Multi-scope Analysis Driven Hierarchical Graph Transformer for Whole Slide Image Based Cancer Survival Prediction

Wentai Hou[1], Yan He[2], Bingjian Yao[2], Lequan Yu[3], Rongshan Yu[2], Feng Gao[4], and Liansheng Wang[2](✉)

[1] Department of Information and Communication Engineering at School of Informatics, Xiamen University, Xiamen, China
`houwt@stu.xmu.edu.cn`

[2] Department of Computer Science at School of Informatics, Xiamen University, Xiamen, China
`{yanhe56,yaobingjian}@stu.xmu.edu.cn`, `{rsyu,lswang}@xmu.edu.cn`

[3] Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong SAR, China
`lqyu@hku.hk`

[4] The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China
`gaof57@mail.sysu.edu.cn`

Abstract. Cancer survival prediction requires considering not only the biological morphology but also the contextual interactions of tumor and surrounding tissues. The major limitation of previous learning frameworks for whole slide image (WSI) based survival prediction is that the contextual interactions of pathological components (*e.g.*, tumor, stroma, lymphocyte, *etc.*) lack sufficient representation and quantification. In this paper, we proposed a multi-scope analysis driven Hierarchical Graph Transformer (HGT) to overcome this limitation. Specifically, we first utilize a multi-scope analysis strategy, which leverages an in-slide superpixel and a cross-slide clustering, to mine the spatial and semantic priors of WSIs. Furthermore, based on the extracted spatial prior, a hierarchical graph convolutional network is proposed to progressively learn the topological features of the variant microenvironments ranging from patch-level to tissue-level. In addition, guided by the identified semantic prior, tissue-level features are further aggregated to represent the meaningful pathological components, whose contextual interactions are established and quantified by the designed Transformer-based prediction head. We evaluated the proposed framework on our collected Colorectal Cancer (CRC) cohort and two public cancer cohorts from the TCGA project, *i.e.*, Liver Hepatocellular Carcinoma (LIHC) and Kidney Clear Cell Carcinoma (KIRC). Experimental results demonstrate that our proposed method yields superior performance and richer interpretability compared to the state-of-the-art approaches.

Keywords: Whole slide image · Survival prediction · Contextual interaction · Graph neural network · Transformer

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 745–754, 2023.
https://doi.org/10.1007/978-3-031-43987-2_72

1 Introduction

The ability to predict the future risk of patients with cancer can significantly assist clinical management decisions, such as treatment and monitoring [21]. Generally, pathologists need to manually assess the pathological images obtained by whole-slide scanning systems for clinical decision-making, *e.g.*, cancer diagnosis and prognosis [20]. However, due to the complex morphology and structure of human tissues and the continuum of histologic features phenotyped across the diagnostic spectrum, it is a tedious and time-consuming task to manually assess the whole slide image (WSI) [12]. Moreover, unlike cancer diagnosis and subtyping tasks, survival prediction is a future state prediction task with higher difficulty. Therefore, automated WSI analysis method for survival prediction task is highly demanded yet challenging in clinical practice.

Over the years, deep learning has greatly promoted the development of computational pathology, including WSI analysis [9,17,24]. Due to the huge size, WSIs are generally cropped to numerous patches with a fixed size and encoded to patch features by a CNN encoder (*e.g.*, Imagenet pretrained ResNet50 [11]) for further analysis. The attention-dominated learning frameworks (*e.g.*, ABMIL [13], CLAM [18], DSMIL [16], TransMIL [19], SCL-WC [25], HIPT [4], NAGCN [9]) mainly aim to find the key instances (*e.g.*, patches and tissues) for WSI representation and decision-making, which prefers the needle-in-a-haystack tasks, *e.g.*, cancer diagnosis, cancer subtyping, *etc.* To handle cancer survival prediction, some researchers integrated some attribute priors into the network design [5,26]. For example, Patch-GCN [5] treated the WSI as point cloud data, and the patch-level adjacent relationship of WSI is learned by a graph convolutional network (GCN). However, the fixed-size patches cropped from WSI mainly contain single-level biological entities (*e.g.*, cells), resulting in limited structural information. DeepAttnMISL [26] extracted the phenotype patterns of the patient via a clustering algorithm, which provides meaningful medical prior to guide the aggregation of patch features. However, this cluster analysis strategy only focuses on a single sample, which cannot describe the whole picture of the pathological components specific to the cancer type. Additionally, existing learning frameworks often ignore the capture of contextual interactions of pathological components (*e.g.*, tumor, stroma, lymphocyte, *etc.*), which is considered as important evidence for cancer survival prediction tasks [1,6]. Therefore, WSI-based cancer survival prediction still remains a challenging task.

In summary, to better capture the prognosis-related information in WSI, two technical key points should be fully investigated: (1) an analysis strategy to mine more comprehensive and in-depth prior of WSIs, and (2) a promising learning network to explore the contextual interactions of pathological components. To this end, this paper presents a novel multi-scope analysis driven learning framework, called Hierarchical Graph Transformer (HGT), to pertinently resolve the above technical key points for more reliable and interpretable

W. Hou and Y. He—Contributed equally to this work.

WSI-based survival prediction. First, to mine more comprehensive and in-depth attribute priors of WSI, we propose a multi-scope analysis strategy consisting of in-slide superpixels and cross-slide clustering, which can not only extract the spatial prior but also identify the semantic prior of WSIs. Second, to explore the contextual interactions of pathological components, we design a novel learning network, *i.e.*, HGT, which consists of a hierarchical graph convolution layer and a Transformer-based prediction head. Specifically, based on the extracted spatial topology, the hierarchical graph convolution layer in HGT progressively aggregate the patch-level features to the tissue-level features, so as to learn the topological features of variant microenvironments ranging from fine-grained (*e.g.*, cell) to coarse-grained (*e.g.*, necrosis, epithelium, *etc.*). Then, under the guidance of the identified semantic prior, the tissue-level features are further sorted and assigned to form the feature embedding of pathological components. Furthermore, the contextual interactions of pathological components are captured with the Transformer-based prediction head, leading to reliable survival prediction and richer interpretability. Extensive experiments on three cancer cohorts (*i.e.*, CRC, TCGA-LIHC and TCGA-KIRC) demonstrates the effectiveness and interpretability of our framework. Our codes are available at https://github.com/Baeksweety/superpixel_transformer.

2 Methodology

Figure 1 illustrates the pipeline of the proposed framework. Due to the huge size, WSIs are generally cropped to numerous patches with a fixed size (*i.e.*, 256×256) and encoded to patch features $V_{patch} \in \mathbb{R}^{n \times d}$ in the embedding space D by a CNN encoder (*i.e.*, ImageNet pretrained ResNet50 [11]) for further analysis, where n is the number of patches, $d = 1024$ is the feature dimension. The goal of WSI-based cancer survival prediction is to learn the feature embedding of V in a supervised manner and output the survival risk $O \in \mathbb{R}^1$.

However, conventional patch-level analysis cannot model complex pathological patterns (*e.g.*, tumor lymphocyte infiltration, immune cell composition, *etc.*), resulting in limited cancer survival prediction performance. To this end, we proposed a novel learning network, *i.e.*, HGT, which utilized the spatial and semantic priors mined by a multi-scope analysis strategy (*i.e.*, in-slide superpixel and cross-slide clustering) to represent and capture the contextual interaction of pathological components. Our framework consists two modules: a hierarchical graph convolutional network and a Transformer-based prediction head.

2.1 Hierarchical Graph Convolutional Network

Unlike cancer diagnosis and subtyping, cancer survival prediction is a quite more challenging task, as it is a future event prediction task which needs to consider complex pathological structures [20]. However, the conventional patch-level analysis is difficult to meet this requirement. Therefore, it is essential to extract and combine higher-level topology information for better WSI representation.

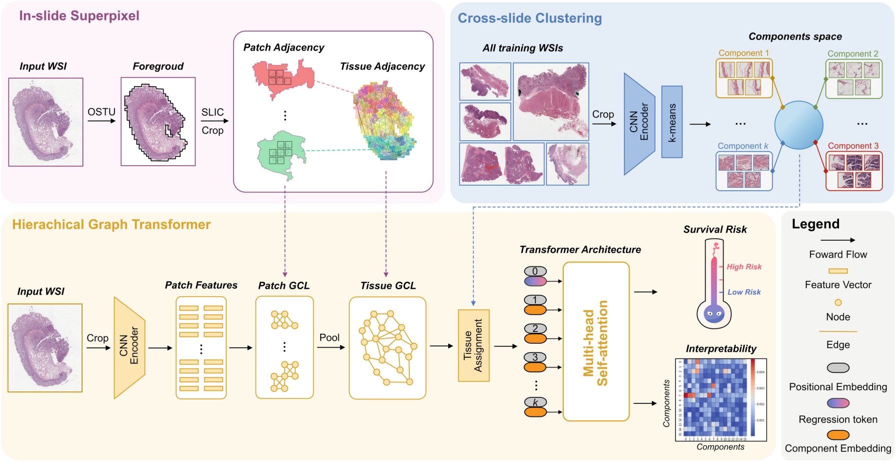

Fig. 1. Overview of the proposed multi-scope analysis (*i.e.*, in-slide superpixel and cross-slide clustering) driven Hierarchical Graph Transformer (HGT). Note that not all nodes and adjacent relationships are shown for visual clarity.

In-slide Superpixel. As shown in Fig .1, we first employ a Simple Linear Iterative Clustering (SLIC) [2] algorithm to detect non-overlapping homogeneous tissues of the foreground of WSI at a low magnification, which can be served as the spatial prior to mine the hierarchical topology of WSI. Intuitively, the cropped patches and segmented tissues in a WSI can be considered as hierarchical entities ranging from fine-grained level (*e.g.*, cell) to coarse-grained level (*e.g.*, necrosis, epithelium, *etc.*). Based on the in-slide superpixel, the tissue adjacency matrix $E_{tissue} \in \mathbb{R}^{m \times m}$ can be obtained, where m denote the number of superpixels. Then, the patches in each superpixel are further connected in an 8-adjacent manner, thus generating patch adjacency matrix $E_{patch} \in \mathbb{R}^{n \times n}$. The spatial assignment matrix between cropped patches and segmented tissues is denoted as $A_{spa} \in \mathbb{R}^{n \times m}$.

Patch Graph Convolutional Layer. Based on the spatial topology extracted by in-slide superpixel, the patch graph convolutional layer (Patch GCL) is designed to learn the feature of the fine-grained microenvironment (*e.g.*, cell) through the message passing between adjacent patches, which can be represented as:

$$V_{patch}^{'} = \sigma(\mathrm{GraphConv}(V_{patch}, E_{patch})), \tag{1}$$

where $\sigma(\cdot)$ denotes the activation function, such as ReLU. GraphConv denotes the graph convolutional operation, *e.g.*, GCNConv [15], GraphSAGE [10], *etc.*

Tissue Graph Convolutional Layer. Third, based on the spatial assignment matrix A_{spa}, the learned patch-level features can be aggregated to the tissue-level features which contain the information of necrosis, epithelium, *etc.*

$$V_{tissue} = [A_{spa}]^{\mathrm{T}} V_{patch}^{'}, \tag{2}$$

where $[\cdot]^{\mathrm{T}}$ denote the matrix transpose operation. The tissue graph convolutional layer (Tissue GCL) is further designed to learn the feature of this coarse-grained microenvironment, which can be represented as:

$$V_{tissue}^{'} = \sigma(\mathrm{GraphConv}(V_{tissue}, E_{tissue})). \tag{3}$$

2.2 Transformer-Based Prediction Head

Clinical studies have shown that cancer survival prediction requires considering not only the biological morphology but also the contextual interactions of tumor and surrounding tissues [1]. However, existing analysis frameworks for WSI often ignore the capture of contextual interactions of pathological components (*e.g.*, tumor, stroma, lymphocyte, *etc.*), resulting in limited performance and interpretability. Therefore, it is necessary to determine the feature embedding of pathological components and investigate their contextual interactions for more reliable predictions.

Cross-Slide Clustering. As shown in Fig. 1, we perform the k-means algorithm on the encoded patch features of all training WSIs to generate k pathological components $P \in \mathbb{R}^{k \times d}$ in the embedding space D. P represents different pathological properties specific to the cancer type. Formally, the feature embedding of each tissue in space D is defined as the mean feature embeddings of the patches within the tissue. And then, the pathological component label of each tissue is determined as the component closest to the Euclidean distance of the tissue in space D. The semantic assignment matrix between segmented tissues and pathological components is denoted as $A_{sem} \in \mathbb{R}^{m \times k}$.

Transformer Architecture. Under the guidance of the semantic prior identified by cross-slide clustering, the learned tissue features $V_{tissue}^{'}$ can be further aggregated, forming a series meaningful component embeddings P' specific to the cancer type.

$$P' = [A_{sem}]^{\mathrm{T}} V_{tissue}^{'}. \tag{4}$$

Then we employed a Transformer [22] architecture to mine the contextual interactions of P' and output the predicted survival risk. As shown in Fig. 1, P' is concatenated with an extra learnable regression token R and attached with positional embeddings E_{Pos}, which are processed by:

$$P'_{out} = \mathrm{MLP}(\mathrm{LN}(\mathrm{MHSA}(\mathrm{LN}([R; P'] + E_{Pos})))). \tag{5}$$

where P'_{out} is the output of Transformer, MHSA is the Multi-Headed Self-Attention [22], LN is Layer Normalization and MLP is Multilayer Perceptron. Finally, the representation of the regression token at the output layer of the Transformer, *i.e.*, $[P'_{out}]^0$, is served as the predicted survival risk O.

Loss Function and Training Strategy. For the network training, Cox loss [26] is adopted for the survival prediction task, which is defined as:

$$\mathcal{L}_{Cox} = \sum_{i=1}^{B} \delta_i \left(-O(i) + \log \sum_{j:t_j>=t_i} \exp\left(O(j)\right) \right), \tag{6}$$

where δ_i denote the censorship of i-th patient, $O(i)$ and $O(j)$ denote the survival output of i-th and j-th patient in a batch, respectively.

3 Experiments

3.1 Experimental Settings

Dataset. In this study, we used a **Colorectal Cancer (CRC)(385 cases)** cohort collected from co-operated hospital to evaluate the proposed method. Moreover, two public cancer cohorts from TCGA project, *i.e.*, **Liver Hepatocellular Carcinoma (LIHC)(371 cases)** and **Kidney Clear Cell Carcinoma (KIRC)(398 cases)** are also included as the censorship of these two data sets are relatively balanced. All WSIs are analyzed at ×10 magnification and cropped into 256 × 256 patches. The average patch number of each WSI is 18727, 3680, 3742 for CRC, TCGA-LIHC, TCGA-KIRC, respectively. It should be noted that the largest WSI (from CRC) contains 117568 patches.

Implementation Details. All trials are conducted on a workstation with two Intel Xeon Silver 4210R CPUs and four NVIDIA GeForce RTX 3090 (24 GB) GPUs. Our graph convolutional model is implemented by Pytorch Geometric [7]. The initial number of superpixels of SLIC algorithm is set to {600, 700, 600}, and the number of clusters of k-means algorithm is set to {16, 16, 16} for CRC, TCGA-LIHC and TCGA-KICA cohorts. The non linearity of GCN is ReLU. The number of Transformer heads is 8, and the attention scores of all heads are averaged to produce the heatmap of contextual interactions. HGT is trained with a mini-batch size of 16, and a learning rate of $1e-5$ with Adam optimizer for 30 epochs.

Evaluation Metric. The concordance index (CI) [23] is used to measure the fraction of all pairs of patients whose survival risks are correctly ordered. CI ranges from 0 to 1, where a larger CI indicates better performance. Moreover, to evaluate the ability of patients stratification, the Kaplan-Meier (KM) analysis is used [23]. In this study, we conduct a 5-fold evaluation procedure with 5 runs to evaluate the survival prediction performance for each method. The result of $mean \pm std$ is reported.

3.2 Comparative Results

We compared seven state-of-the-art methods (SOTAs), *i.e.*, DeepSets [27], ABMIL [13], DeepAttnMISL [26], CLAM [18], DSMIL [16], PatchGCN [5], and TransMIL [19]. We also compared three baselines of our method, *i.e.*, w/o Patch GCL, w/o Tissue GCL and w/o Transformer. For fair comparison, same CNN extractor (*i.e.* ImageNet pretrained Resnet50 [11]), and survival prediction loss (*i.e.* Cox loss [26]) is adopt for all methods.

Table 1 and Fig. 2 show the results of CI and KM-analysis of each method, respectively. Generally, most MIL methods, *i.e.*, DeepSets, ABMIL, DSMIL, TransMIL mainly focus on a few key instances for prediction, but they do not have significant advantages in cancer prognosis. Furthermore, due to the large size of CRC dataset and relatively high model complexity, Patch-GCN and TransMIL encountered a memory overflow when processing the CRC dataset, which limits their clinical application. DeepAttnMISL has a certain semantic perception ability for patch, which achieves better performance in LIHC cohort. PatchGCN is capable to capture the local contextual interactions between patch, which also achieves satisfied performance in KIRC cohort. As our method has potential to explore the contextual interactions of pathological components, which more in line with the thinking of pathologists for cancer prognosis. Our method achieves higher CI and relatively low P-Value (< 0.05) of KM analysis on both three cancer cohorts, which consistently outperform the SOTAs and baselines. In addition, the feature aggregation of the lower levels (*i.e.*, patch and tissue) are guided by the priors, and the MHSA is only executed on pathological components, resulting in high efficiency even on the CRC dataset.

Table 1. Experimental results of CI. Results not significantly worse than the best (P-Value $>$ 0.05, Two-sample t-test) are shown in bold. The second best results of SOTA methods are underlined. "-" denotes that the algorithm cannot be executed in this cohort due to memory overflow.

Type	Method	CRC	TCGA-LIHC	TCGA-KIRC
SOTAs	DeepSets [27]	0.504 ± 0.004	0.511 ± 0.011	0.483 ± 0.033
	ABMIL [13]	0.580 ± 0.005	0.634 ± 0.005	0.617 ± 0.094
	DeepAttnMISL [26]	0.570 ± 0.001	0.644 ± 0.009	0.584 ± 0.019
	CLAM [18]	0.575 ± 0.010	0.641 ± 0.002	0.635 ± 0.006
	DSMIL [16]	0.550 ± 0.016	0.626 ± 0.005	0.603 ± 0.022
	PatchGCN [5]	-	0.643 ± 0.003	**0.637**±**0.010**
	TransMIL [19]	-	0.641 ± 0.023	0.616 ± 0.014
Ablation study	w/o Patch GCL	0.597 ± 0.007	0.640 ± 0.002	0.626 ± 0.008
	w/o Tissue GCL	0.584 ± 0.010	0.644 ± 0.002	0.636 ± 0.008
	w/o Transformer	0.592 ± 0.010	0.647 ± 0.002	0.616 ± 0.005
Ours	HGT	**0.607** ± **0.004**	**0.657** ± **0.003**	**0.646** ± **0.003**

3.3 Interpretability of the Proposed Framework

We selected the CRC dataset for further interpretable analysis, as it is one of the leading causes of mortality in industrialized countries, and its prognosis-related factors have been widely studied [3,8]. We trained an encoded feature based classification model (*i.e.*, a MLP) on a open-source colorectal cancer dataset (*i.e.*, NCT-CRC-HE-100K [14]), which is annotated with 9 classes, including: adipose tissue (ADI); background (BACK); debris (DEB); lymphocytes (LYM); mucus (MUC); muscle (MUS); normal colon mucosa (NORM); stroma (STR); tumor (TUM). The trained classification model can be used to determine the biological semantics of the pathological components extracted by our model with a major voting rule. Figure 3 shows the original image, spatial topology, proportion and

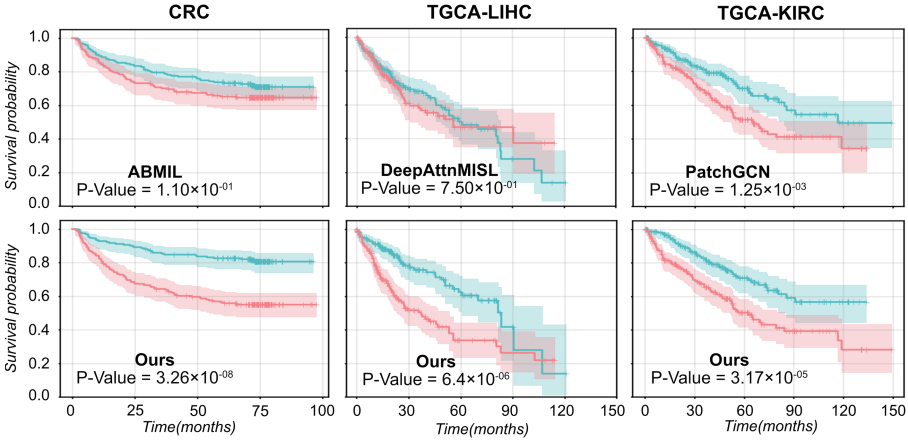

Fig. 2. KM analysis of second best SOTA method and our proposed framework for different datasets. All the patients across the five test folds are combined and analysis here. For each cohort, patients were stratified into high-risk (red curves) and low-risk (green curves) groups by the median score output by predictive models. (Color figure online)

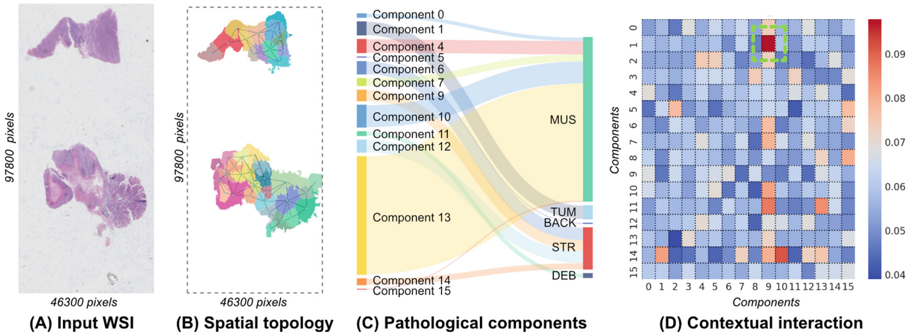

Fig. 3. Interpretability of the proposed method. A typical case in the test fold of CRC cohort is used for illustration. Best viewed by zoom in.

biological meaning of pathological components, and its contextual interactions of a typical case from CRC cohort. It can be seen that the interaction between component 1 (TUM) and component 9 (STR) has gained the highest attention of the network, which is consistent with the existing knowledge [3,8]. Moreover, there is also concentration of interaction in some other interactions, which may potentially imply some new biomarkers.

4 Conclusion

In this paper, we propose a novel learning framework, *i.e.*, multi-scope analysis driven HGT, to effectively represent and capture the contextual interaction of pathological components for improving the effectiveness and interpretability of WSI-based cancer survival prediction. Experimental results on three clinical cancer cohorts demonstrated our model achieves better performance and richer interpretability over the existing models. In the future, we will evaluate our framework on more tasks and further statistically analyze the interpretability of our model to find more pathological biomarkers related to cancer prognosis.

Acknowledgement. This work was supported by Ministry of Science and Technology of the People's Republic of China (2021ZD0201900)(2021ZD0201903).

References

1. AbdulJabbar, K., et al.: Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nat. Med. **26**(7), 1054–1062 (2020)
2. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
3. Bilal, M., et al.: Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digit. Health **3**(12), e763–e772 (2021)
4. Chen, R.J., et al.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16144–16155 (2022)
5. 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
6. Diao, J.A., et al.: Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat. Commun. **12**(1), 1613 (2021)
7. Fey, M., Lenssen, J.E.: Fast graph representation learning with PyTorch geometric. In: ICLR Workshop on Representation Learning on Graphs and Manifolds (2019)
8. Foersch, S., et al.: Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. Nat. Med. **29**, 1–10 (2023)
9. Guan, Y., et al.: Node-aligned graph convolutional network for whole-slide image representation and classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18813–18823 (2022)

10. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, pp. 1024–1034 (2017)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
12. Hou, W., et al.: H^2-mil: exploring hierarchical representation with heterogeneous multiple instance learning for whole slide image analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 933–941 (2022)
13. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136. PMLR (2018)
14. Kather, J.N., et al.: Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med. **16**(1), e1002730 (2019)
15. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2017). https://openreview.net/forum?id=SJU4ayYgl
16. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF Conference On Computer Vision and Pattern Recognition, pp. 14318–14328 (2021)
17. Liu, P., Fu, B., Ye, F., Yang, R., Ji, L.: DSCA: a dual-stream network with cross-attention on whole-slide image pyramids for cancer prognosis. Expert Syst. Appl. **227**, 120280 (2023)
18. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomed. Eng. **5**(6), 555–570 (2021)
19. 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. Adv. Neural. Inf. Process. Syst. **34**, 2136–2147 (2021)
20. Sterlacci, W., Vieth, M.: Early colorectal cancer. In: Baatrup, G. (ed.) Multidisciplinary Treatment of Colorectal Cancer, pp. 263–277. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58846-5_28
21. Sung, H., et al.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **71**(3), 209–249 (2021). https://doi.org/10.3322/caac.21660
22. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems 30 (2017)
23. Wang, P., Li, Y., Reddy, C.K.: Machine learning for survival analysis: a survey. ACM Comput. Surv. **51**(6), 3214306 (2019). https://doi.org/10.1145/3214306
24. Wang, X., et al.: RetCCL: clustering-guided contrastive learning for whole-slide image retrieval. Med. Image Anal. **83**, 102645 (2023)
25. Wang, X., et al.: SCL-WC: cross-slide contrastive learning for weakly-supervised whole-slide image classification. In: Thirty-Sixth Conference on Neural Information Processing Systems (2022)
26. 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. Med. Image Anal. **65**, 101789 (2020)
27. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J.: Deep sets. In: Advances in Neural Information Processing Systems 30 (2017)