Dynamic Graph Representation with Knowledge-aware Attention for Histopathology Whole Slide Image Analysis

Jiawen Li[1*], Yuxuan Chen[1*], Hongbo Chu[1*], Qiehe Sun[1], Tian Guan[1], Anjia Han[2†], Yonghong He[1†]

[1]Shenzhen International Graduate School, Tsinghua University
[2]Department of Pathology, The First Affiliated Hospital of Sun Yat-sen University

{lijiawen21, chenyx23, zhu-hb23}@mails.tsinghua.edu.cn

hananjia@mail.sysu.edu.cn, heyh@sz.tsinghua.edu.cn

Abstract

Histopathological whole slide images (WSIs) classification has become a foundation task in medical microscopic imaging processing. Prevailing approaches involve learning WSIs as instance-bag representations, emphasizing significant instances but struggling to capture the interactions between instances. Additionally, conventional graph representation methods utilize explicit spatial positions to construct topological structures but restrict the flexible interaction capabilities between instances at arbitrary locations, particularly when spatially distant. In response, we propose a novel dynamic graph representation algorithm that conceptualizes WSIs as a form of the knowledge graph structure. Specifically, we dynamically construct neighbors and directed edge embeddings based on the head and tail relationships between instances. Then, we devise a knowledge-aware attention mechanism that can update the head node features by learning the joint attention score of each neighbor and edge. Finally, we obtain a graph-level embedding through the global pooling process of the updated head, serving as an implicit representation for the WSI classification. Our end-to-end graph representation learning approach has outperformed the state-of-the-art WSI analysis methods on three TCGA benchmark datasets and in-house test sets. Our code is available at https://github.com/WonderLandxD/WiKG.

1. Introduction

Pathology is the gold standard for disease diagnosis. Traditional histopathological diagnostics are performed manually by examining pathological slides using an optical microscope. However, due to social concerns such as environmental changes and an increasing aging population,
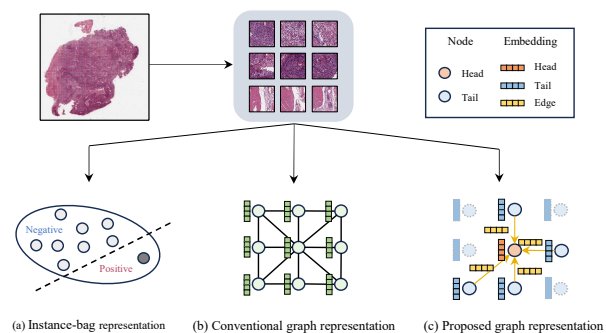

Figure 1. Illustration of three different frameworks of instance-bag representation, conventional undirected graph representation, and the proposed dynamic graph representation.

those diagnostic approaches are facing significant challenges [11]. With the advancement of digital scanning technologies, physical slides can now be converted into high-resolution WSIs that preserve the complete pathological tissues. This advancement has markedly improved the work efficiency of pathologists [33] and stimulated demand for intelligent WSI diagnostic tools [26, 27, 30].

In recent years, many studies have shown promising results by continuously exploring deep learning in WSIs. However, the professional interpretation of WSIs and their large-scale data type make obtaining manual annotation at the pixel level difficult. To alleviate this issue, researchers have developed weakly supervised algorithms that can be trained using only slide-level labels [35, 37]. Many existing weakly supervised studies on WSI analysis are based on embedding-based multiple instance learning (MIL) [28], which partitions WSIs into patches (instances) and aggregates their embedding information. However, MIL overly emphasizes the impact of patches on the global representation but ignores the interconnections. Such interconnections often indicate the potential development trends of tumors and provide pathologists with deeper relational in-

---

* Contributed equally. †Corresponding authors.

sights, contributing to the construction of tumor microenvironments. For instance, the interactions between tumor cells and inflammatory cells assist in analyzing tumor staging [4] and clinical prognosis [6, 22]. Graph Neural Networks (GNNs) are emerging as promising tools for WSI analysis, primarily because they focus more on local similarities within the entity topology [1]. GNNs represent WSIs as graphs, treat patches as nodes, and transfer and aggregate feature information of patches through the interaction between nodes to obtain graph-level representations.

Graph-based WSI analysis methods are currently devoted to modeling through the spatial positional relationships of patches. While these approaches have achieved some success, they exhibit some deficiencies. (1) Explicit spatial topology restricts the ability of distant entities to explore information with one another. (2) GNNs for WSIs are typically modeled as undirected graphs, which overlooks the directed contributions of adjacent entities. (3) The over-parameterization of GNNs in WSI analysis may lead to over-fitting, which reduces the generalization ability.

For this purpose, we introduce a dynamic graph representation based on knowledge-aware attention called WiKG, which fully leverages the interactions among entities to analyze WSIs. Specifically, we parameterize the head and tail embeddings for each patch, which are instrumental in modeling the directed graphs of WSIs (Figure 1 shows the formal differences between proposed directed graph representation and other methods). Following this, we design edge embeddings based on the interactions between the heads and tails, and construct a knowledge-aware attention mechanism that aggregates neighbor information. This mechanism injects the embeddings of head, tail, and edge triplets into neighbor knowledge attributes, allowing more valuable entities to propagate more practical information. To demonstrate the effectiveness of our method, we conduct extensive evaluations on three public benchmark TCGA datasets, including typing and staging for esophageal carcinoma, kidney cancer, and lung cancer. In addition, we collect in-house frozen section lung cancer WSIs for external testing to showcase the generalization ability of our method. Through comparisons with various state-of-the-art WSI analysis methods and ablation studies, we demonstrate the superior performance of WiKG in histopathological WSI analysis tasks.

2. Related Works

2.1. Application of MIL in Histopathology WSIs

Given the consistency between MIL theory and the diagnosis process of pathologists, the development of MIL algorithms for WSI analysis has become a principal research trend [13], where models are constructed in an instance-bag format. First, researchers attempt to generate WSI predictions based on instance scores [3, 16, 23, 40]. As the embedding-based MIL method has developed, the research trend gradually turned to it because of its more stable convergence and better performance. Conventional embedding-based MIL approaches involve compressing patches into embeddings and aggregating these to learn WSI-level features for final predictions [8, 19, 24, 25]. For example, ABMIL [19] learns an attention score for each instance and performs a weighted aggregation to obtain a bag representation at the WSI level. Lu et al. [25] propose an improved ABMIL by introducing clustering-constraint loss and single/multiple-branch attention to enhance WSI classification capabilities. Li et al. [24] propose a dual-stream network to capture the latent relationships between the most suspicious positive instances and others. Shao et al. [34] integrate transformer networks into the MIL paradigm to update inter-instance embeddings efficiently. Zhang et al. [42] build upon ABMIL by introducing the concept of pseudo-bags to enhance the attention information and proposed a feature distillation strategy to fuse representations at the pseudo-bag level. Though these methods have endeavored to improve the analytical performance of WSIs, they still lack representation of the informative interactions to inter-patch interactions.

2.2. Graph Representation in Digital Pathology

The tumor microenvironment, as delineated by histological phenotypes and topological distribution, plays a pivotal role in pathological diagnosis. Graph-based deep learning approaches, outstanding in encoding representations and capturing intra- and inter-nodal interactions, have attracted widespread attention in digital pathology [1]. Specifically, Zhou et al. [45] and Chen et al. [5] construct edges through the spatial position of cells, coupled with hand-crafted features to form the cell graph representations in tissue. As the paradigm of efficient feature extraction through deep learning, training neural networks at the cellular level can enhance the performance of the cell graph [32]. Some studies have also designed topological structures by integrating representations of tissue regions and cells. For example, HACT-Net [31] combines tissue and cell graph representations to classify histopathological regions of interest in the breast subtype. SHGNN [17] constructs a spatially hierarchical GNN framework for joint representation learning on tissue and cell structures.

With the necessity to analyze whole slide images (WSIs) in tissue pathology, graph representations that use patches as entities have also been widely discussed. For instance, Chen et al. [6] model the local and global topological structures within the tumor microenvironment by hierarchically aggregating patch-level histological features. Zheng et al. [44] combine the spatial representation of WSIs with Vision Transformers to predict disease grades. Hou et al. [18]
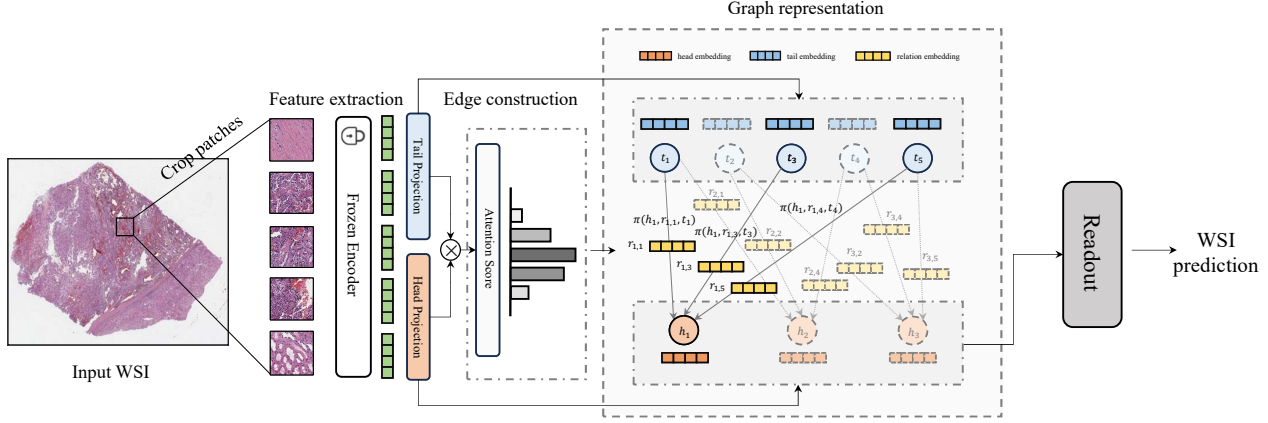
Figure 2. The framework of our proposed method for WSI analysis, including patch feature extraction, dynamic edge construction based on head and tail embeddings, graph representation learning, and the prediction of WSIs.

develop a multiscale hierarchical representation using heterogeneous graphs with resolution attributes to explicitly model the spatial scale relationships of patches. Zhao *et al.* [43] establish a multi-task learning framework for WSI analysis using graph neural networks and Transformers as the building base. Chan *et al.* [4] analyze WSIs by representing them as heterogeneous graphs, utilizing the relationships between different types of nuclei. While these methods have demonstrated impressive results, most rely on fixed spatial constructs for edge creation, limiting the ability to explore mutual information freely. Simultaneously, they fail to fully account for the directional complexity of contextual interactions between patches.

3. Methodology

In this section, we introduce our proposed WiKG, including how dynamic graph construction works and the knowledge-aware attention mechanism for updating node features. Figure 2 presents our proposed framework.

3.1. Dynamic Graph Construction

Unlike conventional methods that construct graphs using spatial relationships [6, 18, 44], our approach is dedicated to quantifying positional relationships between patches through learnable implicit features. For a given WSI, we first use the Otsu threshold algorithm [29] to distinguish the foreground tissue regions and apply the sliding window operation to divide them into non-overlapping patches $X = \{x_1, x_2, \ldots, x_n\}$, which are denoted as graph nodes. Subsequently, we utilize a feature encoder f (e.g., Vision Transformer [10] pretrained on ImageNet [9]) to obtain embeddings for each patch. Following this, we project these into head and tail embeddings using two separate linear projection layers, where the head aims to explore the correlation between other patches and itself, and the tail aims to

explore its contribution to other patches.

$$h_i = W_h f(X), \quad t_i = W_t f(X), \quad (1)$$

where h_i and t_i respectively denote the head and tail embeddings of patch i. We then compute the dot product of these and employ a softmax function to quantify the similarity between heads and tails, which is expressed as:

$$\omega_{i,j} = \frac{h_i^{\mathrm{T}} t_j}{\sum_{j=1}^{N}(h_i^{\mathrm{T}} t_j)}, \quad (2)$$

where $\omega_{i,j}$ represents the similarity score between the head of patch i and the tail of patch j. For each patch i, the top k patches with the highest similarity score are selected as the neighbors of patch i, which is described as:

$$\mathcal{N}(i) = \{j \in V : \omega_{i,j} \in \mathrm{Topk}\{\omega_{i,j}\}_{j=1}^{N}\}, \quad (3)$$

where V represents the patch set, and $|V| = N$, $|\mathcal{N}(i)| = k$. Our topological structure also assigns embeddings for directed edges, obtained using head and tail embeddings:

$$r_{i,j} = \omega_{i,j} t_j + (1 - \omega_{i,j}) h_i, \quad \text{for every } j \in \mathcal{N}(i), \quad (4)$$

where $r_{i,j}$ represents the edge embedding from patch j to patch i. Through the above operations, we delineate a WSI as a dynamic graph representation $G = (V, \mathcal{E}, \mathcal{F}, \mathcal{R})$, where V denotes the set of nodes, $\mathcal{E}$ denotes the set of edges, $\mathcal{F}$ denotes the set of head and tail embeddings and $\mathcal{R}$ denotes the set of directed edge embeddings. $\mathcal{E} = \{(h, r, t) : (h, t) \in \mathcal{F}, r \in \mathcal{R}\}$ introduces the triplet of head, tail, and their high dimensional relation on each directed edge. Algorithm 1 provides the pseudo-code for our proposed dynamic graph construction in a Pytorch-like style.

Algorithm 1: Directional Edge Construction

```
# input:  patch embedding (B, N, C)
# output:  relation embedding (B, N, k, C)
# B: batch size
# N: patch number
# C: feature dimension
# K: neighbor nodes number

# linear projection of head, tail
e_h, e_t = linear_ht(input).chunk(2, dim=-1)

# similarity of head and tail
attn_logit = (e_h * scale) @ e_t.transpose(-2, -1)

# chooses top k weight and index
topk_weight, topk_index = torch.topk(attn_logit,
  k=k, dim=-1)

# expand topk_index to match dimensions
topk_idx_expand = topk_index.expand(B, -1, -1)

# create a range tensor for secondary indexing
batch_indices = torch.arange(B).view(-1, 1, 1)

# generate neighbor features (B, N, k, C)
Nb_h = e_t[batch_indices, topk_idx_expand, :]

# weight probability
topk_prob = F.softmax(topk_weight, dim=2)

# generate edge embeddings (B, N, k, C)
eh_r = topk_prob.unsqueeze(-1) * Nb_h + (1 -
  topk_prob).unsqueeze(-1) @ e_h.unsqueeze(2)
```

3.2. Knowledge-aware Attention Mechanism

To take full advantage of the node relationships of the above graph construction, inspired by graph for recommendation [12, 38], we propose a knowledge-aware attention mechanism for information propagation and aggregation across nodes. For patch i, we compute a linear combination of the tail embeddings of its neighbors $\mathcal{N}(i)$ to characterize its first-order connectivity structure:

$$h_{\mathcal{N}(i)} = \sum_{j \in \mathcal{N}(i)} \pi(h_i, r_{i,j}, t_j) t_j, \quad (5)$$

where $\pi(h, r, t)$ is a weighting factor that guides how much information from each tail will be propagated to the head. We use the nonlinear combination of triples to calculate $\pi(h, r, t)$, which is expressed as follows:

$$u(h_i, r_{i,j}, t_j) = t_j^{\mathrm{T}} \tanh(h_i + r_{i,j}), \quad (6)$$

where the hyperbolic tangent $\tanh(\cdot)$ serves as a non-linear activation function, facilitating an appropriate gradient flow encompassing both negative and positive values. This combination permits the assessment of the proximity between tails and heads to reveal differences between neighbors. Subsequently, we employ the softmax function to normalize these combinations:

$$\pi(h_i, r_{i,j}, t_j) = \frac{\exp\{u(h_i, r_{i,j}, t_j)\}}{\sum_{j \in \mathcal{N}(i)} \exp\{u(h_i, r_{i,j}, t_j)\}}. \quad (7)$$
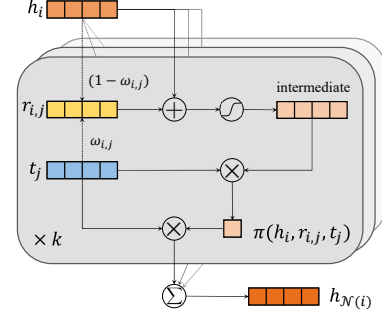

Figure 3. Illustration of our proposed knowledge-aware attention mechanism, including aggregation between head, tail, and edge embeddings.

By modeling relationships of triplets and describing them as knowledge information for edges, head nodes are allowed to measure signals from tail nodes and capture them efficiently. Figure 3 illustrates the implementation of our proposed knowledge-aware attention mechanism. Ultimately, we fuse the aggregated neighbor information with the original head to form a new head representation. We adopt a dual-interaction mechanism to facilitate the exchange of more messages between nodes:

$$h_i = \sigma_1(W_1(h_i + h_{\mathcal{N}(i)})) + \sigma_2(W_2(h_i \odot h_{\mathcal{N}(i)})), \quad (8)$$

where σ is the activation function, such as LeakyReLU, and W denotes a learnable transformation matrix. Finally, we employ a $\mathrm{Readout}$ function to generate graph-level embeddings and a softmax function to obtain the probability score of the WSI:

$$\hat{Y} = \mathrm{Softmax}(\mathrm{Readout}(G)), \quad (9)$$

where $\mathrm{Readout}$ is a global pooling layer, such as mean or max pooling, and $\hat{Y}$ denotes the predicted probabilities. During training, the cross-entropy loss function is utilized as the objective loss for the WSI classification task, which is expressed as:

$$\mathcal{L}_{ce} = -\frac{1}{M} \sum_{m=1}^{M} \sum_{c=1}^{C} Y_{m,c} \ln(\hat{Y}_{m,c}), \quad (10)$$

where C is the number of categories, M is the number of training samples, and Y is the one-hot label.

4. Experiments

In this section, we introduce the performance of the proposed WiKG on three public histopathological benchmark datasets and compare it with other state-of-the-art WSI analysis algorithms. Moreover, we also conduct ablation studies on edge construction, know-aware attention, the number of neighbor nodes, convergence, and efficiency. Finally, we perform additional experiments on the in-house test set to show the generalization of the trained models.

Method	TCGA-ESCA			TCGA-KIDNEY			TCGA-LUNG		
	Accuracy	AUC	F1-score	Accuracy	AUC	F1-score	Accuracy	AUC	F1-score
ABMIL [19]	$85.61_{2.00}$	$90.74_{0.90}$	$85.70_{1.93}$	$94.97_{0.85}$	$98.69_{0.65}$	$94.98_{0.88}$	$80.67_{2.37}$	$84.80_{3.77}$	$80.35_{2.76}$
CLAM-SB [25]	$88.27_{2.30}$	$93.56_{1.40}$	$88.28_{2.28}$	$96.68_{0.93}$	$99.49_{0.20}$	$96.67_{0.94}$	$82.89_{1.54}$	$89.25_{1.87}$	$82.86_{1.52}$
CLAM-MB [25]	$87.21_{4.05}$	$92.83_{1.64}$	$87.27_{4.03}$	$96.92_{0.42}$	$99.55_{0.08}$	$96.91_{0.42}$	$80.81_{1.08}$	$88.66_{1.67}$	$80.77_{0.98}$
DSMIL [24]	$87.20_{0.82}$	$91.91_{1.34}$	$87.25_{0.79}$	$96.19_{1.31}$	$99.28_{0.26}$	$96.17_{1.31}$	$81.80_{1.20}$	$88.30_{1.60}$	$81.78_{1.27}$
TransMIL [34]	$86.70_{2.04}$	$92.85_{2.06}$	$86.81_{1.94}$	$95.38_{1.47}$	$99.22_{0.51}$	$95.36_{1.49}$	$78.31_{2.37}$	$85.89_{3.48}$	$78.03_{2.66}$
DTFD-MIL [42]	$88.81_{2.49}$	$93.79_{2.22}$	$88.88_{2.43}$	$96.27_{1.00}$	$99.41_{0.30}$	$96.26_{1.01}$	$82.41_{0.82}$	$88.70_{1.45}$	$82.37_{0.88}$
HIPT [7]	$89.24_{4.25}$	$93.79_{3.12}$	$89.27_{4.27}$	$96.84_{0.85}$	$99.48_{0.14}$	$96.83_{0.86}$	$80.39_{1.13}$	$86.91_{0.58}$	$80.35_{1.14}$
GTP [44]	$78.03_{5.80}$	$84.61_{5.33}$	$77.76_{6.09}$	$83.54_{4.35}$	$93.71_{3.33}$	$82.76_{5.03}$	$78.07_{1.31}$	$84.01_{2.06}$	$77.98_{1.25}$
Patch-GCN [6]	$88.30_{2.13}$	$93.98_{0.83}$	$88.37_{2.06}$	$96.68_{1.11}$	$99.53_{0.22}$	$96.67_{1.10}$	$79.72_{3.67}$	$87.83_{4.30}$	$79.88_{3.59}$
WiKG (ours)	$\mathbf{90.37_{3.14}}$	$\mathbf{95.23_{2.90}}$	$\mathbf{90.40_{3.13}}$	$\mathbf{97.08_{0.71}}$	$\mathbf{99.65_{0.12}}$	$\mathbf{97.08_{0.71}}$	$\mathbf{84.02_{0.72}}$	$\mathbf{90.78_{1.24}}$	$\mathbf{83.93_{0.64}}$

Table 1. Cancer typing results [%] of various methods on TCGA-ESCA, TCGA-KIDNEY, and TCGA-LUNG datasets.

4.1. Datasets

Datasets	Type	Number of WSIs			
		Stage I	Stage II	Stage III	Stage IV
TCGA-ESCA	1	30	56	75	19
	2	13	115	55	12
TCGA-KIDNEY	1	129	109	62	26
	2	270	58	124	83
	3	231	44	70	27
TCGA-LUNG	1	644	378	232	20
	2	464	200	141	42
FROZEN-LUNG	1	65			
	2	105			

Table 2. Statistics of three public datasets and in-house datasets.

We evaluate our WiKG and other methods on three public clinical WSI benchmark datasets from The Cancer Genome Atlas (TCGA) project [39], including ESCA, KIDNEY, and LUNG. We performed two-group classification experiments on tumor type and stage for each dataset. TCGA-ESCA comprises an esophageal cancer cohort of 375 cases, including adenocarcinomas (1) and squamous cell carcinomas (2). TCGA-KIDNEY comprises a kidney cancer cohort of 1,233 cases, including chromophobe renal cell carcinoma (1), renal clear cell carcinoma (2), and renal papillary cell carcinoma (3). TCGA-LUNG comprises a lung cancer cohort of 2121 cases, including squamous cell carcinomas (1) and adenocarcinomas (2). For staging tasks, all cases are divided into "I", "II", "III" and "IV" categories according to TNM standards [14].

For preprocessing, each WSI is divided into 256×256 non-overlapping patches at $20\times$ magnifications. The Area Under the Curve (AUC) is reported for all experiments due to its comprehensiveness and insensitivity to class imbalance. Additionally, Accuracy and weighted F1-score are also reported. These metrics are determined by a threshold of 0.5. All results on TCGA datasets are obtained through 4-fold cross-validation, with each metric recorded as a percentage value accompanied by the standard error. We visualize the constructed graph structure in the Camelyon16 dataset [2], chosen for its pixel-level annotations of tumor mask. To further evaluate the generalization performance of our proposed method, we also conduct in-house testing with 170 real cases of frozen section WSIs collected from the First Affiliated Hospital of Sun Yat-sen University, including squamous cell carcinomas (1) and adenocarcinomas (2), documented under the FROZEN-LUNG. Table 2 shows more details of the dataset.

4.2. Implementation Details

The proposed methodology uses the Pytorch library in Python on an Nvidia RTX 3090 GPU. To adopt a single GPU and facilitate future work, we utilize a ViT Small model [10] pretrained on ImageNet [9] as the encoder. For comparative experiments utilizing domain-specific encoders trained on pathology images, please refer to the Supplementary material. We extract a 384-dimensional feature vector for each patch and then expand it to 512 dimensions via a fully connected layer, forming the final feature representation of the patch. The batch size is set to 1, indicating that only one slide is processed per iteration. The k in Top-k is chosen to be 6, which means that each patch will have 6 patches connected to it. We set the number of epochs to 100 and employ the Adam [20] optimizer with a learning rate of 10^{-4} and a weight decay of 10^{-5} for model weight updates. A dropout ratio of 0.3 is applied before obtaining graph-level embeddings. All other baseline methods are experimented with the same parameters and settings.

Method	TCGA-ESCA			TCGA-KIDNEY			TCGA-LUNG		
	Accuracy	AUC	F1-score	Accuracy	AUC	F1-score	Accuracy	AUC	F1-score
ABMIL [19]	$51.21_{4.10}$	$64.28_{4.00}$	$48.51_{2.75}$	$51.99_{2.08}$	$63.42_{2.24}$	$47.10_{2.45}$	$52.57_{0.72}$	$57.33_{1.42}$	$45.39_{0.98}$
CLAM-SB [25]	$51.47_{3.24}$	$65.08_{1.46}$	$48.70_{3.84}$	$51.99_{2.62}$	$66.07_{2.97}$	$47.83_{2.44}$	$51.02_{2.19}$	$57.22_{2.86}$	$44.73_{2.95}$
CLAM-MB [25]	$54.15_{4.65}$	$65.81_{2.30}$	$51.56_{3.76}$	$51.91_{3.27}$	$67.67_{2.50}$	$48.63_{1.85}$	$50.21_{1.66}$	$57.32_{2.01}$	$43.99_{1.52}$
DSMIL [24]	$51.75_{4.58}$	$66.03_{6.05}$	$49.07_{5.85}$	$52.32_{3.02}$	$66.85_{2.80}$	$48.16_{2.29}$	$52.33_{1.21}$	$57.93_{2.74}$	$44.96_{2.07}$
TransMIL [34]	$53.61_{3.44}$	$65.43_{4.07}$	$49.86_{5.96}$	$50.37_{1.43}$	$61.96_{2.64}$	$43.73_{2.60}$	$52.15_{1.06}$	$55.09_{0.94}$	$42.51_{1.49}$
DTFD-MIL [42]	$50.68_{5.29}$	$65.60_{5.18}$	$48.21_{5.26}$	$51.75_{2.00}$	$64.03_{4.26}$	$45.63_{3.38}$	$52.52_{0.78}$	$57.56_{1.71}$	$44.88_{1.76}$
HIPT [7]	$48.82_{6.16}$	$63.46_{3.89}$	$35.55_{12.13}$	$51.81_{2.96}$	$61.83_{1.40}$	$37.40_{1.53}$	$52.24_{4.18}$	$55.26_{0.46}$	$35.92_{4.74}$
GTP [44]	$46.38_{4.03}$	$63.55_{3.88}$	$45.90_{3.96}$	$46.05_{2.14}$	$57.01_{1.00}$	$39.73_{2.14}$	$50.85_{3.77}$	$55.22_{4.10}$	$43.34_{2.12}$
Patch-GCN [6]	$52.28_{3.98}$	$68.03_{3.72}$	$50.47_{5.18}$	$54.26_{3.26}$	$68.96_{2.22}$	$49.89_{2.26}$	$50.63_{0.58}$	$53.82_{1.43}$	$42.62_{2.64}$
WiKG (ours)	$\mathbf{57.50_{3.99}}$	$\mathbf{69.96_{4.28}}$	$\mathbf{55.80_{4.39}}$	$\mathbf{55.49_{2.12}}$	$\mathbf{69.71_{1.40}}$	$\mathbf{51.23_{1.43}}$	$\mathbf{52.85_{0.74}}$	$\mathbf{60.34_{1.37}}$	$\mathbf{47.52_{1.52}}$

Table 3. Cancer staging results [%] of various methods on TCGA-ESCA, TCGA-KIDNEY, and TCGA-LUNG datasets.

4.3. Comparison with Other Existing Works

We present our tumor typing and staging results on three datasets in Table 1 and Table 3, respectively, and compare them with the following nine state-of-the-art methodologies: (1) ABMIL [19], a MIL framework that aggregates instance information via an attention mechanism to obtain bag-level embeddings. (2) CLAM-SB [25], a single-gated attention-based MIL framework optimized with clustering constraint loss. (3) CLAM-MB [25], an enhanced version of CLAM-SB employing multi-gated attention. (4) DSMIL [24], a dual-stream MIL method aggregating patch embeddings through non-local attention pooling and max pooling. (5) TransMIL [34], a MIL-like paradigm using transformers and multiscale position encoding modules. (6) DTFD-MIL [42], a double-tier MIL framework developed by introducing pseudo-bags and feature distillation. (7) HIPT [7], a hierarchical ViT architecture for leveraging Pyramid WSI structure. (8) GTP [44], a graph representation integrating graph and transformer. (9) Patch-GCN [6], a hierarchical graph-based method for WSI survival prediction, with global attention pooling introduced as a graph representation model. Our method demonstrates superior performance compared to MIL and graph representation methods, suggesting that our WiKG might better capture the interactions between patches, thus obtaining more comprehensive global information. Notably, in staging tasks, the improvement in F1-score is more significant than in Accuracy, indicating the advantage of WiKG in delineating the malignancy degree of subdivided tumors. It is necessary to acknowledge that the staging of WSIs is generally influenced by factors including the type of tumor, its dimension, the extent of invasion into adjacent organs, the number of cancerous cells in lymph nodes, and the presence of distant metastases. Most current algorithms employed in this domain are adaptations from those developed for typing tasks. In our future work, we will explore deep learning methods for tumor staging.

4.4. Effectiveness of Proposed Edge Construction

Datasets	Task	Edge Construction		
		k-NN (cos)	k-NN (dist)	WiKG (ours)
ESCA	Type	$94.98_{2.52}$	$92.36_{1.11}$	$\mathbf{95.23_{2.90}}$
	Stage	$64.36_{4.38}$	$66.27_{4.69}$	$\mathbf{69.96_{4.28}}$
KIDNEY	Type	$99.44_{0.28}$	$99.47_{0.24}$	$\mathbf{99.65_{0.12}}$
	Stage	$68.62_{1.97}$	$68.01_{1.77}$	$\mathbf{69.71_{1.40}}$
LUNG	Type	$90.62_{1.40}$	$88.31_{2.22}$	$\mathbf{90.78_{1.24}}$
	Stage	$58.72_{1.58}$	$57.92_{1.99}$	$\mathbf{60.34_{1.37}}$

Table 4. Cancer typing and staging AUC results [%] of our proposed dynamic graph construction compared to graph construction based on the k-nearest neighbor algorithm.

Our strategy to construct directed edges based on the heads and tails relation demonstrates efficiency in establishing topological structures. In Table 4, we report a comparative analysis of the AUC performance between our method and two conventional dynamic edge construction algorithms based on the k-nearest neighbor (k-NN). Specifically, k-NN (dist) and k-NN (cos) identify the top k indexes with the highest similarity by calculating the Euclidean distance and cosine similarity between feature embeddings respectively, and keep other structures unchanged. Notably, our approach outperforms the other two k-NN-based methods by nearly one to three percentage points across two WSI analysis tasks in all three datasets, suggesting that our edge-building strategy may be more consistent with representations of the organizational microenvironment.

4.5. Effectiveness of Knowledge-aware Attention

GNN	ESCA	KIDNEY	LUNG
GCN [21]	$94.24_{2.84}$	$99.44_{0.14}$	$90.33_{1.20}$
GIN [41]	$93.58_{2.37}$	$99.56_{0.18}$	$90.37_{1.32}$
SAGE [15]	$93.52_{2.60}$	$99.51_{0.23}$	$90.40_{1.14}$
GAT [36]	$93.77_{2.68}$	$99.54_{0.21}$	$90.41_{1.11}$
WiKG(ours)	$\mathbf{95.23_{2.90}}$	$\mathbf{99.65_{0.12}}$	$\mathbf{90.78_{1.24}}$

Table 5. Cancer typing AUC results [%] of our proposed WiKG compared to other GNN architecture.

We conduct a comparative performance analysis with other prevalent GNN architectures to further substantiate the performance of WiKG. We substituted our knowledge-aware attention mechanism update strategy with four different GNN architectures, which are (1) graph convolutional network (GCN) [21], graph isomorphism network (GIN) [41], graphSage (SAGE) [15], and graph attention network (GAT) [36]. All other structures of the model remain unchanged. Table 5 shows the performance comparison of our proposed WiKG with different GNN architectures, demonstrating the superiority of our proposed method in terms of AUC across three WSI typing tasks. Especially in the cancer analysis of esophageal, the proposed method improves up to nearly 2%. It is worth noting that GAT is also a GNN framework that aggregates node features through attention scores. Still, its performance is not as good as WiKG, which shows that WiKG may have more advantages in capturing effective information of neighbors.

4.6. Effect of Numbers of Neighbor Nodes

We also conduct analysis experiments on the different numbers of neighbor nodes. Figure 4 illustrates the Accuracy and AUC scores of WiKG in the typing and staging experiments across three TCGA datasets with varying numbers of neighbor nodes. Overall, the number of neighbor nodes has little impact on the performance of WiKG. Although the Accuracy metric fluctuates in the staging experiment of TCGA-ESCA, this fluctuation is reasonable considering the uneven number of each staging. At the same time, this may also be related to the inherent morphological diversity and complexity of esophageal cancer.

4.7. Convergence and Time Efficiency

The over-parameterization problem can lead to unstable convergence of GNNs and increased training time. We compare WiKG with Patch-GCN and GTP in the staging task of TCGA-ESCA. Figure 5 compares our convergence curve and the training time of each epoch. WiKG converges quickly during training and shows no over-fitting
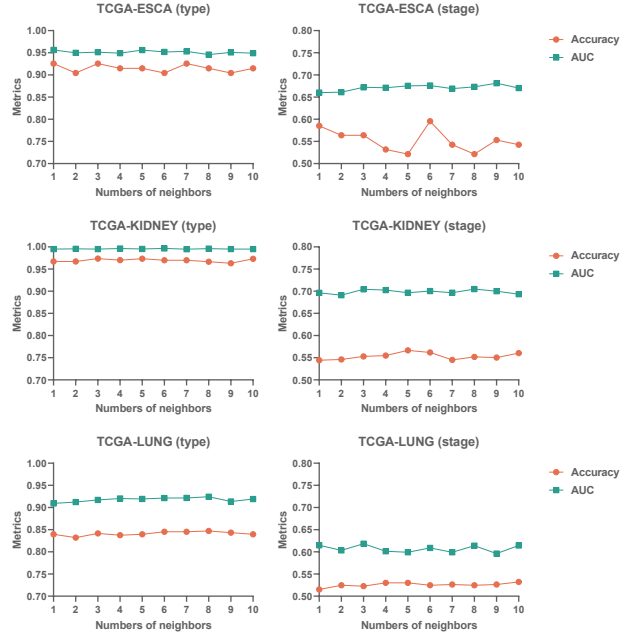

Figure 4. Typing and staging results of AUC and Accuracy scores with different numbers of neighbor nodes on three TCGA datasets.

Figure 5. Convergence curves of validation AUC and training time of each epoch in TCGA-ESCA.

phenomenon compared with Patch-GCN and GTP. At the same time, in terms of training speed in a single epoch, WiKG is also the fastest and has the smallest number of parameters.

4.8. Visualization and Further Experiment

Through the visualization of the directed graph structure constructed by WiKG, we gain an intuitive understanding of the potential relationships between patches, further helping to build relationship networks of cancer cells within tissues and enriching the representation of the tumor microenvironment. Figure 6 displays our visualization results on the Camelyon16 dataset. We observe that the patch with metastatic cancer is more likely to select patches containing a large number of lymphocytes as neighbors and some adipose and fibrous tissue as supplements. Notably, the patch with the highest knowledge-aware attention score (number 5) contains a greater concentration of lymphomas, indicating that the patch containing metastatic cancer pays more
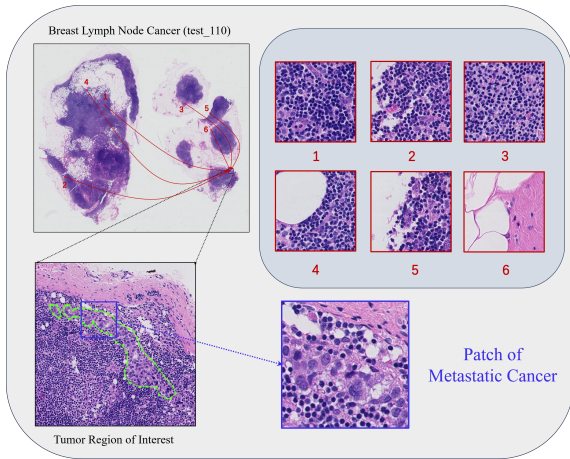
Figure 6. Visualization of the constructed graph structure.

Figure 7. Mean and standard deviation of per-class Accuracy with various methods on FROZEN-LUNG datasets.

attention to patches containing more lesion features.

We also conduct further testing experiments on the FROZEN-LUNG dataset with models previously trained well on the TCGA-LUNG dataset. The purpose of these experiments is to evaluate the generalization performance and application prospects of various methods. Table 6 provides a comparative analysis with other methods, demonstrating the superior overall performance of WiKG. At the same time, we also show the results of the Accuracy metric for each category, which is demonstrated in Figure 7. Specifically, WiKG has excellent recognition capabilities for both lung subtypes, and the standard deviation of each category is also much smaller than other methods. However, the generalization capabilities of other MIL methods are not comparable to WiKG.

Method	FROZEN-LUNG		
	Accuracy	AUC	F1-score
ABMIL [19]	$70.15_{7.15}$	$80.12_{10.73}$	$70.73_{8.41}$
CLAM-SB [25]	$84.71_{2.40}$	$92.09_{3.11}$	$84.41_{2.43}$
CLAM-MB [25]	$83.82_{4.91}$	$93.71_{0.98}$	$83.64_{4.91}$
DSMIL [24]	$86.18_{2.61}$	$93.91_{0.29}$	$85.66_{2.90}$
TransMIL [34]	$70.29_{8.07}$	$\mathbf{94.10_{1.66}}$	$69.64_{8.91}$
DTFD-MIL [42]	$77.50_{4.25}$	$87.28_{3.46}$	$77.70_{4.21}$
WiKG (ours)	$\mathbf{87.06_{1.66}}$	$92.31_{0.93}$	$\mathbf{87.05_{1.68}}$

Table 6. Results [%] of various methods on FROZEN-LUNG datasets. All methods are trained well on TCGA-LUNG datasets.

Some interesting results are observed in this experiment. For most methods, there is a significant difference between AUC and accuracy/F1-score, especially for TransMIL. Although its Accuracy and F1-score are relatively low, its AUC is unusually high, specifically for squamous cell carcinomas, which is nearly 44 percent Accuracy higher than
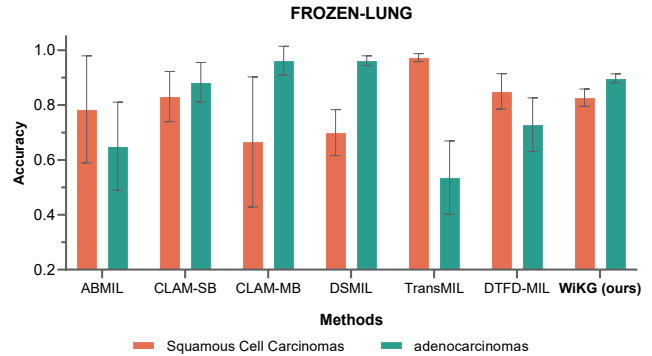
that of adenocarcinomas. This may be due to the imbalance in the size of the training data, which is that the greater the number of a specific category in the training set, the higher the Accuracy of its test. However, other methods show the opposite results except TransMIL, ABMIL, and DTFD-MIL. We also observe that other compared methods typically encountered two issues: a substantial discrepancy in Accuracy across different categories and a high standard deviation in Accuracy of specific categories. This suggests that some existing works might suffer from overfitting issues and exhibit increased sensitivity to particular categories. Our proposed WiKG, in contrast, significantly mitigated both these issues.

5. Conclusion and Future Work

In this paper, we introduce a novel dynamic graph representation approach called WiKG to analyze WSIs. By modeling interactions between head and tail embeddings and constructing a knowledge-aware attention mechanism that aggregates neighbor information, WiKG liberates the ability of patches to explore mutual relationships in their topological structures and also uses the directional contributions between entities to better interact between patches, thereby achieving better WSI analysis performance. Through extensive experiments and ablation studies of three public benchmark datasets and in-house test sets, we demonstrate the effectiveness and better generalization performance of WiKG.

In future works, we will focus on the interpretability of WiKG in WSIs and explore the impact of graph pooling on WSIs. We will also conduct in-depth research on the generalization of WSI analysis tasks, because existing research methods may not have good generalization performance.

References

[1] David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. A survey on graph-based deep learning for computational histopathology. *Computerized Medical Imaging and Graphics*, 95:102027, 2022. 2

[2] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017. 5

[3] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019. 2

[4] Tsai Hor Chan, Fernando Julio Cendra, Lan Ma, Guosheng Yin, and Lequan Yu. Histopathology whole slide image analysis with heterogeneous graph representation learning. In *CVPR*, pages 15661–15670, 2023. 2, 3

[5] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 41(4):757–770, 2020. 2

[6] Richard J Chen, Ming Y Lu, Muhammad Shaban, Chengkuan Chen, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In *MICCAI*, pages 339–349, 2021. 2, 3, 5, 6

[7] Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *CVPR*, pages 16144–16155, 2022. 5, 6

[8] Yang Chen, Zhuchen Shao, Hao Bian, Zijie Fang, Yifeng Wang, Yuanhao Cai, Haoqian Wang, Guojun Liu, Xi Li, and Yongbing Zhang. dmil-transformer: Multiple instance learning via integrating morphological and spatial information for lymph node metastasis classification. *IEEE Journal of Biomedical and Health Informatics*, 2023. 2

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. Ieee, 2009. 3, 5

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3, 5

[11] E Christopher Ellison, Timothy M Pawlik, David P Way, Bhagwan Satiani, and Thomas E Williams. The impact of the aging population and incidence of cancer on future projections of general surgical workforce needs. *Surgery*, 163 (3):553–559, 2018. 1

[12] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *WWW*, pages 417–426, 2019. 4

[13] Michael Gadermayr and Maximilian Tschuchnig. Multiple instance learning for digital pathology: A review on the state-of-the-art, limitations & future potential. *arXiv preprint arXiv:2206.04425*, 2022. 2

[14] Peter Goldstraw, John Crowley, Kari Chansky, Dorothy J Giroux, Patti A Groome, Ramon Rami-Porta, Pieter E Postmus, Valerie Rusch, Leslie Sobin, International Association for the Study of Lung Cancer International Staging Committee, et al. The iaslc lung cancer staging project: proposals for the revision of the tnm stage groupings in the forthcoming (seventh) edition of the tnm classification of malignant tumours. *Journal of thoracic oncology*, 2(8):706–714, 2007. 5

[15] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017. 7

[16] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel H Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *CVPR*, pages 2424–2433, 2016. 2

[17] Wentai Hou, Helong Huang, Qiong Peng, Rongshan Yu, Lequan Yu, and Liansheng Wang. Spatial-hierarchical graph neural network with dynamic structure learning for histological image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 181–191. Springer, 2022. 2

[18] Wentai Hou, Lequan Yu, Chengxuan Lin, Helong Huang, Rongshan Yu, Jing Qin, and Liansheng Wang. Hˆ2-mil: Exploring hierarchical representation with heterogeneous multiple instance learning for whole slide image analysis. In *AAAI*, pages 933–941, 2022. 2, 3

[19] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136, 2018. 2, 5, 6, 8

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[21] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2016. 7

[22] Yongju Lee, Jeong Hwan Park, Sohee Oh, Kyoungseob Shin, Jiyu Sun, Minsun Jung, Cheol Lee, Hyojin Kim, Jin-Haeng Chung, Kyung Chul Moon, et al. Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning. *Nature Biomedical Engineering*, pages 1–15, 2022. 2

[23] Marvin Lerousseau, Maria Vakalopoulou, Marion Classe, Julien Adam, Enzo Battistella, Alexandre Carré, Théo Estienne, Théophraste Henry, Eric Deutsch, and Nikos Paragios. Weakly supervised multiple instance learning histopatho-

logical tumor segmentation. In *MICCAI*, pages 470–479. Springer, 2020. 2

[24] Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *CVPR*, pages 14318–14328, 2021. 2, 5, 6, 8

[25] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021. 2, 5, 6, 8

[26] Anant Madabhushi. Digital pathology image analysis: opportunities and challenges. *Imaging in medicine*, 1(1):7, 2009. 1

[27] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33:170–175, 2016. 1

[28] Oded Maron and Tomás Lozano-Pérez. A framework for multiple-instance learning. *Advances in neural information processing systems*, 10, 1997. 1

[29] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979. 3

[30] JD Pallua, A Brunner, B Zelger, M Schirmer, and J Haybaeck. The future of pathology is digital. *Pathology-Research and Practice*, 216(9):153040, 2020. 1

[31] Pushpak Pati, Guillaume Jaume, Lauren Alisha Fernandes, Antonio Foncubierta-Rodríguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio, Nadia Brancati, Daniel Riccio, Maurizio Di Bonito, et al. Hactnet: A hierarchical cell-to-tissue graph neural network for histopathological image classification. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*, pages 208–219. Springer, 2020. 2

[32] Pushpak Pati, Guillaume Jaume, Antonio Foncubierta-Rodriguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio, Nadia Brancati, Maryse Fiche, Estelle Dubruc, Daniel Riccio, et al. Hierarchical graph representations in digital pathology. *Medical image analysis*, 75:102264, 2022. 2

[33] Robert Pell, Karin Oien, Max Robinson, Helen Pitman, Nasir Rajpoot, Jens Rittscher, David Snead, Clare Verrill, UK National Cancer Research Institute (NCRI) Cellular-Molecular Pathology (CM-Path) quality assurance working group, Owen J Driskell, et al. The use of digital pathology and image analysis in clinical trials. *The Journal of Pathology: Clinical Research*, 5(2):81–90, 2019. 1

[34] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021. 2, 5, 6, 8

[35] Yash Sharma, Aman Shrivastava, Lubaina Ehsan, Christopher A Moskaluk, Sana Syed, and Donald Brown. Cluster-

to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In *MIDL*, pages 682–698, 2021. 1

[36] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018. 7

[37] Xi Wang, Hao Chen, Caixia Gan, Huangjing Lin, Qi Dou, Efstratios Tsougenis, Qitao Huang, Muyan Cai, and Pheng-Ann Heng. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE transactions on cybernetics*, 50(9):3950–3962, 2019. 1

[38] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *KDD*, pages 950–958, 2019. 4

[39] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013. 5

[40] Gang Xu, Zhigang Song, Zhuo Sun, Calvin Ku, Zhe Yang, Cancheng Liu, Shuhao Wang, Jianpeng Ma, and Wei Xu. Camel: A weakly supervised learning framework for histopathology image segmentation. In *CVPR*, pages 10682–10691, 2019. 2

[41] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2018. 7

[42] Hongrun Zhang, Yanda Meng, Yitian Zhao, Yihong Qiao, Xiaoyun Yang, Sarah E Coupland, and Yalin Zheng. Dtfdmil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In *CVPR*, pages 18802–18812, 2022. 2, 5, 6, 8

[43] Weiqin Zhao, Shujun Wang, Maximus Yeung, Tianye Niu, and Lequan Yu. Mulgt: Multi-task graph-transformer with task-aware knowledge injection and domain knowledge-driven pooling for whole slide image analysis. *arXiv preprint arXiv:2302.10574*, 2023. 3

[44] Yi Zheng, Rushin H Gindra, Emily J Green, Eric J Burks, Margrit Betke, Jennifer E Beane, and Vijaya B Kolachalama. A graph-transformer for whole slide image classification. *IEEE transactions on medical imaging*, 41(11):3003–3015, 2022. 2, 3, 5, 6

[45] Yanning Zhou, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir Rajpoot. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *ICCV*, pages 388–398, 2019. 2