}$ are rows of X'_n and S_n, respectively, and the final representation $(\hat{h})$ is generated as $\hat{h} = cat(\hat{h}_l, \hat{h}_g)$.

– **Mixed Guided Attention (MGA):** In the first strategy, the information flows from local and global features to the final representations in parallel without mixing any knowledge. The purpose of this policy is the heavy fusion of fine and coarse knowledge by exploiting shared weights ($W_{\phi,shared}$, $W_{k,shared}$, $W_{q,shared}$, $W_{v,shared}$) in both routes and benefiting from the guidance of local representation on learning the global one by modifying Eq. (7) to:

$$\alpha_g = softmax\left[W_{\phi,g}\left(tanh(W_{q,g}S_n^{T}\hat{h}_{l,n}) \cdot sigm(W_{k,g}S_n^{T}\hat{h}_{l,n})\right)\right] \tag{8}$$

– **Mixed Co-Attention (MCA):** While the first strategy allows the extreme separation of two paths, the second one has the highest level of mixing information. Here, we take a balanced policy between the independence and knowledge mixture of the two routes by only sharing the weights without using any guidance.

4 Experiments and Results

4.1 Dataset

We utilize two prostate cancer (PCa) datasets to evaluate the performance of our proposed model. The first set (PCa-AS) includes 179 PCa patients who were managed with Active Surveillance (AS). Radical therapy is considered overtreatment in these patients, so they are instead monitored with regular serum prostate-specific antigen (PSA) measurements, physical examinations, sequential biopsies, and magnetic resonance imaging [23]. However, AS may be over- or under-utilized in low- and intermediate-risk PCa due to the uncertainty of current methods to distinguish indolent from aggressive cancers [11]. Although majority of patients in our cohort are classified as low-risk based on NCCN guidelines [21], a significant subset of them experienced disease upgrade that triggered definitive therapy (range: 6.2 to 224 months after diagnosis).

The second dataset (PCa-BT) includes 105 PCa patients with low to high risk disease who went through brachytherapy. This treatment involves placing a radioactive material inside the body to safely deliver larger dose of radiation at

Table 1. Comparison of our method against baselines and ablation study on policies.

Model	c-index ↑		p-value ↓		High ↓ - Low ↑ Median Time		Parameters
	PCa-AS	PCa-BT	PCa-AS	PCa-BT	PCa-AS	PCa-BT	PCa-AS
DeepSet	0.495 ± 0.017	0.50 ± 0.0	0.837	0.912	67.78–71.87	24.62–24.89	329K
AMIL	0.544 ± 0.06	0.533 ± 0.060	0.820	0.148	48.99–89.10	21.86–30.71	592K
DGC	0.522 ± 0.113	0.572 ± 0.150	0.494	0.223	47.61–96.66	23.44–24.85	626K
Patch-GCN	0.555 ± 0.059	0.541 ± 0.118	0.630	0.981	37.72–94.95	23.05–25.25	1,302K
ALL-IN + DA (ours)	0.631 ± 0.058	0.596 ± 0.062	< 0.01	< 0.01	37.72–115.91	21.86–35.77	850K
ALL-IN + MGA (ours)	0.632 ± 0.060	0.589 ± 0.074	< 0.01	< 0.01	47.61–101.39	21.86–35.77	653K
ALL-IN + MCA (ours)	$\mathbf{0.639 \pm 0.048}$	$\mathbf{0.600 \pm 0.077}$	$\mathbf{< 0.01}$	$\mathbf{< 0.01}$	**36.5–131.71**	**21.86–35.77**	653K

one time [25]. The recorded endpoint for this set is biochemical recurrence with time to recurrence ranging from 11.7 to 56.1 months.

We also utilized the Prostate cANcer graDe Assessment (PANDA) Challenge dataset [7] that includes more than 10,000 PCa needle biopsy slides (no outcome data) as an external dataset for training the encoder of our model.

4.2 Experiments

We evaluate the models' performance in two scenarios utilizing several objective metrics. Implementation details are available in supplementary material.

Hazard (Risk) Prediction. We utilize concordance-index (c-index) that measures the relative ordering of patients with observed events and un-censored cases relative to censored instances [2]. Using c-index, we compare the quality of hazard ranking against multiple methods including two MIL (DeepSet [31], AMIL [14]) and graph-based (DGC [17] and Patch-GCN [10]) models that were utilized recently for histopathology risk assessment. C-index values are available in Table 1. The proposed model with all strategies outperforms baselines across all sets and is able to achieve 0.639 and 0.600 on PCa-AS and PCa-BT, while the baselines, at best, obtain 0.555, and 0.572, respectively. Statistical tests (paired t-test) on c-indices also show that our model is statistically better than all baselines in PCa-AS and also superior to all models, except DGC, in PCa-BT. Superior performance of our MCA policy implies that balanced exploitation of fine and coarse features with shared weights may provide more robust contextual information compared to using mixed guided information or utilizing them independently.

Patient Stratification. The capacity of stratifying patients into risk groups (e.g., low and high risk) is another criterion that we employ to assess the utility of models in clinical practice. We evaluate model performances via Kaplan-Meier curve [15] (cut-off set as the ratio of patients with recurrence within 3

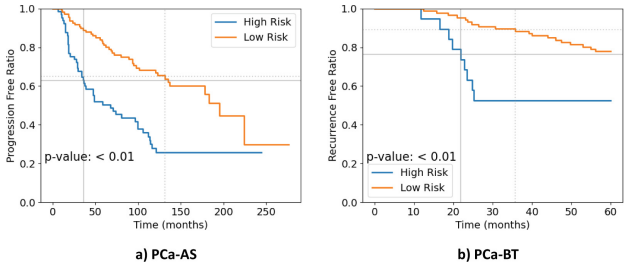

Fig. 2. Kaplan-Meier curves of mixed co-attention model for PCa-AS and PCa-BT.

years of therapy initiation for PCa-BT and the ratio of upgraded cases for PCa-AS), LogRank test [6] (with 0.05 as significance level), and median outcome associated with risk groups (Table 1 and Fig. 2). Our model stratified PCa-AS patients into high- and low-risk groups with median time to progression of 36.5 and 131.7 months, respectively. Moreover, PCa-BT cases assigned to high- and low-risk groups have median recurrence time of 21.86 and 35.7 months. While none of the baselines are capable of assigning patients into risk groups with statistical significance, our distillation policies achieve significant separation in both PCa-AS and PCa-BT datasets; suggesting that global histo-morphological properties improve patient stratification performance. Furthermore, our findings have significant clinical implications as they identify, for the first time, high-risk prostate cancer patients who are otherwise known to be low-risk based on clinico-pathological parameters. This group should be managed differently from the rest of the low-risk prostate cancer patients in the clinic. Therefore, providing evidence of the predictive (as opposed to prognostic) clinical information that our model provides. While a prognostic biomarker provides information about a patient's outcome (without specific recommendation on the next course of action), a predictive biomarker gives insights about the effect of a therapeutic intervention and potential actions that can be taken.

Ablation Study. We perform ablation study (Table 2) on various components of our framework including local nodes, self-supervised ViT-based encoder, and most importantly, super-nodes in addition to fine-coarse distillation module. Although our local-only model is still showing superior results compared to baselines, this analysis demonstrates that all modules are essential for learning the most effective representations. We also assess the impact of our ViT on the baselines (full-results in appendix), showing that it can, on average, improve their performance by an increase of ~ 0.03 in c-index for PCa-AS. However, the best baseline with ViT still has poorer performance compared to our model in both datasets, while the number of parameters (reported for ViT embeddings' size in Table 1) in our full-model is about half of this baseline. Achieving higher c-indices in our all model versions indicates the important role of coarse features and global context in patient risk estimation in addition to local patterns.

Table 2. Ablation study on different modules.

Modules				c-index ↑	
Model	Local-node	our KDM-ViT	Super-node + Distillation Model	PCa-AS	PCa-BT
Patch-GCN	✓	✓	✗	0.627 ± 0.046	0.588 ± 0.067
Ours	✓	✗	✗	0.584 ± 0.072	0.550 ± 0.109
	✓	✓	✗	0.622 ± 0.055	0.597 ± 0.045
	✓	✓	✓	$\mathbf{0.639 \pm 0.048}$	$\mathbf{0.600 \pm 0.077}$

5 Conclusion

While risk assessment is relatively under-explored, most existing methods are focused only on small fields of view. In this work, we introduce a novel graph-based model for integrating global and local features, which utilizes interactions at a larger scale for improved risk stratification. Using two cancer datasets, we evaluated the effectiveness of our model against the baseline methods for hazard prediction and patients stratification. Our results suggest that the proposed model outperforms them in risk assessment and is capable of separating patients into statistically significant risk groups with actionable clinical utility. The full capacity of this work can be revealed by extending it to other histology tasks.

Acknowledgment:. This work was supported by a Canadian Institutes of Health Research grant to AB, PB, and LG and Michael Smith Health Research BC Scholar grant to AB.

References

1. Abu-Rustum, N.R., et al.: The revised 2009 figo staging system for endometrial cancer: should the 1988 figo stages ia and ib be altered? Int. J. Gynecol. Cancer **21**(3) (2011)
2. Alabdallah, A., Ohlsson, M., Pashami, S., Rögnvaldsson, T.: The concordance index decomposition-a measure for a deeper understanding of survival prediction models. arXiv preprint arXiv:2203.00144 (2022)
3. Alon, U., Yahav, E.: On the bottleneck of graph neural networks and its practical implications. arXiv preprint arXiv:2006.05205 (2020)
4. Angell, H., Galon, J.: From the immune contexture to the immunoscore: the role of prognostic and predictive immune markers in cancer. Curr. Opin. Immunol. **25**(2), 261–267 (2013)
5. Bianchi, F.M., Grattarola, D., Alippi, C.: Spectral clustering with graph neural networks for graph pooling. In: International Conference on Machine Learning, pp. 874–883. PMLR (2020)
6. Bland, J.M., Altman, D.G.: The logrank test. BMJ **328**(7447), 1073 (2004)
7. Bulten, W., et al.: Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. Nat. Med. **28**(1), 154–163 (2022)

8. Carbonneau, M.A., Cheplygina, V., Granger, E., Gagnon, G.: Multiple instance learning: a survey of problem characteristics and applications. Pattern Recogn. **77**, 329–353 (2018)
9. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9650–9660 (2021)
10. Chen, R.J., et al.: Whole slide images are 2D point clouds: context-aware survival prediction using patch-based graph convolutional networks. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12908, pp. 339–349. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_33
11. Cooperberg, M.R., et al.: Outcomes of active surveillance for men with intermediate-risk prostate cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. **29**(2), 228–234 (2011)
12. Darbandsari, A., et al.: Identification of a novel subtype of endometrial cancer with unfavorable outcome using artificial intelligence-based histopathology image analysis (2022)
13. Dosovitskiy, A., et al.: An image is worth 16×16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
14. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136. PMLR (2018)
15. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**(282), 457–481 (1958)
16. Lee, Y., et al.: Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. Nat. Biomed. Eng., 1–15 (2022)
17. Li, R., Yao, J., Zhu, X., Li, Y., Huang, J.: Graph CNN for survival analysis on whole slide pathological images. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 174–182. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_20
18. Liu, W., He, Q., He, X.: Weakly supervised nuclei segmentation via instance learning. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–5. IEEE (2022)
19. 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. Nat. Biomed. Eng. **5**(6), 555–570 (2021)
20. Mobadersany, P., et al.: Predicting cancer outcomes from histology and genomics using convolutional networks. Proc. Natl. Acad. Sci. **115**(13), E2970–E2979 (2018)
21. Moses, K.A., et al.: Nccn guidelines® insights: prostate cancer early detection, version 1.2023: featured updates to the nccn guidelines. J. Natl. Comprehens. Cancer Netw. **21**(3), 236–246 (2023)
22. Oono, K., Suzuki, T.: Graph neural networks exponentially lose expressive power for node classification. arXiv preprint arXiv:1905.10947 (2019)
23. Ouzzane, A., et al.: Magnetic resonance imaging targeted biopsy improves selection of patients considered for active surveillance for clinically low risk prostate cancer based on systematic biopsies. J. Urol. **194**(2), 350–356 (2015)
24. Rony, J., Belharbi, S., Dolz, J., Ayed, I.B., McCaffrey, L., Granger, E.: Deep weakly-supervised learning methods for classification and localization in histology images: a survey. arXiv preprint arXiv:1909.03354 (2019)
25. Skowronek, J.: Current status of brachytherapy in cancer treatment-short overview. J. Contemp. Brachyther. **9**(6), 581–589 (2017)

26. Son, B., Lee, S., Youn, H., Kim, E., Kim, W., Youn, B.: The role of tumor microenvironment in therapeutic resistance. Oncotarget **8**(3), 3933 (2017)
27. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: a survey. Med. Image Anal. **67**, 101813 (2021)
28. Tang, S., Chen, D., Bai, L., Liu, K., Ge, Y., Ouyang, W.: Mutual crf-gnn for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2329–2339 (2021)
29. Wetstein, S.C., et al.: Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. Sci. Rep. **12**(1), 1–12 (2022)
30. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? arXiv preprint arXiv:1810.00826 (2018)
31. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. Adv. Neural Inf. Process. Syst. **30**, 1–11 (2017)
32. 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)