



# Graph neural networks in histopathology: Emerging trends and future directions



Siemen Brussee <sup>a</sup>\*, Giorgio Buzzanca <sup>a</sup>, Anne M.R. Schrader <sup>a</sup>, Jesper Kers <sup>a,b</sup>

<sup>a</sup> Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, The Netherlands

<sup>b</sup> Amsterdam University Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

## ARTICLE INFO

Dataset link: <https://zenodo.org/records/1459598>

MSC:  
05C62  
05C90  
62M45  
92C55

**Keywords:**

Graph neural networks  
Computational pathology  
Graph representation learning  
Hierarchical graph representation learning  
Adaptive graph structure learning  
Multimodal graph representation learning  
Higher-order graph representation learning

## ABSTRACT

Histopathological analysis of whole slide images (WSIs) has seen a surge in the utilization of deep learning methods, particularly Convolutional Neural Networks (CNNs). However, CNNs often fail to capture the intricate spatial dependencies inherent in WSIs. Graph Neural Networks (GNNs) present a promising alternative, adept at directly modeling pairwise interactions and effectively discerning the topological tissue and cellular structures within WSIs. Recognizing the pressing need for deep learning techniques that harness the topological structure of WSIs, the application of GNNs in histopathology has experienced rapid growth. In this comprehensive review, we survey GNNs in histopathology, discuss their applications, and explore emerging trends that pave the way for future advancements in the field. We begin by elucidating the fundamentals of GNNs and their potential applications in histopathology. Leveraging quantitative literature analysis, we explore four emerging trends: *Hierarchical GNNs*, *Adaptive Graph Structure Learning*, *Multimodal GNNs*, and *Higher-order GNNs*. Through an in-depth exploration of these trends, we offer insights into the evolving landscape of GNNs in histopathological analysis. Based on our findings, we propose future directions to propel the field forward. Our analysis serves to guide researchers and practitioners towards innovative approaches and methodologies, fostering advancements in histopathological analysis through the lens of graph neural networks.

## 1. Introduction

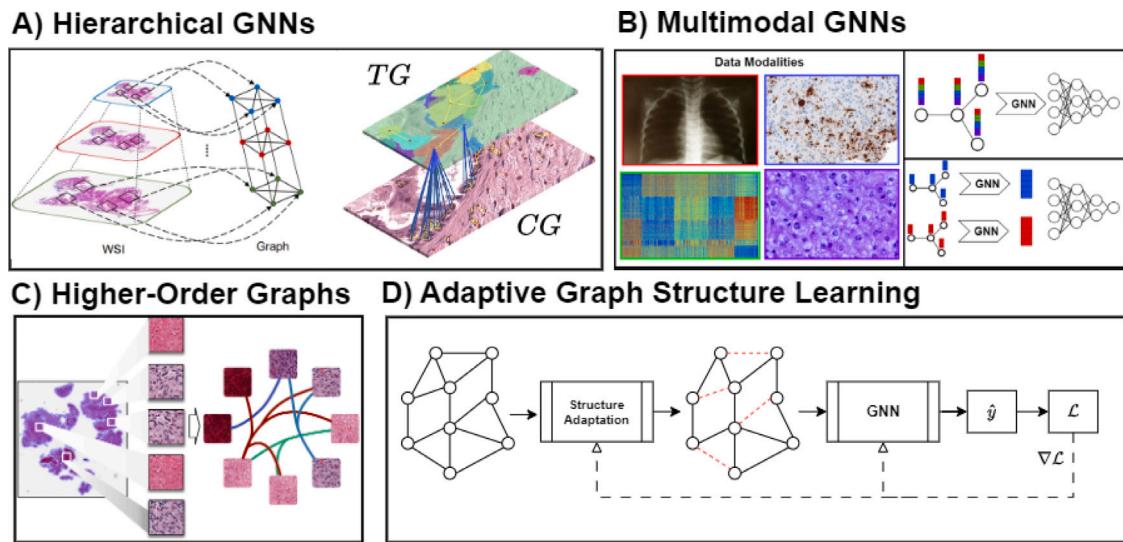
Histopathology analysis is an important examinatory tool that can be used for disease diagnosis, estimating disease prognosis, and monitoring therapeutic effects. Since the digitization of whole slide images (WSIs) in the early 2000s, the computational analysis of histopathology images has become an increasingly important part of histopathology. Starting with image analysis algorithms, the field moved to a deep learning approach after the rise of the convolutional neural network in the 2010s, which can be largely contributed to the availability of large datasets (e.g. ImageNet Deng et al., 2009) and deeper convolutional architectures (e.g. AlexNet Krizhevsky et al., 2012). In the last 5 years, paradigms in the field have become more heterogeneous, with the advent of attention-based multiple instance learning (Ilse et al., 2018; Sudharshan et al., 2019), vision transformers (Dosovitskiy et al., 2020; Wang et al., 2021a), self-supervised learning (Chen et al., 2020b; Ciga et al., 2022) and graph neural network (Scarselli et al., 2008; Li et al., 2018) approaches.

The emergence of Graph Neural Networks (GNNs) (Scarselli et al., 2008) has allowed effective modeling of naturally graph-structured

data, such as social networks, (bio)chemical molecules (Schütt et al., 2018; Li et al., 2021), geospatial data (Cui et al., 2019; Zhu et al., 2020), and tabular data which can be effectively modeled as a graph, such as in recommendation systems (Ying et al., 2018a) and drug interactions (Zitnik et al., 2018). GNNs are well suited for problems involving pairwise interactions between entities in data. Furthermore, the topological inductive bias that can be encoded in the graph structure allows GNN models to learn based on the topology of the problem. We can define the graph neural network as an optimizable transformation on all graph attributes that preserves graph symmetries by being permutation invariant. Fundamental for the graph neural network is the notion of *message-passing* in which we use a learned transformation that exchanges feature information between entities in the graph, leading to topology-aware feature vectors. How the message-passing function is defined depends on the type of GNN used, of which many varieties exist (e.g. GCN (Kipf and Welling, 2016), GAT (Veličković et al., 2017), GIN (Xu et al., 2018)). In 2018, Graph Neural Networks were also introduced to histopathology (Li et al., 2018) and have gained tremendous popularity in the field.

\* Corresponding author.

E-mail address: [s.brussee@lumc.nl](mailto:s.brussee@lumc.nl) (S. Brussee).



**Fig. 1.** Overview of the covered emerging subtopics of GNNs in Histopathology, covered in this review. Hierarchical GNNs, Multimodal GNNs, Higher-order Graphs, and Adaptive Graph Structure Learning (Mirabadi et al., 2024; Pati et al., 2020; Chen et al., 2020a; Di et al., 2022a; Zhu et al., 2021b).

While review papers on the application of graph neural networks exist, they either give a general overview (Ahmedt-Aristizabal et al., 2022) or focus on the clinical applications of GNNs in histopathology (Meng and Zou, 2023). Instead, we focus on identifying and quantifying emerging trends in the application of GNNs in histopathology and use them to provide future directions in the field.

Our review is organized into three main sections: First, we introduce graph theory, graph neural networks, and their applications in histopathology. Then, we identify emerging trends in the application of GNNs in histopathology and select some emerging paradigms from these trends, which are discussed in more depth (Fig. 1). Lastly, based on our findings, we provide future directions for the field.

## 2. Graph neural networks in histopathology

### 2.1. Graph neural networks

A graph  $G$  is defined as a set of nodes  $N$  connected by edges  $E$ :  $G = (V, E)$ . The set of edges is defined as a tuple of nodes:  $E = \{(x, y) | x, y \in V \text{ and } x \neq y\}$ . The connectivity of the nodes in a graph is captured in the adjacency matrix  $A^{n \times n}$ , where  $n$  is the number of nodes in  $G$ . Each entry  $a_{ij} \in A$  denotes the existence of an edge  $e_{ij} \in E$  as follows:

$$a_{ij} = \begin{cases} 1, & \text{if } e_{ij} \in E \\ 0, & \text{if } e_{ij} \notin E \end{cases} \quad (1)$$

Alternatively, the values of  $a_{ij}$  can denote edge weights ranging from 0 to 1, representing the connectivity strength between nodes  $i$  and  $j$ . Given an undirected graph  $G = (V, E)$ , we can define the  $k$ -neighborhood of any node  $v \in V$ , noted as  $N_k(v)$  recursively as follows:

$$N_0(v) = \{v\}, \quad (2)$$

$$N_1(v) = \{u \mid (v, u) \in E \text{ or } (u, v) \in E\}, \quad (3)$$

$$N_k(v) = \{u \mid \exists w \in N_{k-1}(v) \text{ such that } (w, u) \in E \text{ or } (u, w) \in E\}. \quad (4)$$

In Graph Neural Networks, we aggregate feature information from the  $k$ -neighborhood of each node, where  $k$  denotes the number of hops from the target node and also directly corresponds to the number of GNN layers used. This aggregated information is used to update the

node feature representation,  $h$ , in each GNN layer. Mathematically, the node representation update is defined as follows:

$$\begin{aligned} h_u^{k+1} &= \text{UPDATE}^{(k)}(h_u^{(k)}, \text{AGGREGATE}^{(k)}(\{h_v^{(k)} \mid \forall v \in N_k(u)\})) \\ &= \text{UPDATE}^{(k)}(h_u^{(k)}, m_{N(u)}^{(k)}) \end{aligned} \quad (5)$$

where UPDATE and AGGREGATION denote the functions that update node representation  $h_u$  and aggregate the hidden representations from  $u$ 's neighborhood  $N_k(u)$ , respectively. The exact definition of UPDATE and AGGREGATION functions is dependent on the message passing scheme used and is usually parameterized by two learnable weight matrices. However, all message-passing schemes employ a permutation-invariant AGGREGATION function (e.g., sum, mean). We can generally distinguish two types of message-passing schemes: *Spectral* message-passing, based on the spectral graph properties (e.g., eigenvalues) calculated using the graph Fourier transform, and *Spatial* message-passing, which are directly applied on the connectivity structure present in the input graphs. We focus on the spectral ChebNet method and the spatial GCN, GAT, GIN, and GraphSAGE methods, as these are applied in the vast majority of histopathology applications using GNNs. We first denote a tuple  $(G, A, X)$ , where  $G$  denotes the input graph,  $A$  is the associated adjacency matrix, and  $X$  the input node feature matrix. During message passing, we transform the feature matrix  $X$  into a hidden feature representation matrix  $H$ , typically using a learned weight matrix  $W$  and a non-linear activation function  $\sigma$ .

One of the most widely adopted and earliest graph neural network schemes is the Graph Convolutional Network (GCN) (Kipf and Welling, 2016). The message passing function uses a normalized adjacency matrix to update the hidden representations of nodes based on their neighborhood. To acquire the hidden representation matrix  $H$ , the message passing function in GCN layer  $l$  is defined as follows:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l) \quad (6)$$

in which  $\tilde{D}$  denotes the degree matrix of  $G$  and  $\tilde{A}$  represents the adjacency matrix with added self-loops for each node.

The Graph Attention Network (GAT) (Veličković et al., 2017) extends the graph convolutional network scheme by adding attention weights to each edge of the graph. This essentially allows models to learn the importance of nodes during message passing. For each edge  $e_{vu}$  connecting nodes  $v$  and  $u$ , we first calculate an attention score:

$$e_{vu} = \sigma(\vec{d}^T [W h_v^{(l)} \parallel W h_u^{(l)}]) \quad (7)$$

where  $\parallel$  denotes concatenation and  $\vec{a}$  is a trainable, shared parameter vector. Using this score, we can calculate the corresponding edge attention weight as follows:

$$\alpha_{vu} = \frac{\exp(e_{vu})}{\sum_{u' \in N(v)} \exp(e_{vu'})} \quad (8)$$

We then update the hidden node representation of  $h_v^{(l)} \in H^l$  as follows:

$$h_v^{(l+1)} = \sigma \left( \sum_{u \in N(v)} \alpha_{vu} \cdot W^l h_u^{(l)} \right) \quad (9)$$

GraphSAGE (Hamilton et al., 2017) provides a scalable and flexible framework to decide how neighboring nodes should be aggregated. It differs from other message-passing schemes in that it samples  $S$  neighbors in the neighborhood of each node, instead of using all neighbors. Given a hidden node representation  $h_v^{(l)} \in H^l$ , we can define its message-passing scheme as follows:

$$h_v^{(l+1)} = \sigma \left( \mathbf{W}^{(l)} \cdot \text{AGG}^{(l)} (\{h_u^{(l)} : u \in S_v\}) \right) \quad (10)$$

where  $\text{AGG}$  denotes an aggregation function at layer  $l$ , which can be any permutation invariant function (e.g., sum, mean).

Xu et al. introduced the Graph Isomorphism Network (GIN) (Xu et al., 2018), which has an expressive message-passing scheme to optimally differentiate between isomorphic graph structures. For any hidden node representation  $h_v^l \in H^l$ , the message-passing is defined as follows:

$$h_v^{(l+1)} = \text{MLP}^{(l)} \left( (1 + \epsilon^{(l)}) \cdot h_v^{(l)} + \sum_{u \in N(v)} h_u^{(l)} \right) \quad (11)$$

Here, the  $\text{MLP}$  denotes a multilayer perceptron which process each node's aggregated feature vector.  $\epsilon^l$  is a learnable parameter that learns how to scale the node's own feature vector.

The spectral ChebNet (Tang et al., 2019) method uses Chebyshev polynomials to approximate spectral graph convolution. First, we define the graph Laplacian as follows:  $L = D - A$  where  $D$  is the degree matrix and  $A$  is the adjacency matrix of the graph. We then rescale the graph Laplacian  $L$  using the largest eigenvector of  $L$ ,  $\lambda_{max}$ :  $\hat{L} = (2L/\lambda_{max}) - I$ . Given the approximation parameter  $k$ , we can compute the approximated Chebyshev polynomial  $Z^{(k)}$  as follows:

$$\begin{aligned} Z^{(1)} &= \mathbf{X} \\ Z^{(2)} &= \hat{L} \cdot \mathbf{X} \\ Z^{(k)} &= 2 \cdot \hat{L} \cdot Z^{(k-1)} - Z^{(k-2)} \end{aligned} \quad (12)$$

Finally, our message-passing function to update the hidden representation matrix in layer  $l$ ,  $H^l$ , is defined as follows:

$$H^{l+1} = \sum_{k=1}^K Z^{(k)} \cdot \mathbf{W}^{(l)} \quad (13)$$

Prediction tasks using GNNs can be categorized into node-level, edge-level, and graph-level prediction tasks. Node-level tasks, such as node classification, predict labels of target nodes based on the transformed representations after message passing. Edge-level tasks include edge classification, where edge labels are predicted, and link prediction. In link prediction, the aim is to predict whether links between nodes should exist based on the node features after message passing. Lastly, graph-level tasks that predict graph-level labels. These tasks rely on a global *pooling* step, which aggregates information from node and/or edge level into a global representation. Let us define a graph  $G = (V, E)$  with an associated node feature matrix  $X$ . We can then use any permutation-invariant function to pool the node features into a global representation:

$$\text{pool}(G) = \bigoplus_{v \in V} X(v) \quad (14)$$

where  $\bigoplus$  is any permutation invariant function (e.g. sum).

## 2.2. GNNs in histopathology

Graphs have been used in digital pathology since the 1990s (Sharma et al., 2015) and have later been combined with classical machine learning algorithms (e.g., SVMs) for diagnosis tasks (Bilgin et al., 2007). Since then, Graph Neural Networks have been gaining popularity throughout the 2010s to become the primary method for graph-based machine learning tasks. Since the first application of GNNs in histopathology, in 2018 (Li et al., 2018), the use of GNNs in histopathology has grown rapidly to more than 200 publications in 2024. Applying GNN to histopathology requires some decision-making and algorithmic steps (Fig. 4). First, we preprocess the WSI (e.g., quality control, stain normalization). Now, either a cell segmentation algorithm can be applied from which a cell graph can be constructed, or one extracts patches from which a patch graph can be constructed. Using the extracted image entities, a graph can be defined using a chosen graph construction algorithm. This graph can then be used as input for a GNN model. The predictions given by the GNN model can be explained using various GNN explainability methods. We will further explore this typical workflow of GNNs in histopathology in the following sections.

### 2.2.1. Defining the input graph

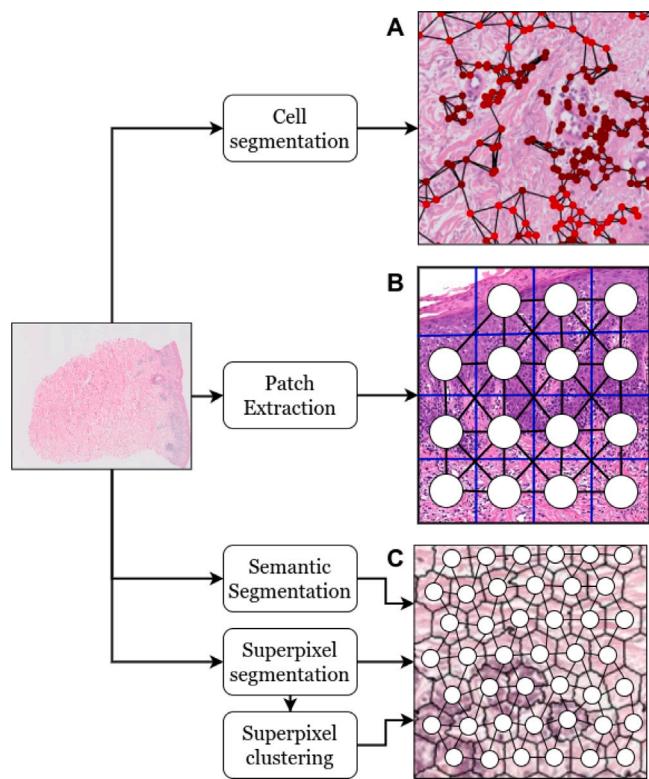
For GNNs to be applied to histopathology images, one first needs to define what entities the nodes in the input graph will represent. Most GNNs applied to histopathology use one of 3 types of input graphs, as shown in Fig. 2: *Cell Graphs* where nodes represent cells or nuclei, segmented using a segmentation algorithm or model (for example, HoVerNet). *Patch Graphs*, where nodes represent patches of the image, and lastly, *Tissue Graphs* where nodes represent larger-scale semantic entities in the image. These tissue graph entities can be acquired from a semantic segmentation map, superpixels (usually generated using the SLIC algorithm Achanta et al., 2012), or clustered superpixels, which represent similar regions in the input image. Some alternate approaches also exist: Notably, approaches that treat image pixels as nodes and approaches that construct a patch-based hypergraph.<sup>1</sup>

This flexibility in defining the notion of a local neighborhood is what sets GNNs apart from other model architectures like CNNs and Transformers, which can be viewed as specific cases of a GNN where the graph structure is fixed. CNNs, for example, can be viewed as a special case of GNNs, where the underlying graph is a regular grid and each node (pixel) is connected to its immediate neighbors, defined by the convolutional kernel. In contrast, Transformers can be seen as a GNN operating on a fully connected graph, where every input token or feature (node) is connected to all others.

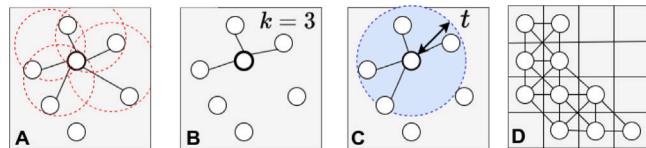
Once the nodes have been established, one needs to decide how the nodes should be connected. For this, histopathology GNNs usually apply one of four graph construction strategies or combinations of these strategies (as shown in Fig. 3). First, we can use a simple *distance threshold*, where for each node, we connect it to all other nodes having a distance (e.g. Euclidean) from the target node less than a set threshold  $t$ . Second, we can use the *k-Nearest Neighbor* (k-NN) algorithm. Here, we set a parameter  $k$ , which denotes how many neighbors each node will have. Then, we connect the  $k$  closest neighbors of each node to the target node. Note that for both approaches, we can base our notion of distance on spatial distance or distance between the node-associated feature vectors. Third, we can construct a *Region Adjacency Graph* (RAG), where we connect all entities that share a border.<sup>2</sup> Typically, this approach is used for patch or tissue graphs, where there is a clear border between entities. Lastly, we can use *Delaunay triangulation*. Here, we define the graph edges as all possible triangles between the nodes, such that the circumcircle of each triangle does not contain other nodes than the 3 nodes the triangle consists of.

<sup>1</sup> Graph where edges can connect any number of nodes instead of the pairwise edges seen in regular graphs.

<sup>2</sup> In patch graphs, this is equivalent to using a  $k = 4$  k-NN without diagonal neighbors and  $k = 8$  k-NN with diagonal neighbors.



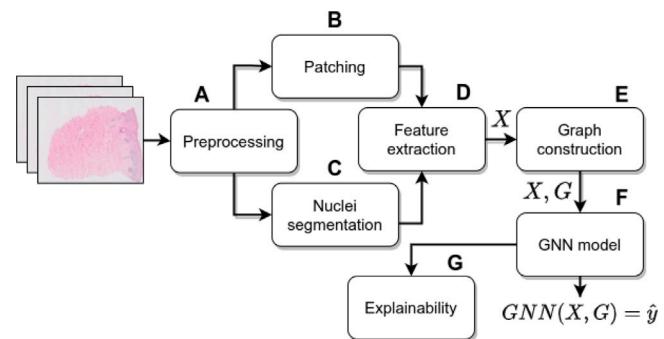
**Fig. 2.** Most widely used graph types in GNNs for histopathology. (A) Cell Graph, (B) Patch Graph, (C) Tissue graph (based on superpixels, clustered superpixels, or a semantic segmentation mask). The superpixel image was acquired from [Bejnordi et al., 2015](#).



**Fig. 3.** Most widely used graph construction techniques in GNNs for histopathology. (A) Delaunay triangulation, (B) K-NN with  $k = 3$ , (C) Distance threshold with threshold  $t$ , (D) RAG with diagonal neighbors ( $k = 8$ ).

### 2.2.2. Extracting features

To allow a GNN to use the image-based features present in whole slide images, one usually extracts features associated with the entity of the node and attaches that to the node as node features. Similarly, features can also be added to the graph edges, which the GNN can use in the message-passing function. Backbones of pretrained<sup>3</sup> CNN (e.g. ResNet [He et al., 2016](#)) or Vision Transformer ([Dosovitskiy et al., 2020](#)) models are primarily used for node feature extraction, where we use an image patch corresponding to the node entity, process it using the feature extraction model, and extract the feature vectors of this image in the intermediate layers of the model as node features. Sometimes, the feature extraction model is trained in a supervised manner on the histopathology images for the problem at hand, or fine-tuned for the prediction task at hand, which allows for more problem-specific features. More recently, self-supervised training has been applied for feature extraction, providing model features that generalize better across prediction tasks ([Tendle and Hasan, 2021](#)). Handcrafted features, based on morphology, texture, or intensity measurements, can also be used as node features. Furthermore, (spatial)



**Fig. 4.** Overview of a typical workflow of applying GNNs to histopathology whole slide images. (A) First, preprocessing steps such as slide quality thresholds and tissue segmentation (e.g., using Otsu thresholding) are applied. (B) Then, for patch graph approaches, the WSI images are divided into smaller image patches. (C) When a cell graph approach is used, nuclei segmentation algorithms are applied to acquire a mask of the nuclei in the WSI. (D) For each acquired entity (patch, nucleus) features are extracted, typically using a pretrained CNN model (e.g., ResNet) to acquire a feature matrix  $X$ . (E) Using a graph construction strategy (e.g., k-NN), entities are connected to other entities to form a cell/patch graph,  $G$ . (F) The resulting graph, along with its associated feature matrix, can be used as input for a GNN model applying message-passing operations to learn a representation and provide predictions. (G) (Graph) explainability methods can be applied to the GNN model to acquire interpretable information on the model behavior, and its predictions.

graph features (e.g. node degree) can be calculated on a node, edge, or graph level to more directly incorporate topological information in the model.

### 2.2.3. Graph neural network architectures

Most message-passing schemes used in histopathology GNNs are not specific to histopathology. Popular schemes used include Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), GraphSAGE, or GINs (Graph Isomorphism Networks). Some approaches invented schemes specific for their problem ([Gao et al., 2021](#); [Zhang et al., 2022](#); [Hou et al., 2022b](#); [Hasegawa et al., 2023](#); [Nakhli et al., 2023](#); [Wang et al., 2023b](#)) and more recently, Graph Transformers models have gained traction as a popular alternative or addition to regular message passing. In particular, many approaches combine message-passing layers with other neural network modules, such as transformers, LSTMs, MLPs, and MIL aggregation layers. For graph-level prediction problems, global pooling layers are applied, sometimes combined with sequentially applied local pooling layers, which hierarchically coarsen the graph.

### 2.2.4. Applications

GNNs in histopathology have been applied to a wide variety of tasks, primarily on supervised prediction tasks such as survival prediction, region-of-interest (ROI) classification, cancer grading, cancer subtyping, cell classification, and the prediction of treatment response. Some applications aim to predict data in other modalities, such as genetic mutations or (spatial) gene expression. Although most of the use cases are classification problems, some research has used GNNs for semantic segmentation ([Anklin et al., 2021](#); [Zhang et al., 2021](#); [He et al., 2023](#)) or nuclei detection ([Bahade et al., 2023](#); [Wang et al., 2023c](#)). Generative approaches also exist, such as H&E image-based text generation or IHC staining generation. Another interesting application is *Content-Based Histopathological Image Retrieval* (CBHIR) ([Zheng et al., 2019, 2020](#)). Here, we first use GNNs to extract and save a graph representation of an ROI in a WSI. When pathologists grade new cases, we can use these embeddings to retrieve similar ROIs, helping in the diagnostic process. Most GNN applications focus on cancer as a disease, with a few exceptions ([Wojciechowska et al., 2021](#); [Nair et al., 2022](#); [Hasegawa et al., 2023](#); [Gallagher-Syed et al., 2023](#); [Lee et al., 2023](#); [Su et al., 2023](#); [Acharya et al., 2024](#)).

<sup>3</sup> Usually on the ImageNet dataset ([Deng et al., 2009](#)).

### 2.2.5. Explainability

One major advantage GNNs have over other model types in histopathology is interpretability. The model output can be explained on an entity level and visualized using a graph overlay. For example, one can pool nodes in a cell graph using an attention mechanism, calculate the attention scores for each node, assign a color based on the attention score per node, and then visualize the attention scores on a cellular level as a graph overlay on the WSI. Many methods for GNN explainability have emerged since the inception of the GNN (e.g., GNNExplainer (Ying et al., 2019), GCExplainer (Magister et al., 2021)). There have also been efforts to develop histopathology-specific GNN explainability methods (Jaume et al., 2020; Yu et al., 2021; Abdous et al., 2023) or to use combinations of existing GNN explainability techniques to extract a clinically interpretable model output (di Villaforesta et al., 2023).

### 2.2.6. Advantages of GNNs for histopathology

Having explained how GNNs can be used in histopathology, it is important to point out why GNNs are used in histopathology. We highlight several important advantages of GNN-based modeling in histopathology compared to other model types (e.g., CNNs, Transformers):

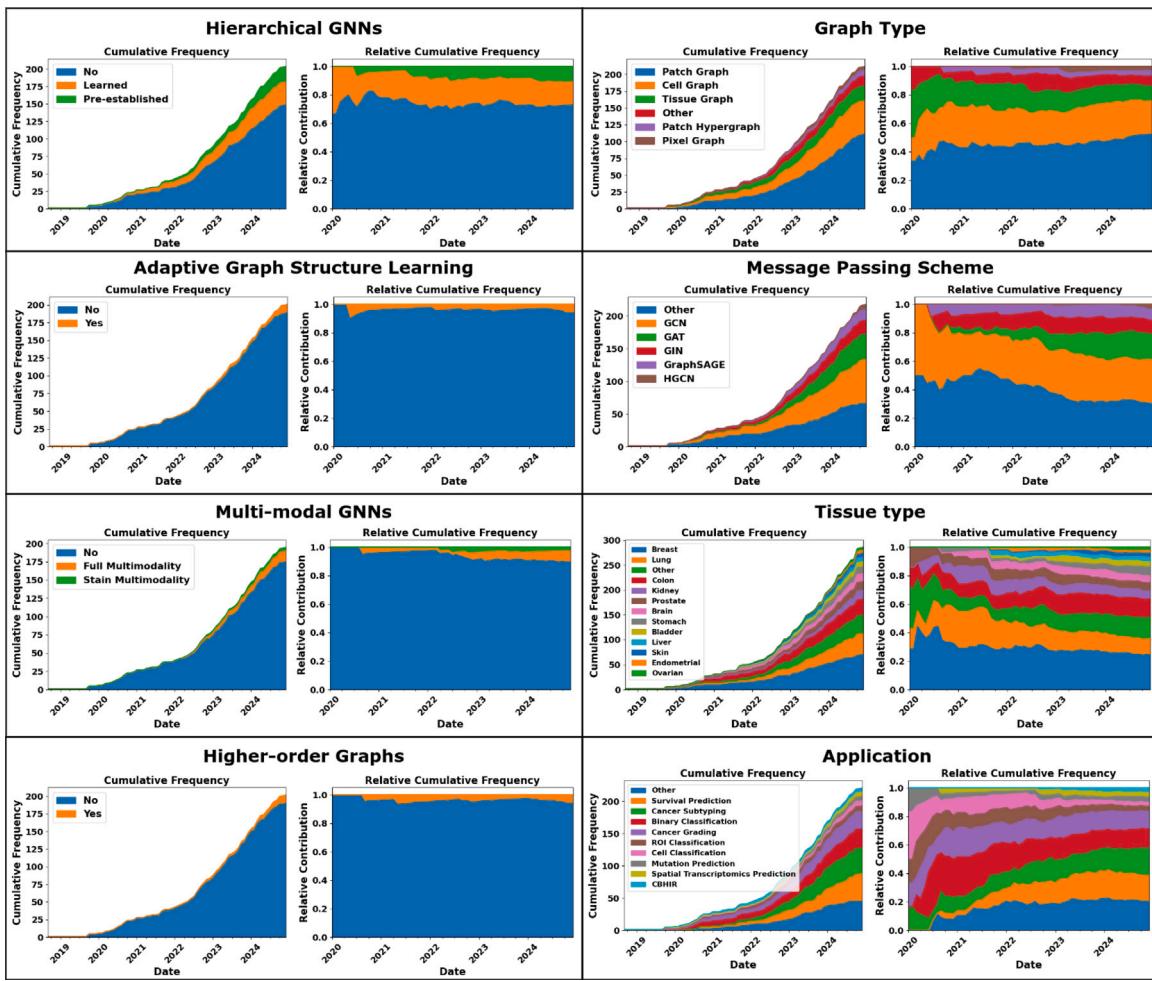
1. *GNNs acquire relationship-aware representations:* By exchanging information between nodes on the input graph, GNNs learn context-aware representations. This is important in pathology, where the importance of certain biological entities often depends on the cellular or regional context (Santoiemma and Powell, 2015).
2. *GNNs can capture the topological information in the WSI:* Graphs are a natural way to capture topology. In histopathology, factors such as cellular density can be important in diagnosis, which can be captured through topological information in the graph structure (Ali et al., 2013; Reynolds et al., 2014).
3. *GNNs can model the entire WSI at once:* Due to the sheer size of whole slide images, traditional deep learning methods usually split the WSI into patches, which are pooled to obtain a global slide representation. GNNs can instead model the entire WSI as a single graph, effectively capturing the global structure of the WSI without having to reconstruct this from a combination of patches (Adnan et al., 2020).
4. *GNNs allow for hierarchical modeling:* In pathology image analysis, diagnosis often relies on information acquired from multiple spatial scales of the WSI (e.g., global patterns combined with atypical cell structures). GNNs allow modeling both of these scales in a single model, either by connecting graphs on different scales or by learning the global structure through pooling operations (Pati et al., 2020; Zheng et al., 2019).
5. *GNNs allow for entity-wise interpretability:* Whereas CNN-based methods usually rely on pixel-level explainability, GNNs allow for entity-wise explainability. This allows pathologists to investigate the dependence of the model prediction on certain biological entities, such as cells or substructures, in the WSI (Sureka et al., 2020).
6. *GNNs allow for injecting task-specific inductive biases:* The input graph structure can be modified based on prior information about the task at hand. This, in turn, allows for more specific explainability and efficient modeling of the problem (Hasegawa et al., 2023).
7. *GNNs enable the efficient fusion and joint modeling of spatially-coordinated multimodal data:* In cases where multimodal information is spatially aligned (e.g., IHC staining w.r.t. H&E staining), GNNs can combine the information from the same location as a single feature vector attached to a graph entity (e.g., node, edge). The multimodal features are then jointly modeled in the message-passing layers, allowing for spatially-aware fusion of the modality information. The graph structure provides a spatial

inductive bias for modality fusion, which allows a precise definition of the spatial context that should be modeled. This approach is harder with other model types, where modality fusion is either defined on a very local (e.g., CNNs) or a very global scale (e.g., Transformers). Furthermore, this approach is efficient as no additional model modules are required for the injection of multimodal information into the GNN (Ektefaie et al., 2023; Li et al., 2022).

### 2.2.7. Challenges of applying GNNs to histopathology

Despite the advantages discussed, one should also be aware of the challenges specific to the application of GNNs in histopathology. We discuss these challenges and provide suggestions on how to overcome them.

1. *Graph structure definition:* Constructing a graph that adequately captures the phenomenon of interest using a WSI can be non-trivial. Since WSIs alone cannot accurately decide whether tissue elements (e.g., cells) are biologically interacting, graphs are usually constructed somewhat arbitrarily based on spatial proximity alone. This can be a problem, since the input graph has a large effect on the patterns learned by the GNN and thus on the downstream performance (Thang et al., 2022). As such, careful graph construction is vital to adequately capture the information required for the downstream task.
2. *Computational Complexity:* When constructing graphs based on nuclei, the number of nodes in the graph can grow very large, limiting the number of graphs that can fit into GPU memory during training. This can lead to slow model training and potentially introduce numerous nodes without predictive value to the model. To mitigate this, it is often helpful to select a subset of nodes from the graph (e.g., using random or node-label-based sampling) or only define a graph at a specified region of interest in the WSI. Another solution would be to use scalable sampling-based GNN methods (e.g., GraphSAGE) (Hamilton et al., 2017), removing the need to load the entire graph into memory at once.
3. *Dependence on segmentation output:* When defining an input graph, one often relies on the output of a segmentation model to define the graph entities (e.g., cell segmentation, semantic segmentation). Consequently, the quality of the graph will be largely dependent on the quality of the provided segmentation. Therefore, it is vital to assess whether the segmentation quality is sufficient for the downstream task. For example, missing a subset of nuclei can be detrimental in tasks involving inflammatory patterns, where a small subset of cells can characterize the state of the tissue. In contrast, in most cancerous tissues, this problem is less severe, as missing cells can be compensated for by an overwhelming number of cancerous cells detected during segmentation.
4. *Overreliance on nuclei-based features:* Since cell graphs are traditionally based on nuclei segmentations, one should carefully consider how to integrate information on the surrounding cytoplasm and non-cellular features in the cellular neighborhood. One way to accomplish this is by extracting features from a patch centered on the nucleus, taking into account the surrounding area. Recent efforts propose another solution by segmenting the membrane as well as the nucleus (Gu et al., 2024), allowing cytoplasm-specific features to be measured.
5. *Obtaining accurate fine-grained labels:* For tasks relying on fine-grained labels (e.g., cell classification using GNN), many annotations are required per WSI, which is very labor-intensive if manually labeled. Automatic cell classification pipelines mitigate this problem (Graham et al., 2019), but can often be inaccurate, leading to incorrect node labels, especially in tissue types not seen in the training data. Similarly to the point above, we suggest carefully inspecting the quality of the labels obtained and assessing whether it is adequate for the downstream task.



**Fig. 5.** Cumulative and relative cumulative frequency of publications on GNNs applied to histopathology. The left four panels show the usage of the emerging trends we explore further in this review. The right four panels depict more general characteristics of the publications, such as graph type and application.

### 3. Methodology

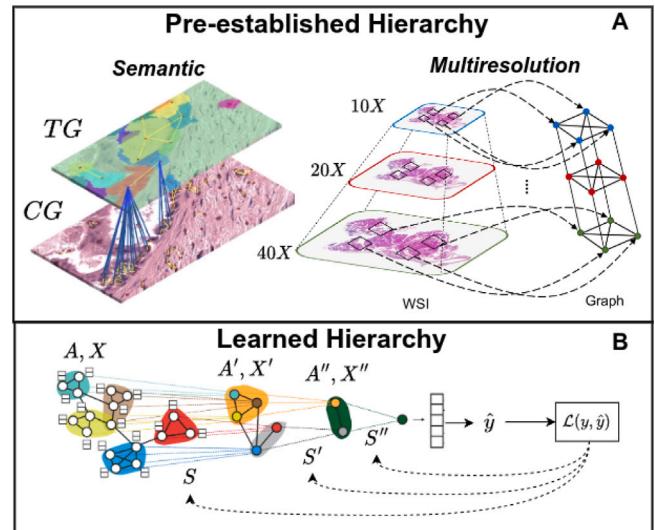
Using Google Scholar, we identified 204 papers applying GNNs to histopathology, dating from September 2018 up to November 2024. We included all papers applied on H&E stained whole slide images or tissue microarrays (TMAs) where GNNs (i.e. message passing) were part of the methodology. The papers were categorized based on the following properties:

- Message-passing scheme
- Type(s) of input graph
- Application(s)
- Tissue type(s)
- Uses hierarchy
- Uses adaptive graph structure learning
- Uses multimodality
- Uses higher-order graphs

We quantified the frequencies in each of these properties to identify emerging trends in the literature (Fig. 5).

From our quantification, we identified upcoming trends and selected four to explore more deeply:

1. Hierarchical GNNs
2. Adaptive Graph Structure Learning
3. Multimodal GNNs
4. Higher-order graphs



**Fig. 6.** (A) *Pre-established hierarchy*, where different graphs are constructed at different semantic- or magnification levels, which are connected hierarchically (e.g. using an assignment matrix) (Pati et al., 2020; Mirabadi et al., 2024) (B) *Learned Hierarchy*, where trainable local pooling operations sequentially coarsen the graph structure (Ying et al., 2018a).

The remaining trends we identified are discussed in a more concise format in the *Other trends* section.

## 4. Emerging trends

### 4.1. Hierarchical GNNs

Diagnostic- and prognostic information present on WSIs often exists on multiple levels of coarseness. For example, the cellular microenvironment can be an important diagnostic factor but can depend on where this microenvironment is globally located in the tissue. Cellular graphs are suitable for capturing the microenvironment but can miss the global tissue information present in the WSI. Similarly, patch- or tissue-based graphs can capture global information in the WSI, but miss the topological information present in the cellular structures (Pati et al., 2022). To connect information on different levels of coarseness, we can either apply local pooling layers which learn a hierarchical representation of the input graph in an end-to-end manner, which we denote as *Learned Hierarchy*, or, we can define the hierarchy between graphs prior to model training: which we denote as *Pre-established hierarchy*. Both are illustrated in Fig. 6.

In a learned hierarchy, we apply *local pooling* layers that iteratively coarsen the graph structure hierarchically. Let us define our input graph with associated node features as  $G_0 = (V_0, E_0, X_0)$ . Assuming that we have  $k$  local pooling layers in our GNN architecture, we sequentially coarsen our input graph to  $G_1, G_2, \dots, G_k$  where  $G_k$  is the final pooled graph representation. Mathematically, we define a local pooling to coarsen the graph  $G_i$  to  $G_{i+1}$  as follows:

$$G_{i+1} = \text{pool}_i(G_i), \quad \forall i \in \{0, 1, 2, \dots, k-1\} \quad (15)$$

where  $\text{pool}_i$  is defined by any permutation-invariant pooling function. Prominently used examples include DiffPool (Ying et al., 2018a), SAGPool (Lee et al., 2019), and MinCutPool (Bianchi et al., 2020). Apart from local pooling, we also classify methods that learn the hierarchy using a cross-hierarchical transformer (Hou et al., 2022a; Shi et al., 2023a; Azadi et al., 2023) layer as learned hierarchy methods.

Learned hierarchy methods learn a node assignment matrix  $S^{(l)}$  which denotes the changes in the graph structure after applying the pooling operation. Often, multiple local pooling layers are applied subsequently to coarsen the graph. One sets a *pooling ratio* hyperparameter, denoted as  $k$ , which determines how many nodes should be present after the pooling operation. For a pooling layer,  $l$ , the pooling operation updates the adjacency matrix of the input graph,  $A$ , and its corresponding node attributes  $X$ . The hidden representations are denoted  $H$ , where  $X = H^0$ . We denote the pooling operation as:

$$(A^{l+1}, H^{l+1}) = \text{POOL}(A^{(l)}, H^l) \quad (16)$$

The exact pooling operation is dependent on the pooling function used. DiffPool (Ying et al., 2018b) applies GNNs to learn a differentiable cluster assignment matrix which maps nodes to clusters, which are then used as the new nodes after pooling. DiffPool uses two GNNs: One for obtaining node embeddings,  $GNN_{l,\text{embed}}$ , and one for assigning the nodes to cluster nodes:  $GNN_{l,\text{pool}}$ . In each DiffPool layer  $l$ , we use the embedding GNN for extracting a feature matrix  $Z$ :

$$Z^l = GNN_{l,\text{embed}}(A^l, H^l) \quad (17)$$

Then, we calculate the assignment matrix using the pooling GNN:

$$S^l = \text{softmax}(GNN_{l,\text{pool}}(A^l, H^l)) \quad (18)$$

Now, we update both the hidden node representations and a new adjacency matrix:

$$H^{l+1} = S^{lT} Z^l \quad (19)$$

$$A^{l+1} = S^{lT} A^l S^l$$

Self-Attention Graph Pooling (SAGPool) (Lee et al., 2019) uses the self-attention mechanism to learn which nodes are important and to discard unimportant ones. First, we calculate the self-attention score using a graph convolution operation:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l) \quad (20)$$

Here,  $W^l$  is a learned weight matrix that we use to calculate the attention score. For each node  $v \in V$ , we calculate:

$$a_i^l = \text{softmax}(W^l h_i^l) \quad (21)$$

where  $h_i$  is the feature embedding of  $v_i$ . SAGPool then ranks the nodes on their attention scores and selects the top  $k$  nodes to retain. These  $k$  nodes are used to mask the adjacency matrix to obtain  $H_{\text{mask}}$ , which gets multiplied with the original adjacency matrix to coarsen the graph:  $A^{l+1} = A \odot H_{\text{mask}}$ .

Lastly, MinCutPool (Bianchi et al., 2020) uses the mincut partition objective function to decide the assignment matrix  $S$ . Similarly to the DiffPool method, we first generate a GNN-based node feature matrix  $H^{l+1}$ :

$$H^{l+1} = GNN(H^l, A^l, W_{GNN}^l) \quad (22)$$

where  $W_{GNN}^l$  is the learned weight matrix of the GNN. Using the updated representation, we can use a multilayer perceptron (MLP) to calculate the node assignment matrix  $S$ :

$$S = \text{MLP}(H^{l+1}, W_{MLP}^l) \quad (23)$$

where  $W_{MLP}^l$  are the learned weights of the MLP. Both  $W_{GNN}^l$  and  $W_{MLP}^l$  are trained by minimizing two loss terms  $L_c$ , denoting the cut loss term, and  $L_o$ , denoting the orthogonality loss term. The cut loss term approximates the MinCut objective, by aiming to minimize the number of edges between clusters while maximizing the edges within clusters. The orthogonality loss term encourages orthogonal cluster assignments and similarly sized clusters. Together, these loss functions form the objective loss  $L_u$ :

$$L_u = L_c + L_o = -\frac{\text{Tr}(S^T \tilde{A} S)}{\text{Tr}(S^T D \tilde{S})} + \frac{\text{Tr}(S^T S - I K)}{\sqrt{K}} \quad (24)$$

where  $D$  is the degree matrix of the normalized adjacency matrix  $\tilde{A}$ ,  $I$  is the identity matrix and  $K$  is the number of desired clusters.

The pooling operation is performed as follows:

$$\begin{aligned} A^{l+1} &= S^T \tilde{A} S \\ H^{l+1} &= S^T H \end{aligned} \quad (25)$$

#### 4.1.1. Learned hierarchy

As Table 1 shows, the vast majority of GNN applications in histopathology use existing local pooling functions such as in the examples above. In this section, we give some examples of newly designed learned hierarchy methods, specifically for problems in histopathology.

**Local Pooling:** Hou et al. proposed *Iterative Hierarchical Pooling* (IH-Pool), which they combined with a pre-established hierarchy (Hou et al., 2022b). As input, the authors used a pyramidal heterogeneous patch graph, with one graph existing on 10x resolution, one on 5x resolution, and one on thumbnail resolution. IHPool was designed to filter redundant information for the downstream prediction task while retaining this pyramidal structure when applying the pooling operation. The method achieves this by conditioning the set of nodes to be pooled on each resolution level on the pooling outcome of the lower-resolution nodes. Let  $X$  be a matrix of node features,  $A$  be the adjacency matrix of the input graph,  $k$  be the ratio of nodes to retain after pooling and  $P$  be a learnable projection layer. Now, let us denote the input graph  $G = (V, E, R)$  where  $R$  represents the set of different resolutions in the graph. For each resolution  $r \in R$ , nodes are pooled hierarchically, such that nodes in higher magnification levels are subordinate to nodes in lower levels. Thumbnail nodes are individually assigned, while other

**Table 1**  
Publications applying GNNs to histopathology which used learned hierarchies.

Publication	Date	Application	Learned hierarchy method
Zheng et al. (2019)	2019/10	CBHIR	DiffPool
Zhou et al. (2019)	2019/10	Cancer grading	DiffPool
Sureka et al. (2020)	2020/10	Binary classification	DiffPool
Zheng et al. (2020)	2020/12	CBHIR	DiffPool
Chen et al. (2020a)	2020/09	Survival prediction, Cancer grading	SAGPool
Jiang et al. (2021)	2021/01	Cancer grading	DiffPool
Zheng et al. (2021)	2021/04	CBHIR	DiffPool
Wang et al. (2021b)	2021/09	Survival prediction	SAGpool
Xiang and Wu (2021)	2021/10	Binary classification	DiffPool
Xie et al. (2022a)	2022/01	Treatment response prediction	TopKPooling
Dwivedi et al. (2022)	2022/04	Cancer grading	SAGPool
Hou et al. (2022b)	2022/06	Binary classification	IHPool
Bai et al. (2022)	2022/08	Cancer subtyping	MinCutPool
Zuo et al. (2022)	2022/09	Survival prediction	SAGPool
Hou et al. (2022a)	2022/09	Cancer subtyping	Hierarchical attention mechanism
Lim and Jung (2022)	2022/10	Survival prediction	SAGPool
Wang et al. (2023a)	2023/02	Cancer subtyping	Scattering Cell Pooling
Zhao et al. (2023)	2023/02	Cancer subtyping, Cancer grading	GCMinCut
Ding et al. (2023b)	2023/02	Cancer subtyping, Cancer grading	Fractal paths
Ding et al. (2023a)	2023/04	Survival prediction	SAGPool
Li et al. (2023)	2023/09	Node classification	Graph V-Net
Gallagher-Syed et al. (2023)	2023/09	Rheuma subtyping	SAGPool
Shi et al. (2023a)	2023/09	Cancer subtyping, mutation prediction	Hierarchical attention mechanism
Wu et al. (2023)	2023/10	Survival prediction	SAGPool
Nakhl et al. (2023)	2023/10	Survival prediction	SAGPool
Azadi et al. (2023)	2023/10	Survival prediction	MinCutPool, Hierarchical attention mechanism
Hou et al. (2023)	2023/10	Survival prediction	Matrix multiplication
Abbas et al. (2023)	2023/12	Cancer grading	DiffPool
Xu et al. (2023)	2023/12	Cancer subtyping	DiffPool
Azher et al. (2023)	2024/01	Cancer grading, Survival prediction	SAGPool
Yang et al. (2024)	2024/03	Binary classification, Survival prediction	MinCutPool
Breen et al. (2024)	2024/07	Cancer subtyping	SAGPool
Gindra et al. (2024)	2024/07	Cancer subtyping, binary classification	Multihead cross-attention pooling

nodes are pooled iteratively. For all nodes, a fitness score is calculated and nodes are assigned to clusters based on spatial distance and fitness difference between nodes. Specifically, for each node  $n \in N$  on resolution  $r$ , we use a learnable projection matrix  $P$  to calculate the fitness set as follows:

$$\phi_n^r = \tanh\left(\frac{V_n^r \cdot P}{\|P\|}\right) \quad (26)$$

where  $V_n^r$  is the set of nodes to be pooled, based on the hierarchical edges between resolutions. Based on the calculated node assignments, we create a new node feature matrix  $X'$ . The adjacency matrix  $A'$  is updated to maintain graph connectivity based on the node assignments.

Wang et al. proposed a new module for pooling information from cell graphs to use as embeddings for clusters of cells, called *cell community forests* (Wang et al., 2023a). Their method incorporates two parallel pooling operations: a max pool and an average pool, whose results are concatenated and processed using a linear layer. The authors applied DBSCAN clustering to cell embeddings where they applied different density values  $d$  to acquire a vector of densities  $d_1, \dots, d_N$  with which they acquired different clusters  $j \in J$ . Mathematically, given a cell embedding in cluster  $j$  with clustering density  $d_i$ ,  $C_{i,j}^x, x \in S_{i,j}$ , where  $S_{i,j}$  denotes the set of cells with density  $d_i$  in cluster  $j$ , we calculate the pooling outcome  $P_{ij}$  for cluster  $j$  with density  $d_i$  as follows:

$$P_{ij} = W^0 \times (\text{Max}(C_{i,j}^x) \| \text{Mean}(C_{i,j}^x)) + b^0 \quad (27)$$

where  $W^0$  and  $b^0$  are learnable parameters.

Zhao et al. proposed an extension of the popular MinCutPool by adding a message-passing layer in the pooling equation (Zhao et al., 2023). For acquiring the cluster assignment matrix  $S$ , where each node  $s \in S$  will be a single node in the coarsened graph, the authors used the following equation:

$$S = H(\sigma(\hat{A}H W_{pool})) \quad (28)$$

where  $\hat{A}$  is the LaPlacian-normalized adjacency matrix,  $H$  denotes the hidden representation matrix of the nodes,  $W_{pool}$  denotes a learnable

pooling weight matrix and  $\sigma$  denotes a nonlinear activation function (e.g. ReLU).

**Attention-based Interaction Modeling:** Azadi et al. proposed 2 attention-based methods for exchanging information between different levels of graph coarseness (Azadi et al., 2023). The authors used a local graph, where nodes represent patches in the WSI, and a global graph, where nodes represent MinCutPool-based clusters of nodes in the local graph. Now, attention scores are calculated for each node in the local- and global graph. The first method the authors proposed for exchanging information between the local- and global graph was *Mixed Co-Attention* (MCA), in which the information is not mixed directly, but weight sharing is applied between parallel processing of the local- and global nodes. The second method, *Mixed Guided Attention*, we expanded MCA by directly infusing the calculated local node feature representation into the attention score calculation of the global nodes. The authors found that the mixed co-attention strategy worked optimally for their use case.

Starting with a 20x resolution patch graph for each WSI, Gindra et al. first process each nodes using multiple stacked GIN-layers (Gindra et al., 2024). The authors then learned node cluster assignments using multihead cross-attention pooling. Each cluster was used as input to a transformer block whose output was processed using a multilayer perceptron to obtain a slide-level prediction.

**Alternative Approaches:** Ding et al. did not learn a hierarchical representation using pooling layers, instead using a *FractalNet* architecture (Ding et al., 2023b). Here, the input graph is given to separate processing paths which consist of different numbers of GNN layers, thereby representing different semantic levels in the tissue. The hierarchy between the paths is encoded using a combination of a gated bimodal unit and an MLP mixer architecture. The former calculates a weighted combination of representations, while the latter enhances communication between the path representations and strengthens connections among different path features.

Li et al. propose a hierarchical Graph V-Net to encode hierarchy in a patch graph input (Li et al., 2023). First, attention-based message-passing is used to exchange information between adjacent patches. Then, the authors used a graph coarsening operation where the node features are arranged as a 2D grid based on the spatial location of the patches. This grid is then evenly divided into submatrices and each submatrix is projected to a single feature vector using a learnable layer, which will act as a node after the coarsening operation. Notably, the Graph V-Net also uses graph upsampling layers, which add nodes until the size of the input graph has been restored, similar to what is done in UNet-architectures.

#### 4.1.2. Pre-established hierarchy

In pre-established hierarchy, we encode the hierarchy prior to model training. For example, we can construct multiple graphs at different levels of coarseness in the WSI, and connect them using assignment matrices, denoting how nodes are connected between the hierarchical levels. Often, during message-passing, the learned representations of the lower hierarchy level are aggregated and used as input for the corresponding nodes at the higher hierarchy level. We differentiate between approaches connecting graphs on different semantic levels (e.g. cells and tissues), and approaches connecting different magnifications of the WSI (e.g. 40x, 20x). An overview of publications using this approach is given in Table 2.

**Semantic Hierarchies:** Pati et al. were the first to introduce a pre-established hierarchy in the graph to use as input for a GNN model (Pati et al., 2020). They constructed a cell graph,  $CG$ , using a nuclei segmentation map and a tissue graph,  $TG$ , constructed by clustering superpixels based on similarity. To introduce the hierarchy, they introduced an assignment matrix  $S_{CG \rightarrow TG}$ , such that  $S_{CG \rightarrow TG}(i, j) = 1$  if a cellular node  $i$  from the cell graph belongs to tissue node  $j$  in the tissue graph.

Wang et al. introduced hierarchy by applying separate message-passing operations on both a cell graph and a patch graph (Wang et al., 2021b). For each patch, they used the pooled cell graph representation on that patch as the node feature in the corresponding patch in the patch graph. The authors combined learned hierarchy learning with pre-established hierarchy by also applying self-attention graph pooling on both the cell- as well as the patch graph.

Sims et al. connected a cell graph with a level-1 and level-2 patch graph, which represent patches of increasing size (400  $\mu\text{m}$ , 800  $\mu\text{m}$ ) (Sims et al., 2022). They define their message passing for any cellular node  $i$  as  $CG_i \rightarrow L1_i \rightarrow L2_i \rightarrow L1_i \rightarrow CG_i$ , where each  $\rightarrow$  defines a message-passing function,  $CG_i$  represents the node in the cell graph and  $L1_i$ ,  $L2_i$  represent the node corresponding to the level-1 patch and the level-2 patch on which this cell exists. By applying message-passing in this way, the model can exchange information between distant cells without using many message-passing layers, as the cellular nodes belonging to the same layer-2 node can be 800  $\mu\text{m}$  away.

Guan et al. proposed a *Node-aligned* hierarchical graph-to-local clustering approach, inspired by the Bag-Of-Visual-Words (BOVW) methodology in NLP (Guan et al., 2022). Starting with a set of WSIs, the authors first clustered the patches for each WSI, using a codebook with patches from all other WSIs in the dataset. Then, a local clustering approach is used that samples global clusters from  $B$  to divide the global clusters into local subclusters  $K_G$ , such that  $B = sub_1, sub_2, \dots, sub_{K_G}$ . From each subcluster, they use K-means clustering to divide the subcluster  $sub_k$  into bins  $S_k$  and randomly select one patch in each bin. Now, for each WSI, we have a patch for each bin  $S_k \in K_G$  in each subcluster  $K_G \in B$ . The patches in each subcluster are connected using inner-sub-bag edges, and the subclusters themselves are connected using outer-sub-bag edges. This graph structure allows hierarchical information flow during message-passing in the GNN model.

Hou et al. proposed constructing a cell graph along with superpixel-based tissue graphs at two levels ( $CG, TG_{l1}, TG_{l2}$ ) (Hou et al., 2022a).

They generated features for the cell graph by using a pretrained ResNet on a patch around the nucleus centroid while generating tissue graph presentations by averaging ResNet embeddings from all crops belonging to a superpixel. The hierarchical information flow is modeled using a Transformer block that calculates the cross-attention between the graphs at different levels.

Shi et al. used graphs at 4 different levels of hierarchy: A tissue graph on 5x resolution, consisting of superpixels constructed using the SLIC algorithm, and 3 patch graphs at 5x, 10x, and 20x resolution, respectively (Shi et al., 2023b). The 5x resolution patch graph is used to generate features for the tissue graph. Then, after applying message-passing to the 10x- and 20x patch graphs, the interaction between the different hierarchical levels is modeled using a hierarchical attention module. This module produces a tissue graph where the interactions are captured in the node features. Message-passing layers, global attention layers, and a fully connected layer are applied subsequently to the tissue graph to come to a final prediction.

Gupta et al. modeled a tissue graph and a cell graph together as a heterogeneous graph with cellular nodes, tissue nodes, cell-cell edges, tissue-tissue edges, and cell-tissue edges:  $H = \{C, T, E_{cell \rightarrow cell}, E_{tissue \rightarrow tissue}, E_{cell \rightarrow tissue}\}$  (Gupta et al., 2023). After applying message-passing layers, they calculated the cross-attention between the cellular and tissue nodes using the transformer architecture to model the hierarchical relationships.

Abbas et al. established four separate hierarchical levels, where one level is a global image analyzed using a CNN model and the other levels are cell graphs constructed at different levels (global, spanning the entire WSI ( $G^{(0)}$ ), 512  $\times$  512px ( $G^{(1)}$ ), 256  $\times$  256px ( $G^{(2)}$ )) (Abbas et al., 2023). For each level, a subset of the segmented cells is randomly selected to build a cell graph. After applying message-passing layers on each level separately, the outputs are combined and processed using a fully connected layer. The combined representation and the representations gathered at each cell graph level separately are combined using an entropy weighting strategy, which weights the different representations based on the uncertainty of the model prediction given that representation.

**Multiresolution Hierarchies:** Xing et al. constructed hierarchical patch graphs at several levels of image resolution, thus aggregating information from multiple resolution levels. Starting with a single patch, they subsampled the same patch at increasingly lower resolution and connected the lower-resolution patches to the corresponding higher-resolution patch it was sampled from. This input graph was then used for a GNN model (Xing et al., 2021).

Bazargani et al. introduced hierarchy into their approach by constructing separate patch graphs on 5x, 10x and 20x resolution and then performing message-passing operations both on each graph separately as well as between graphs with different resolutions (Bazargani et al., 2022).

Bontempo et al. used a knowledge distillation approach combined with two patch graphs at different resolutions (high, low) (Bontempo et al., 2023). They performed message-passing both hierarchically between high and low resolution and in each resolution graph itself. They treated the high-resolution graph as a ‘teacher’ and the low-resolution graph as a ‘student’ network, between which they optimize the KL-divergence for the bag-level predictions at each resolution.

Mirabadi et al. proposed modeling the pyramidal multi-magnification structure in whole slide images as a multiresolution graph, where information on both the inner-magnification and the intra-magnification levels could be modeled (Mirabadi et al., 2024). They extracted patches from three magnification levels (20x, 10x and 5x), such that the patches on the higher resolutions are spatially equivalent to center crops of the patches at the lower resolutions. A RAG-graph was constructed such that nodes on each level were connected to both their adjacent neighbors on the same resolution as well as the spatially corresponding lower- and higher-level patch nodes. This allowed information to be exchanged between resolutions during

**Table 2**

Publications applying GNNs to histopathology which used a pre-established hierarchy. All hierarchies are shown small to large, such that when  $X \rightarrow Y$ , entities in  $X$  are subordinate to the entities in  $Y$ . CG: Cell Graph, PG: Patch Graph, TG: Tissue Graph, PHG: Patch Hypergraph.

Publication	Date	Application	Hierarchy
Pati et al. (2020)	2020/07	ROI classification	$CG \rightarrow TG$
Xing et al. (2021)	2021/08	Cancer subtyping	$PG_{40x} \rightarrow PG_{10x} \rightarrow PG_{5x}$
Wang et al. (2021b)	2021/09	Survival prediction	$CG \rightarrow PG$
Sims et al. (2022)	2022/01	ROI classification	$CG \rightarrow PG_1 \rightarrow PG_2$
Hou et al. (2022b)	2022/06	Binary classification	$PG_{10x} \rightarrow PG_{5x} \rightarrow PG_{\text{thumbnail}}$
Guan et al. (2022)	2022/06	Cancer subtyping	$S_k \rightarrow K_G \rightarrow B$
Hou et al. (2022a)	2022/09	Cancer subtyping	$CG \rightarrow TG_{i1} \rightarrow TG_{i2}$
Shi et al. (2023b)	2023/01	Cancer grading	$PG_{20x} \rightarrow PG_{10x} \rightarrow TG_{5x}$
Wang et al. (2023a)	2023/02	Cancer subtyping	$CG \rightarrow CCFG$
Gupta et al. (2023)	2023/07	Cancer subtyping, binary classification	$CG \rightarrow TG$
Bazargani et al. (2022)	2023/08	Cancer subtyping	$PG_{20x} \rightarrow PG_{10x} \rightarrow PG_{5x}$
Bontempo et al. (2023)	2023/10	Binary classification	$PG_{high} \rightarrow PG_{low}$
Abbas et al. (2023)	2023/12	Cancer grading	$CG_{256px} \rightarrow CG_{512px} \rightarrow CG_{global} \rightarrow WSI_{\text{thumbnail}}$
Mirabadi et al. (2024)	2024/02	Cancer subtyping	$PG_{20x} \rightarrow PG_{10x} \rightarrow PG_{5x}$
Godson et al. (2023)	2024/03	Binary classification	$PG_{40x} \rightarrow PG_{20x} \rightarrow PG_{10x}$
Ibañez et al. (2024)	2024/03	Cancer grading	$PG_{1x} \rightarrow PG_{5x} \rightarrow PG_{10x} \rightarrow PG_{20x}$
Liang et al. (2024b)	2024/04	Binary classification	$PHG_{40x} \rightarrow PG_{20x}$
Han et al. (2024)	2024/04	Survival prediction	$PG_{20x} \rightarrow PHG_{10x}$
Paul et al. (2024)	2024/05	Cancer grading	$CG, PG$
Cai et al. (2024)	2024/05	Survival prediction	$PG \rightarrow TG$
Breen et al. (2024)	2024/06	Cancer subtyping	$PG_{10x} \rightarrow PG_{5x}$

message passing. After message passing, a mean pooling operation was applied on each resolution level, resulting in a 3 node graph. This three-node graph embedding is then used for the downstream classification task.

Godson et al. hierarchically connected RAG patch graphs defined on 40x, 20x and 10x resolution (Godson et al., 2023). Nodes in each resolution were assigned a resolution-specific one-hot encoded vectors to uniquely barcode the nodes for each resolution. Similarly, all edges were assigned unique barcode vectors, depending on their resolution and whether the edges were hierarchical connections or intra-resolution connections. GATv2 layers were used to process the resulting hierarchical graph, where information was exchanged both between resolutions as well as inside each resolution graph.

Ibañez et al. constructed patch graphs across 4 resolution levels: 1x, 5x, 10x and 20x which were hierarchically connected with edges (Ibañez et al., 2024). During message-passing, information is exchanged between all resolution levels and through spatially adjacent patches. The authors introduced an *average scaled influence score* to allow comparative heatmaps across resolutions. They compute this core by calculating the median attention scores per resolution and then normalizing these across all WSIs.

Paul et al. used a cell-level graph for each patch and constructed a patch-level graph, but did not directly connect the two graphs using hierarchical edges (Paul et al., 2024). Instead, they use handcrafted global graph features gained from the cell-graph on each patch as the basis for the feature embedding for each patch. The patch-level graph was then constructed based on the cosine similarity between these features, effectively conditioning the patch graph structure on the cellular structure seen in each patch.

#### 4.1.3. Discussion and future prospects

We observed that hierarchical graph neural networks are an increasingly popular modeling technique for histopathology WSIs, due to the information in whole slide images existing on different levels of coarseness. One future approach will be to learn the necessary level of coarsening to establish an effective hierarchical structure end-to-end, which is currently controlled using a pooling ratio hyperparameter. We argue that different levels of graph coarseness might be optimal for different problems, as some problems in histopathology rely more on cell-level information, while others on larger tissue structures. Lastly, in most current approaches, message-passing occurs on each level of hierarchy separately, not directly between hierarchies. We argue that the field could move to message-passing schemes that are more effective at taking into account the hierarchical graph structure (Zhong et al., 2023).

#### 4.2. Adaptive graph structure learning

Most GNN applications in histopathology use a fixed input graph with fixed edge connectivity. While successful results have been achieved using this approach, we argue that it is suboptimal. Whether connections between nodes should exist is not clearly defined in the histopathology image, leading to the wide range of different approaches for constructing the input graphs, as previously discussed. These approaches are usually not based on biological or medical information and thus introduce inductive bias which might not reflect the biology in the tissue. To counteract this problem, one can either adjust the message-passing equation such that some edges are given more representative power than others (e.g. using GAT (Veličković et al., 2017)), or one can *learn* the graph structure during model training. The second approach, *Adaptive Graph Structure Learning* (AGSL), has gained more popularity recently (Table 3). In GNNs for histopathology, the strategies used for AGSL can be subdivided into three main paradigms: *Learned transformation* based AGSL, *CNN-filter* based AGSL and *Patch-selection* based AGSL. The first is based on a learned transformation that updates the adjacency matrix based on the model gradients. In the second approach, the graph is constructed using the learned filters inherent to convolutional neural networks, which are updated based on the model gradients. Lastly, some approaches using patch graphs learn to select a discriminative subset of patches from each slide during training, thereby also changing the nodes that will be present in the graph structure.

**Learned Transformation:** In 2020, Adnan et al. introduced adaptive graph learning for the classification of lung cancer subtypes (Adnan et al., 2020). The authors modeled the whole slide image as a fully connected graph of representative patches. Then, they used a pre-trained DenseNet for feature extraction. The graph connectivity is learned end-to-end using both global WSI context and local pairwise context between patches. Let us denote WSI  $W$  with patches  $w_1, \dots, w_n$ , where for each patch  $w_i$  we have a feature vector  $x_i$ . The authors first pooled the patch representations into a global context vector  $c$  using a pooling function  $\phi$  (e.g. sum):

$$c = \phi(x_1, x_2, \dots, x_n) \quad (29)$$

The global vector  $c$  is concatenated to each patch feature vector  $x_i$  and is jointly processed by MLP layers which gives a feature vector  $x_i^*$  that contains both local and global context information. Finally, the matrix  $X^*$ , which holds all feature vectors  $x_i^*$ , is processed using a cross-correlation layer that determines the connectivity of the output graph

**Table 3**

Publications applying GNNs in histopathology and using adaptive graph structure learning strategies.

Publication	Date	Application	Adaptive learning mechanism
Adnan et al. (2020)	2020/05	Binary classification	Learned transformation
Gao et al. (2022)	2022/02	Cancer subtyping	CNN-filter based
Hou et al. (2022a)	2022/09	Cancer subtyping	Learned transformation
Behzadi et al. (2024)	2022/12	Cancer grading	Learned transformation
Ding et al. (2023b)	2023/02	Cancer subtyping, Cancer grading	CNN-filter based
Liu et al. (2023)	2023/04	Survival prediction	Patch selection
Li et al. (2024b)	2024/06	Cancer subtyping, Cancer grading	Learned transformation
Shu et al. (2024)	2024/07	Binary classification	Learned transformation
Kim et al. (2024)	2024/07	Binary classification	Patch selection
Liu et al. (2024)	2024/08	Patch classification	CNN-filter based
Lu et al. (2024)	2024/10	Survival prediction	Patch selection

in  $A$ , where each element  $a_{ij} \in A$  represents the correlation between patches  $w_i$  and  $w_j$ , which are used as edge weights in the learned graph structure.

Hou et al. described a spatial-hierarchical GNN framework that could dynamically learn the graph structure during model training (Hou et al., 2022a). Their *Dynamic Structure Learning* module first embeds the representation of both node features  $V$  and centroid coordinates  $P$  together into a single representing  $J$ , using the following equation:

$$J = \text{Concat}[\sigma(P^T W_1), \sigma(V^T W_2)] \quad (30)$$

where  $W_1$  and  $W_2$  are learned weight matrices and  $\sigma$  denotes a non-linear activation function. Next, the authors applied a distance-thresholded k-NN algorithm on the acquired embedding  $J$  to determine the edge connectivity. Given a set of nodes  $V$ , set of edges  $E$ , distance threshold  $d_{\min}$  and the number of neighbors  $k$ , we use the following equation to determine the edges in  $E$ :

$$e_{uv} \in E \iff \{u, v \in V \mid \|v - u\|_2 \leq \min(d_k, d_{\min})\} \quad (31)$$

Here,  $d_k$  denotes the distance between nodes  $u$  and the  $k$ -closest neighbor. The weight matrices are trained using the overall loss function in the GNN framework, which allows the graph connectivity to adapt to the learning task at hand.

Liu et al. propose learning the graph structure based on the cosine similarity between the transformed patch feature vectors (Liu et al., 2023). Given an input feature matrix  $X$  and a transformation matrix  $T$ , we create a projected matrix  $P = XT$ . They then calculate the cosine similarity between each pair of patches in  $P$ , which are saved as a symmetric adjacency matrix  $A_L$ , which holds the ‘edge strength’ between any two patches in  $P$ . The edge strength is then thresholded using a set threshold  $\epsilon$ :

$$e_{uv} \in E \iff \{u, v \in V \mid \frac{P[u] \cdot P[v]}{\|P[u]\| \cdot \|P[v]\|} \leq \epsilon\} \quad (32)$$

where  $P[u]$  and  $P[v]$  denote the projected feature vectors of nodes  $u$  and  $v$ , respectively. Note that the transformation matrices are learned, which allows the graph structure to be adapted during model training.

Li et al. introduced *tail* and *head* embeddings calculated as two separate linear projections of feature vectors from each WSI patch (Li et al., 2024b). The head projection is trained to capture the correlation between other patches and itself, while the tail captures the contribution of the patch to other patches. The authors then calculate the similarity between each head and tail embeddings to decide which patches should be connected. Note that these connections are represented as directed edges from the patch of the head embedding to the patch of the tail embedding, showing adequate similarity. Additionally, each directed edge is assigned an edge embedding defined as a weighted sum of the head and tail embeddings. The learned directed graph is then processed using a knowledge-aware attention mechanism that takes into account the head, tail and edge embeddings.

Shu et al. implemented a learnable slide-level graph (Shu et al., 2024). By pooling patch embeddings from a pretrained encoder model, they generated a slide-level embedding for each WSI in their dataset.

In addition to a more traditional MIL-classification branch which uses these slide embeddings, the embeddings were also combined into a slide-graph. This graph was processed with GCN layers and treated as a node classification task to obtain slide-level predictions. Knowledge distillation was applied between the MIL classifier and GNN-classifier to regularize the classification. Since the construction of the slide-level graph is conducted using K-NN on a learned projection of the slide embeddings, the slide-level graph structure changes during training.

**Patch-selection:** Kim et al. constructed a patch graph based on learned patch clusters from which representative patches were selected using Gumbel softmax (Kim et al., 2024). The similarity between these patches was captured in a similarity matrix. The top  $k$  most similar patches were connected to form the patch graph. Allowing the input patches and associated embeddings to change during training leads to a learned graph structure.

Behzadi et al. propose selecting a discriminative subset of patches to form a concise patch graph for each WSI (Behzadi et al., 2024). They use a patch image autoencoder to obtain latent features for each WSI patch. These latent features are processed using a multi-head attention mechanism to obtain attention scores per patch, which are used as the selection criterion for inclusion in the patch graph. The selected patches are connected using the K-NN algorithm for subsequent GNN-based processing.

**CNN-filter Based:** Gao et al. and Ding et al. both use a very different approach, where the learned feature maps generated by a CNN are used as basis for the graph construction (Gao et al., 2022; Ding et al., 2023b). More specifically, they treat the pixels in each feature map as nodes, in which the features are spatially concatenated across channels into a node feature vector. Then the k-nn algorithm is used to connect the nodes. By basing the graph structure on learned CNN feature maps, the graph structure is learned by training the CNN. Since each pixel in the feature maps corresponds to a patch in the input WSI, the constructed graph can capture spatial dependencies between regions in the WSI. Given the acquired node embedding matrix  $X \in \mathbb{R}^{N \times C}$  where  $N$  is the number of nodes and  $C$  the amount of feature map channels, we determine the existence of edges as follows:

$$e_{uv} \in E \iff \{u, v \in V \mid \|u_f - v_f\|_2 \leq d_k\} \quad (33)$$

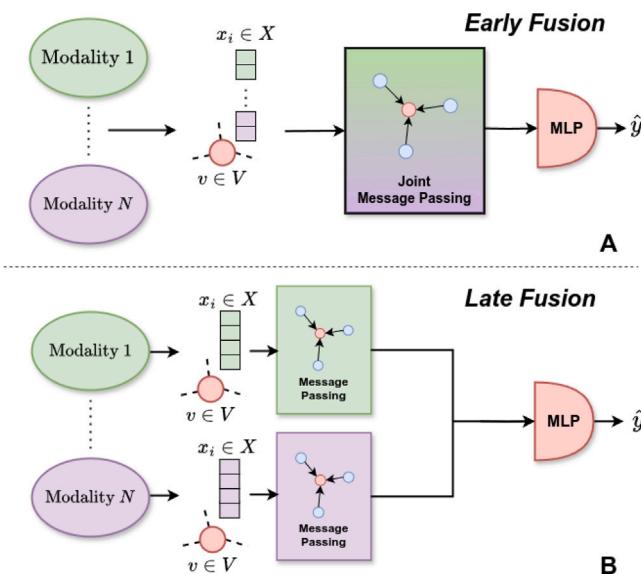
where  $u_f$ ,  $v_f$  are the feature vectors of node  $u$  and  $v$ , and  $d_k$  is the distance between node  $u$  and the  $k$ -closest neighbor of  $u$ .

Liu et al. expanded on this idea of using convolutional feature maps as the basis for graph construction by hierarchically applying convolutional layers, constructing graphs based on the learned feature maps, processing this graph using a message passing layer and finally applying a transformer layer to model long-range dependencies (Liu et al., 2024). This results in a set of feature vectors which can be used as input for the convolutional layer in the next hierarchical layer. This process is repeated 4 times before using several linear layers to obtain the final prediction probabilities. The new graph constructed at every hierarchical layer utilizes *dilated K-NN*, which skips every  $d$ th nearest neighbor, effectively exponentially increasing each node’s receptive field.

**Table 4**

Applications of Multimodal GNNs in histopathology. CD: Clinical Data, CNV: Copy Number Variation, TC: Trichrome, MP: MultiPhoton microscopy, TPEF: two-photon excited fluorescence microscopy, MRI: Magnetic Resonance Imaging, SHG: Second-Harmonic Generation microscopy, ST: Spatial Transcriptomics, IHC: Immunohistochemistry, GE: Gene Expression, miRSeq: microRNA sequencing, DNA-Met: DNA methylation.

Publication	Date	Application	Fusion	Modalities
Chen et al. (2020a)	2020/09	Survival prediction, Cancer subtyping	Late	H&E WSI, GE, CNV
Dwivedi et al. (2022)	2022/04	Cancer grading	Late	H&E WSI, TC WSI
Qiu et al. (2022)	2022/07	Survival prediction	Early	H&E WSI, MP, TPEF
Zuo et al. (2022)	2022/09	Survival prediction	Late	H&E WSI, GE
De et al. (2022)	2022/10	Cancer subtyping	None	H&E WSI, MRI
Li et al. (2022)	2022/11	Cancer subtyping	Early	H&E WSI, SHG
Xie et al. (2022b)	2022/12	Survival prediction	Late	H&E WSI, GE
Fatemi et al. (2023)	2023/03	ST-prediction	None	H&E WSI, ST
Jiang et al. (2023)	2023/03	Mutation prediction	Late	H&E WSI, CD
Gao et al. (2023)	2023/07	ST-prediction, survival prediction	None	H&E WSI, ST
Gallagher-Syed et al. (2023)	2023/09	Rheumatoid Subtyping	Early	H&E WSI, IHC WSI
Pati et al. (2023)	2023/12	Survival prediction, Cancer grading	Both	H&E, virtual IHC
Azher et al. (2023)	2024/01	Survival prediction, Cancer grading	Early	H&E WSI, ST
Zheng et al. (2023)	2024/01	Survival prediction	Late	H&E WSI, GE
Zhang et al. (2024)	2024/04	Survival prediction	Late	H&E WSI, RNA-seq
Chi et al. (2024)	2024/06	ST prediction	None	H&E WSI, GE, ST
Palma et al. (2024)	2024/09	Binary classification	Early	H&E WSI, GE, miRSeq, DNA-Met, CNV, CD
Li et al. (2024a)	2024/10	ST-prediction	None	H&E WSI



**Fig. 7.** Early fusion (A) versus late fusion (B). In early fusion, information from different modalities is typically integrated in the node features before message passing, enabling joint spatial modeling of modalities. In late fusion, meanwhile, modalities are separately processed and combined before the final model layers which calculate the model prediction. FCN: Fully Connected Layer.

#### 4.2.1. Discussion and future prospects

Outside of histopathology, most adaptive graph structure learning assumes graph *homophily* (Zhu et al., 2021a) where similar nodes are likely to be colocated. This is not always the case in histopathology, as some structures might be composed of different cell types that can vary widely in morphology. Furthermore, most applications focus on homogeneous graphs, where a single type of node and edge exists. Work by Zhao et al. (2021) showed that we can learn a heterogeneous graph optimized for downstream tasks, which is suitable for graphs showing heterophily, which is often the case in histopathology. Therefore, we argue that heterogeneous graph learning will be a useful approach for histopathology if we model the whole slide images as a heterogeneous graph.

#### 4.3. Multimodal GNNs

In histopathology diagnostics, different modalities are often combined to assist in clinical decision-making and prognostic predictions. While most applications of GNNs in histopathology focus solely on H&E image data, approaches considering multiple modalities have gained popularity recently. Combining data from multiple modalities helps increase model accuracy and generalization. Graph Neural Networks are especially suitable for multimodal integration, since data from different modalities can be easily combined in the node- and edge feature vectors (Ding et al., 2022b), which can then be jointly modeled with a controllable spatial inductive bias. Multiple approaches combined IHC-stained biopsy images with H&E stained biopsy images, while other approaches incorporated spatial transcriptomics or genetic data in the model input. We differentiate between *Stain multimodality*, where the same whole slide images with different stainings (e.g. IHC) are combined, and *Full multimodality*, where the modalities are not based on whole slide images, (e.g. CT-scans, gene expression data). An overview of multimodal GNNs in histopathology is given in Table 4.

An important challenge in multimodal integration in Deep Learning models is how- and where in the model architecture data from different modalities should be combined, which we call *fusion*. In a GNN context, we broadly differentiate between *early fusion*, where data from different modalities are combined before message passing and thus jointly modeled in the GNN and *late fusion*, where data is combined after the message passing steps (Fig. 7).

We broadly categorize multimodal GNNs into four groups: *Pathomic fusion-based*: This uses the pathomic fusion strategy popularized by Chen et al. (2020a), *Early fusion*, *Late fusion* and *Modality prediction*, encompassing models that predict one modality using another. Models that do not directly fuse modalities but use predictions from one modality to drive how the other modalities are processed are considered late-fusion models.

#### 4.3.1. Full multimodality

**Pathomic Fusion:** Chen et al. integrated whole slide image information together with RNA-Seq count-data and copy number variant (CNV) information (Chen et al., 2020a). They used this combined information for cancer subtyping and survival analysis in TCGA data sets for clear cell renal cell carcinoma and glioma. Their multimodal model fused information from 3 different modules: A CNN-based image module, a GNN-based cell graph module, and a genomic module, which took CNV and RNA-seq information as input. Each of these modalities was first processed individually before fusing the information. Their approach for multimodal fusion, which they call *Pathomic fusion*, models

interactions between modalities via the Kronecker product of attention-gated representation. The attention gating is applied to the hidden representation of modality  $m$ ,  $h_m$ , by learning a transformation  $W_{ign \rightarrow m}$  which assigns an importance score for each modality, which we denote as  $z_m$ :

$$h_{m,gated} = z_m * h_m, \quad \forall m \in \{i, g, n\}$$

where,  $h_m = \text{ReLU}(W_m \cdot h_m)$  (34)

$$z_m = \sigma(W_{ign \rightarrow m} \cdot [h_i, h_g, h_n])$$

where  $h_i$ ,  $h_g$ , and  $h_n$ , are the gated representation vectors of the image module, graph module, and genomic module, respectively. The authors calculated the Kronecker product of these vectors to get a combined representation  $h_{fusion}$ :

$$h_{fusion} = \begin{pmatrix} h_i \\ 1 \end{pmatrix} \otimes \begin{pmatrix} h_g \\ 1 \end{pmatrix} \otimes \begin{pmatrix} h_n \\ 1 \end{pmatrix} \quad (35)$$

where  $\otimes$  denotes the outer product. The result,  $h_{fusion}$  is a three-dimensional tensor that can then be connected to a fully connected layer for classification tasks or survival prediction.

Jiang et al. predicted EGFR-gene mutations in lung cancer (Jiang et al., 2023) by augmenting the approach used by Chen et al. (2020a). The author's approach differs from that of Chen et al. by not using genomic data but instead using clinical information (e.g., gender, age) as the third modality, next to a spatial cell graph and whole slide image. Compared with a previous model from the same group (Xiao et al., 2022), which used a cell graph and image module but no clinical features, the authors found considerable performance increases for the multimodal model.

**Early Fusion:** Azher et al. integrated spatial transcriptomics data with accompanying H&E WSI data to predict survival and cancer grade in colorectal cancer (Azher et al., 2023). The authors first constructed an embedding model that used an ImageNet-pretrained CNN to encode H&E patches and fully connected layers to encode spatial gene expression data at the same location. They then optimized a projection layer to merge the data from these modalities into a single vector using a combination of unimodal and cross-modal loss functions. This effectively trained the model to encode a cross-modal embedding vector. The acquired embeddings were used as node vectors in a GNN model for downstream tasks. The authors showed that the use of expression-aware embeddings improved model performance on all tasks, indicating that pretraining using coupled H&E WSIs and spatial transcriptomics datasets can help retrieve more discriminative embeddings for downstream tasks.

Palmal et al. integrated data from six different modalities: WSI patches, miRSeq, DNA methylation, copy number variation (CNV), mRNASeq and clinical data to predict short- or long-term survival using the TCGA-BRCA dataset (Palmal et al., 2024). Each modality was represented as a graph, which was processed using several GCN layers. The learned embeddings are enhanced using contrastive learning, with the objective of aligning the embeddings of the nodes of neighboring nodes while separating those between nodes further apart. The learned node embeddings are utilized in cross-attention blocks that model the interaction between the modality-specific graphs. Finally, the outputs of the cross-attention blocks are concatenated and used as input in a random forest model to provide the model prediction.

**Late Fusion:** Zuo et al. integrated H&E stained whole slide images with genomic biomarker information (Zuo et al., 2022). Specifically, they constructed a graph of Tumor Infiltrating Lymphocyte (TIL) patches with Tumor patches and analyzed this graph using a GNN. Genomic data consisted of mRNA gene counts, which were transformed to a gene co-expression module matrix using the lmQCM algorithm. They then applied a concrete autoencoder model to the co-expression matrix to identify survival-associated features. The GNN- and autoencoder outputs were then fused using a self-attention layer.

De et al. combined MRI- and H&E stained whole slide images of brain tumors to predict the type of brain cancer (De et al., 2022). The modalities were not directly fused; instead, the authors first used a 3D-CNN model to detect whether the cancer was one of the possible cancer types (Glioblastoma). If this was the case, the model simply outputs glioblastoma as its prediction. When this was not the case, a patch graph was constructed from the H&E image, which was used as input for a GNN model. Finally, this GNN model could predict one of the remaining subtypes (Normal, Astrocytoma, or Oligodendrogloma).

Xie et al. combined gene expression with H&E whole slide image data for survival prediction in gastric cancer (Xie et al., 2022b). Here, the authors first processed the WSI data and gene expression data separately using MLP layers. Then the interaction between each WSI patch and each gene feature vector was calculated using a cross-modal attention layer. After this processing, the data from both modalities was aggregated using a MIL-aggregation module and finally fused using concatenation. The fused embeddings were used to construct a patient graph, based on the similarity of the fused embeddings between the patients. A GNN was used to process this graph, which produced a survival prediction.

Zheng et al. fused gene expression signatures with a WSI patch graph using their *Genomic Attention Module* approach (Zheng et al., 2023). After message-passing on the patch graph, the pairwise interactions between each patch and each individual gene signature are modeled using a self-attention mechanism. This allows the model to learn the interactions between spatial tissue regions and gene signatures, which allowed the authors to visualize which gene signatures were associated with certain regions in the WSI.

Zhang et al. fused pathway-enriched RNASeq-data with H&E stained WSI images by constructing modality-specific graphs for each (Zhang et al., 2024). The RNASeq-data was used to train the model to construct a graph where each node represents a significantly enriched pathway in the gene expression data, where edges are defined based on the correlation of biological function and the overlap in associated genes. This graph is then used in a graph autoencoder model, which reconstructs the adjacency matrix based on its latent representation. The WSI-graph consisted of nodes representing 20x resolution patches along with pretrained ResNet50 features, connected using a RAG-strategy. The pooling of these patch features to a WSI-level embedding was regularized by a gene set variation analysis obtained from the RNA-Seq data. Further interaction between the RNA-Seq and WSI-data was modeled by aligning linear projections of the WSI-graph and the pathway-node graph and by applying cross-modality attention blocks. Subsequently, the outputs were pooled using global attention pooling and mapped to a risk value for survival prediction.

Elforaici et al. integrated clinical data along with GNN-processed cell graphs defined on WSIs in a support vector machine (SVM) classification model (Elforaici et al., 2024). They used their modeling approach for accurate tumor response grading in colorectal liver metastases.

**Modality Prediction:** Fatemi et al. integrated spatial transcriptomic data with co-localized H&E whole slide imaging data to characterize spatial tumor heterogeneity in colorectal cancer (Fatemi et al., 2023). They achieved this by training a model to predict spatial gene expression from the H&E whole slide image. The authors tried to predict spatial gene expression using both a CNN- and a GNN-based network and showed that, for this task, the CNN-based methods performed better.

Gao et al. predicted spatial transcriptomic data using H&E images by integrating image and cell graph data using CNN- and GNN-based models (Gao et al., 2023). The authors showed that integrating the graph and image-based information together did significantly improve over using either one alone.

Chi et al. utilized the *deep graph infomax* (Veličković et al., 2018) GNN-based contrastive learning framework to align spatial transcriptomics data with WSI patches and raw gene expression data (Chi et al.,

2024). During inference using a single WSI patch, the top  $k$  most similar gene expressions were retrieved from a database of gene expression embeddings to predict spatial gene expression.

#### 4.3.2. Stain multimodality

**Early fusion:** Li et al. fused information from Second-Harmonic Generation (SHG) microscopy images and H&E whole slide images together to differentiate between pancreatic ductal adenocarcinoma and chronic pancreatitis in pancreatic cancer (Li et al., 2022). The images from both modalities were registered. The features from each modality were combined into node features for the input graph, where nodes represented registered patches in both modalities. An ImageNet-pretrained ResNet model was used to retrieve features from the H&E patches, while collagen fiber-specific handcrafted features were extracted for each SHG-patch. A H&E-SHG graph was constructed where the node vectors contained the concatenation of the patch features from both modalities. This graph was used in a GNN model which predicted between the two classes.

Gallagher-Syed et al. integrated data from IHC- (CD138, CD68, CD20) and H&E stained synovial biopsy samples to predict a Rheumatoid Arthritis subtype using a GNN model (Gallagher-Syed et al., 2023). Information between the staining modalities was exchanged by modeling each patch, from each staining, as a node and connecting the nodes based on their feature similarity to get a single multistain graph. The authors showed that the features across stains were similar enough to cause nodes from different staining to mix in the graph and, thus, enable information exchange between the modalities in message passing layers of the GNN. The authors used the multimodal graph as input for a GNN model whose output was used to predict the rheuma subtype.

**Late fusion:** Dwivedi et al. combined trichrome- (TC) and H&E stainings of liver biopsies to predict liver fibrosis (Dwivedi et al., 2022). The authors experimented with different mid- and late fusion techniques. Their experiments showed that their late concatenation or addition and the pathomic fusion strategy proposed by Chen et al. (2020a) performed the best for fibrosis prediction. In the late and pathomic fusion strategies, they separately processed both the H&E and TC tissues as graphs using a GNN and then fused the features from both modalities together.

Qiu et al. combined information from H&E stainings, multiphoton microscopy (MP), and two-photon excited fluorescence (TPEF) applied to the same breast cancer biopsies (Qiu et al., 2022). Instead of fusing the modalities in the model itself, the authors determined tumor-associated collagen signatures from the 3 different modalities in different regions to calculate a 8-bit binary vector for each region. The regions sampled were treated as graph nodes having the binary vector as node attributes. Using these nodes, a fully connected graph was constructed and used as input for a survival prediction GNN-model.

**Modality prediction:** Pati et al. used a generative approach to virtually predict IHC stained tissue images from H&E slide images, and then used a multimodal GNN Transformer model to perform survival prediction and cancer grading tasks in prostate cancer, breast cancer, and colorectal cancer (Pati et al., 2023). The authors used three strategies for fusion (no fusion, early fusion, late fusion) and found that early fusion works optimally for both tasks. In early fusion, the authors combined ImageNet-pretrained ResNet features from the same patch in all modalities to form the node features in the input graph. In late fusion, meanwhile, all modalities were assigned a separate input graph, which was processed separately using the GNN Transformer model. Subsequently, the output features were combined. The authors hypothesized that early fusion allowed the model to learn multimodal spatial interactions during message passing, causing a performance gain compared to the other fusion strategy.

#### 4.3.3. Discussion and future prospects

In this section, we highlighted the use of graph-based modeling in multimodal approaches, which have been successfully applied both for integration of multimodal data to improve prediction accuracy and for predicting a translation from one modality to another, particularly in the context of spatial transcriptomics data. We argue that, in cases where the modalities are not spatially aligned, graphs themselves could be utilized more for the multimodal integration itself. For example, several researchers have used the concept of a *Patient graph*, where nodes represent (aggregated) data points from different medical modalities corresponding to the same patient (Kim, 2023) or multiple patients (Gao et al., 2020; Ochoa and Mustafa, 2022). Some approaches use graphs to model time series data, where, for example, medical information on the same patient gathered at different time points can be effectively utilized (Rocheteau et al., 2021; Daneshvar and Samavi, 2022). Zheng et al. proposed a framework in which adaptive graph structure learning and GNNs are combined to integrate data from different medical modalities for disease prediction (Zheng et al., 2022). One major problem in the application of multimodal approaches in histopathology is that, often, not every modality is available for each patient. This effectively creates a missing modality problem. Ma et al. proposed a Bayesian meta-learning framework which mitigates this problem, allowing effective multimodal learning and prediction even when numerous modalities are missing in the data (Ma et al., 2021). We argue that these approaches should be combined to effectively model the relationships between modalities, based on the task at hand, even in settings where modalities are missing.

#### 4.4. Higher-order graphs

While graphs have shown to be adequate formats for the representation of histopathology slides, it is limited by the fact only pairwise relations can be modeled. Furthermore, the entities in the graphs can solely be modeled as nodes and edges. This limitation has inspired extensions to the graph modeling framework, which are suited to model *higher-order graphs*, which move beyond pairwise node-edge interactions. Examples of higher-order graphs are hypergraphs, cellular complexes, and combinatorial complexes. To allow learning from these higher-order graph structures, message-passing frameworks called *topological neural networks* (TNNs) have been developed (Papillon et al., 2023).

In histopathology, TNNs have not yet been widely adopted, but there has been a steadily increasing number of publications that model WSIs as hypergraphs. Hypergraphs extend the graph modeling framework with *hyperedges*, which can connect sets containing an arbitrary number of nodes in the graph. This allows hypergraphs to model relations that rely on more than 2 pairwise entities. Deep learning on hypergraphs is typically achieved using hypergraph neural network architectures, such as HGNN (Feng et al., 2019) and Hyper-GAT (Ding et al., 2020). We provide an overview of publications using higher-order graphs in histopathology in Table 5.

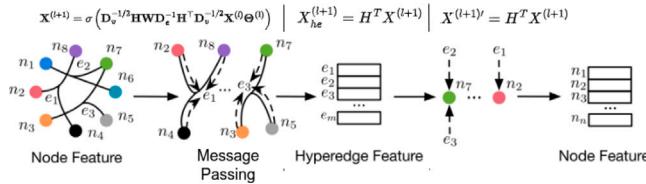
Let us denote a hypergraph as  $G = (V, E_{hyp})$ , which consists of a set of nodes  $V$  and a set of hyperedges  $E_{hyp}$ . Each hyperedge in  $E_{hyp}$  is a subset of  $V$ , connecting any number of vertices. For example, a hypergraph with vertices  $V = \{v_1, v_2, v_3, v_4\}$  and hyperedges  $E_{hyp} = \{\{v_1, v_2\}, \{v_2, v_3, v_4\}, \{v_1, v_3, v_4\}\}$  of  $V$ , expressing relationships between multiple nodes simultaneously. We denote the connectivity of a hypergraph using an incidence matrix  $H^{|V| \times |E|}$  whose entries are defined as:

$$h(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (36)$$

for nodes  $v \in V$  and edges  $e \in E_{hyp}$ . For any node  $v$ , its degree is defined as  $d(v) = \sum_{e \in E_{hyp}} h(v, e)$ , similarly for any edge  $e \in E_{hyp}$ , its degree is defined as  $d(e) = \sum_{v \in V} h(v, e)$ . These degrees are saved in diagonal matrices  $D_e$  and  $D_v$ , which contain the edge degrees and node degrees,

**Table 5**  
Publications which utilized hypergraph neural networks for histopathology WSI analysis.

Publication	Date	Application	Hypergraph type	Message-Passing
Di et al. (2020)	2020/09	Survival prediction	Patch Hypergraph	HGNN
Bakht et al. (2021)	2021/05	Patch classification	Patch Hypergraph	HGNN
Di et al. (2022b)	2022/09	Survival prediction	Patch Hypergraph	HGMConv
Benkirane et al. (2022)	2022/11	Survival prediction	Patch Hypergraph	HGCN, HGAT
Liang et al. (2024a)	2024/02	Binary classification	Patch HyperGraph	Adaptive HGNN
Han et al. (2024)	2024/04	Survival prediction	Patch Hypergraph	HMHA
Liang et al. (2024b)	2024/04	Binary classification	Patch Hypergraph	HGNN
Cai et al. (2024)	2024/05	Survival prediction	Patch Hypergraph	HGNN
Shi et al. (2024)	2024/08	Cancer subtyping, mutation prediction	Patch Hypergraph	HGAT
Li et al. (2024a)	2024/10	Gene expression prediction	Patch Hypergraph	HGNN
Zhou et al. (2024)	2024/10	Metastasis prediction	Patch Hypergraph	(Multi-slice) HGNN
Lu et al. (2024)	2024/10	Survival prediction	Patch Hypergraph	HGNN



**Fig. 8.** Graphical overview of the hypergraph neural network framework (Feng et al., 2019). First, message-passing gets applied between all nodes connected to the same hyperedge. Then the learned features are calculated on a hyperedge-level. Finally, the hyperedge-level features are used to calculate the new node features.

respectively. Lastly, we denote the matrix of node features as  $X$ . The decision on which nodes to connect to a hyperedge is usually based on feature similarity or spatial distance (i.e., closely related nodes are connected together by a single hyperedge). Feng et al. introduced the hypergraph neural network (Feng et al., 2019) (visualized in Fig. 8), which defined a message passing operation on hypergraphs as follows:

$$X^{(l+1)} = \sigma\left(D_v^{-1/2} HWD_e^{-1} H^T D_v^{-1/2} X^{(l)} \Theta^{(l)}\right) \quad (37)$$

where  $W$  is a learned weight matrix,  $\sigma$  denotes a nonlinear activation function, and  $\Theta$  is a learnable filter matrix used for feature extraction. After applying message passing, we have an updated feature matrix  $X$ . This can then be used to obtain features on the hyperedge level, mathematically defined as:  $X_{he}^{(l+1)} = H^T \times X$ . Finally, to get the updated node-level embeddings is acquired by multiplying the hyperedge features with the incidence matrix:  $X^{(l+1)\prime} = X_{he}^{(l+1)} \times H$ .

Di et al. were the first to model WSIs as hypergraphs (Di et al., 2020). They used their hypergraph approach for survival prediction in lung and brain cancer datasets. The authors started by constructing sets of  $K$  similar patches based on the Euclidean distance between the feature vectors, which were retrieved using an ImageNet-pretrained ResNet model.  $N$  hyperedges are then used to connect the patches in each of the sets. The authors then used the node feature matrix  $X$  with the defined hypergraph, captured in  $H$ , and updated the features using a series of convolutional hypergraph layers (HGNN). The acquired representations after message passing were then used for the downstream survival prediction task. The authors show that their hypergraph-based method outperforms other CNN- and GNN-based frameworks for survival prediction.

Bakht et al. extended this idea by introducing a new strategy for constructing the hypergraph (Bakht et al., 2021). Given a fixed neighbor parameter  $k$ , their hypergraph construction strategy starts by defining the distance between any two patches  $i, j$  as:

$$d_k(i, j) = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right) \quad (38)$$

where  $x_i, x_j$  represent the feature vectors of patch  $i$  and  $j$ , respectively, and  $\sigma$  is a bandwidth parameter. Then, the authors calculated the vertex-edge probabilistic incidence matrix which determines the

probability of a node  $v$  to be connected using hyperedge  $e$ :

$$h(n, e) = \begin{cases} \exp\left(-\frac{d}{p_{\max} d_{\text{avg}}}\right), & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (39)$$

Here  $d$  denotes the distance between the current node  $n$  and the neighboring node.  $p_{\max}$  denotes the maximum probability and  $d_{\text{avg}}$  is the average distance between all  $k$  nearest neighbors. Finally, they use this incidence matrix to calculate the node and edge degrees:

$$d(v) = \sum_{v \in V} h(v, e), \quad d(e) = \sum_{e \in E} h(v, e) \quad (40)$$

The degrees are combined into matrices  $D_v$  and  $D_e$ , which are used, together with the incidence matrix  $H$  and node feature matrix  $X$  in 3 HGNN message passing layers. The output of these layers was used to predict the label of patches in the WSI.

Di et al. then expanded on their previous work by using multiple hypergraphs that are fused together to be used as input for message passing layers (Di et al., 2022b). Specifically, they construct a topological hypergraph and a phenotype (feature-based) hypergraph. The authors sampled patches sequentially from the tissue boundary to the tissue center and grouped the patches in the same sequence step in the same topological area. The topological hypergraph is constructed by connecting neighboring patches with a hyperedge if they belong to the same topological area. The phenotype hypergraph meanwhile, is constructed using K-NN based on the vector similarity between the patch features. The two hypergraphs are then concatenated together to form a total incidence matrix  $H$ . For processing the constructed hypergraph the authors use max-mask convolutional layers, which are defined in 4 iterative steps:

- 1. Hyperedge Feature Gathering:** First, hyperedge-level features are formed by multiplying the hypergraph incidence matrix ( $H$ ) and the node feature matrix ( $F_v^{(l)}$ ). This step aggregates the information from nodes connected by each hyperedge, resulting in hyperedge-level features ( $F_e^{(l)}$ ).
- 2. Max-Mask Operation:** After gathering hyperedge-level features, a max-mask operation is performed on each dimensionality of  $F_e^{(l)}$ . This operation aims to avoid overfitting by disregarding the contribution of dominant hyperedges that take the largest values.
- 3. Node Feature Aggregating:** By multiplying the hyperedge features with the transposed incidence matrix ( $H^T \times F_e^{(l)}$ ), we can calculate the node features ( $F_v^{(l+1)}$ ).
- 4. Node Feature Reweighting:** Finally, the output node features are further weighted using learnable parameters ( $t^{(l)}$ ), which are represented as a diagonal matrix. This reweighting is followed by a non-linear activation function ( $\sigma$ ). The reweighting step allows the model to learn the importance of different node features and adaptively adjust them.

Mathematically, the max-mask convolutional layer is defined as follows:

$$\begin{aligned} X^{(l+1)} &= \sigma((I - L)X^{(l)} + H^{-1}(I - L)X^{(\lambda)} \cdot i^{(l)}) \\ F_e^{(l+1)} &= H^{-1}(I - L)X^{(l)} + X^{(\lambda)} \end{aligned} \quad (41)$$

Here,  $L$  is the multigraph Laplacian matrix, and  $I$  denotes the identity matrix.  $H^{-1}(I - L)X^{(\lambda)}$  functionally ensures that the top  $\lambda$  attribute feature dimensionalities are ignored during the gradient calculation.

Bankirane et al. used adaptive agglomerative clustering to construct a patch hypergraph, which was then processed using a combination of HGNN and HGAT layers (Bankirane et al., 2022). In the agglomerative clustering, a similarity kernel was used that took into account both spatial location and feature similarity between patches. This kernel calculated pairwise similarity scores between all patches. If the similarity score was higher than a fixed threshold  $\delta$ , the patches were assigned to the same cluster  $C_k$ . For each cluster, the representation of the patches in the cluster was averaged to obtain cluster-level representations. Each clustered patch was treated as a hypergraph node. The hyperedges connected all nodes with a feature similarity higher than a fixed threshold  $\delta_h$ . We denote the neighborhood of a clustered node  $c_i$  as  $\gamma(c_i) = c_j \in C; \kappa_h(c_i, c_j) \geq \delta_h$ . Here,  $C$  denotes the set of all clusters and  $\kappa_h(c_i, c_j)$  denotes the output of the feature similarity kernel  $\kappa_h$ . Having determined the neighborhood, we can calculate the incidence matrix  $H$  where:

$$h(k, j) = \begin{cases} 1, & \text{if } c_i \in \gamma(c_j) \\ 0, & \text{else} \end{cases} \quad (42)$$

The authors then used the incidence matrix  $H$ , and node feature matrix  $X$  as input for a series of combined HGNN-HGAT layers which were subsequently pooled into a global representation. Finally, this representation was used as input for an MLP layer, which predicted the hazard score for survival prediction.

Liang et al. introduced the *adaptive HGNN* to histopathology, for the classification of sentinel node metastases and the differentiation between lung squamous cell carcinoma and lung adenocarcinoma (Liang et al., 2024a). The authors used the K-NN algorithm on patch-level ImageNet-pretrained ResNet features to construct a hypergraph of patches, where the  $k$  most similar patches were connected using a hyperedge. Their main innovation comes in the form of an adaptive HGNN, which can adjust the correlation strength between nodes and hyperedges on the graph during model training. They first denote a matrix of edge strength in layer  $l$  as  $T^{(l)}$ . Each element  $t_i^{(l)} \in T^{(l)}$ , which denotes the attention score of the node  $i$  and its associated hyperedge  $e_{i,i'}$  in the  $l$ th layer, is defined as:

$$t_i^{(l)} = \frac{\exp(\sigma(\text{sim}(f_i M^{(l)}, e_{i,i'} M^{(l)})))}{\sum_{k \in N_j} \exp(\sigma(\text{sim}(f_i M^{(l)}, e_{i,k} M^{(l)})))} \quad (43)$$

here,  $M^{(l)}$  denotes a feature transformation matrix.  $e_{i,i'}$  denotes the hyperedge connecting nodes  $i$  and  $i'$ . By calculating these edge strength scores, the incidence matrix can be updated as follows:

$$\tilde{H}^{(l+1)} = D_V^{-1/2} (T_i^{(l)} \odot H^i) W D_e^{-1} (T_i^{(l)} \odot H^i)^T D_V^{-1/2} \quad (44)$$

where  $D_v$ ,  $D_e$  denote the node degree and edge degree matrices.  $T^{(l)}$  denotes the edge strength matrix and  $W$  is a learnable weight matrix. This function essentially adapts the interconnection of the nodes in  $H$  using the edge strengths calculated in  $T^{(l)}$ . Note that the edge strength changes depending on the layer  $l$ , as the feature similarities also change between the layer embeddings. The feature matrix is updated as follows:

$$\tilde{F}_i^{(l+1)} = \{\tilde{f}_{i,j}\}_{j=1}^P = \sigma((\tilde{H}^{(l+1)}) F_i^{(l)} P_i^{(l)}) \quad (45)$$

where  $\sigma$  is a nonlinear activation function and  $P_i^{(l)}$  denotes a learned projection matrix.

In a later publication, Liang et al. expanded their approach, utilizing a multiscale hypergraph model enhanced by cross-scale contrastive learning (Liang et al., 2024b).

Cai et al. introduced hierarchy into the hypergraph-GNN framework, by defining hypergraphs at the patch and tissue levels (Cai et al., 2024). On each of these levels, the hypergraph is defined using the K-NN algorithm in both the spatial and the feature space. Patch-level features were pooled hierarchically to obtain tissue-level embeddings. Using attention-based pooling, patch- and tissue-level hypernodes were pooled to obtain a WSI-level representation. The authors show their approach performs competitively for survival prediction tasks on various TCGA datasets.

Han et al. introduced another hierarchical hypergraph approach, defining hypergraph nodes on 10x resolution tissue patches based on the heterogeneity of a subordinate heterogeneous cell graph defined on 20x resolution patches (Han et al., 2024). The authors used a heterogeneous approach, where they separated the hypernodes into mix nodes, boundary nodes and random nodes, connected using mix hyperedges, boundary hyperedges and neighbor hyperedges. They used a heterogeneity-aware hypergraph learning layer to pass messages, additionally incorporating a self-attention mechanism. Their approach explicitly focuses on tissue regions where the cellular composition changes considerably, which play an important factor in the prognostic prediction. As such, their method showed competitive results for survival prediction on TCGA datasets.

Li et al. used hypergraph-based WSI modeling to predict spatial gene expression (Li et al., 2024a). They first extracted image features using a combination of convolutional block attention modules and vision transformers defined at multiple latent stages of the pipeline. The incidence matrix of their hypergraph was defined by a linear combination of both the spatial- and feature-wise distances between image patches:

$$\text{Inc\_Mat}(v_i, v_j) = \text{Norm}(\text{Dis}(v_i, v_j)) + \text{Norm}(\text{Pos}(v_i, v_j)) \quad (46)$$

where  $v_i$  and  $v_j$  are two image patches and  $\text{Dis}$ ,  $\text{Pos}$  denote the feature-wise distance and spatial distance, respectively.

Shi et al. extended the hypergraph-based modeling of WSIs by introducing masked hypergraph modeling, which generates masked views of the hypergraph during model training (Shi et al., 2024). The model was then trained to reconstruct the masked nodes and edges. This strategy aimed to increase the robustness and generalizability of the features learned by the GNN, enabling better performance on unseen samples. Starting with a set of 20x patches from the WSI along with corresponding ViT-based image features, the authors defined their hypergraph by clustering the feature space using k-means and spatially clustering using agglomerative clustering. The obtained clusters are then stacked to define a joined hypergraph.

Zhou et al. moved beyond hypergraphs defined on single WSIs to hypergraphs defined using multiple WSIs from the same patient (Zhou et al., 2024). Starting with WSI-patches and their corresponding features, the authors first defined two hypergraphs: A *intra-hypergraph*, defined using K-NN in both the feature- and coordinate space, and a *cross-hypergraph*, solely defined using K-NN in the feature space. The intra-hypergraph is WSI specific, while the cross-hypergraph contains hypernodes across WSIs. The features for the hypernodes in the cross-hypergraph are obtained by applying HGNN layers to the intra-hypergraphs, leading to spatially aware representations. A specifically designed multi-slice hypergraph block was then applied to the cross-hypergraph to exchange information across WSIs, which was subsequently pooled to obtain patient-level embeddings.

Finally, Lu et al. introduced reinforcement learning (RL) techniques to optimize their hypergraph-GNN framework for survival prediction (Lu et al., 2024). First, the authors clustered 20x resolution patch features and patch coordinates using K-means. A RL-agent was trained to select representative patches from each cluster, which are consequently used as hypernodes in the hypergraph. These hypernodes were connected using inter-cluster edges if they exist within the same K-means clusters. Clusters themselves were connected by intra-cluster edges. The authors used contrastive learning between different hypergraphs of the same slide to improve the data diversity and aid generalization.

#### 4.4.1. Discussion and future prospects

In this section, we described the increasing popularity of hypergraph-based modeling of histopathology slides. Interestingly, this approach has only been applied on a patch level, whereas we argue that hypergraph-based modeling might be very well suited for cell-level modeling. For example, cells can be organized as clusters that can have an important diagnostic context (Chandran et al., 2012). Such cell clusters could be modeled using hypergraphs, where homogeneous clusters are connected using a single hyperedge. Furthermore, there exist many other higher-order graph types, such as cellular complexes and combinatorial complexes (Papillon et al., 2023), which have not seen use in histopathology. We anticipate that these approaches will also be tested in a histopathological context. For example, using cellular complexes, different semantic tissue structures (e.g. hair follicles in the skin) can be modeled jointly with cells, but as separate graph entities.

### 5. Other trends

In addition to the trends explored above, we identified several other trends, which we will discuss more briefly.

#### 5.1. Graph transformers

In the last few years, graph neural networks have been combined with transformer architectures, which has given birth to the Graph Transformer modeling paradigm. Graph Transformers either use the positional embedding of the graph in the input to the transformer module, use the graph structure as a prior to build an attention mask for each input, or directly combine message passing layers with transformer blocks in the model architecture (Min et al., 2022). Graph transformers are especially suited for modeling long-distance relations in graphs, as they do not suffer from *oversmoothing*, where node representations become almost identical throughout the graph when using increased GNN layer depth and *oversquashing*, where the computational costs of adding GNN layers grow exponentially (Kreuzer et al., 2021). In histopathology, all three types of graph transformers have been used (Shi et al., 2023a; Hou et al., 2023; Lou et al., 2024; Wang et al., 2024). One major challenge in the application of graph transformers is their scalability, as the time- and memory complexity of the attention mechanism in Transformers grows exponentially( $O(|V|^2)$ ), where  $V$  is the number of nodes). This is especially a problem in cell graphs in histopathology, as these graphs often pass 10.000 nodes in size. Recently, efforts have been made to greatly mitigate this challenge of scalability (Rampášek et al., 2022; Shirzad et al., 2023; Wu et al., 2024), leading us to believe that the popularity of graph transformers in histopathology will continue to grow.

#### 5.2. Heterogeneous GNNs

Histopathology tissues are composed of different entities (e.g., glands, cells) which can be subdivided further into biological subtypes (e.g., celltypes). In a graph, we can explicitly encode these differences as different node types to obtain a *heterogeneous* graph. Similarly, different interactions between histopathological entities can also be explicitly defined as different types of edges. Heterogeneous GNNs are specifically designed to work on heterogeneous graphs, separately modeling the model parameters based on the node/edge type. Given the large number of different cell types and biological interactions present in tissues, heterogeneous GNNs are a natural model for histopathological tissues. Interestingly, only a handful of publications have made use of heterogeneous GNNs, where most have modeled entities on different hierarchies (e.g., 20x, 40x patches) as different node types (Hou et al., 2022b; Gupta et al., 2023; Bazargani et al., 2024), and 2 articles modeled different node types based on predicted cell types (Chan et al., 2023; Han et al., 2024).

#### 5.3. Self-supervised learning using GNNs

Due to the high costs of annotation in histopathology, adaptation of self-supervised learning (SSL) has been steadily growing for histopathology applications, particularly for image feature extraction. As such, they have become the basis for feature extraction models often used in GNN approaches. Recently, GNN-based SSL methods specifically designed for graphs have been introduced (Liu et al., 2022), which can be subcategorized into *contrastive*, *generative* or *auxiliary* approaches. Contrastive approaches aim to align the embeddings of graph entities (e.g., nodes, subgraphs, graphs) to each other, typically using mutual information (MI) maximization. Generative approaches learn to reconstruct either the graph structure or features defined on the graph based on latent space features or based on different views of the graph. Lastly, auxiliary approaches predict graph properties (e.g., degree) of the graph from its latent representation. In histopathology, contrastive learning has emerged as the predominant strategy for GNN-based SSL (Pina and Vilaplana, 2022; Liang et al., 2024b; Shao et al., 2024; Chi et al., 2024; Lu et al., 2024), while some recent publications have used generative approaches (Pan et al., 2024; Saeidi et al., 2024; Elforaici et al., 2024). The focus on contrastive and generative techniques, leaves room for auxiliary approaches in the field. For example, one could predict global graph statistics such as the density or average clustering coefficient, which could act as a regularization to gain more general-purpose graph-level features before introducing image-based information. Based on the rapid increase of GNN-based SSL approaches and the considerable shortage of labeled data in histopathology, we argue these techniques will continue to gain popularity.

#### 5.4. Semi-supervised learning using GNNs

Label scarcity is a big challenge in histopathology, as evident by the popularity of weakly supervised multiple instance learning strategies. *Semi-supervised* learning aims to utilize both labeled and unlabeled examples in partly labeled datasets, usually by inferring labels from unlabeled examples based on the labeled ones, deemed *label inference*. In a feature similarity-based graph, we can propagate the label of labeled examples to unlabeled examples. Often, edges are weighted based on the strength of the similarity, leading to a more nuanced label propagation process. GNNs implicitly perform label propagation, transferring both structural- and feature information from labeled nodes to unlabeled nodes via message passing, thus bypassing the need for explicit label propagation (Song et al., 2022). When predicting WSI-level labels using GNNs in histopathology, there is often a weakly supervised setting where the graph is labeled while the nodes themselves are not. This can be regarded as a form of semi-supervised learning, where the features learned from the unlabeled nodes are pooled to a labeled graph-level example. On node-level tasks, however, semi-supervised learning is still relatively unexplored in histopathology. We argue this type of learning can be useful for cell classification tasks where some cells have a cell type label (e.g., through the use of IHC-staining) while other nuclei are unlabeled (Turkki et al., 2016).

#### 5.5. Foundation models in computational histopathology

The rise of self-supervised learning as well as increased availability of histopathology datasets, has allowed the construction of very large deep neural networks, termed *Foundation models*, on huge amounts of (unlabeled) histopathology images (Vorontsov et al., 2023). These models can be used for effective feature extraction in a wide variety of tissue types. In both natural language processing and computer vision, there has been a move to foundation models that incorporate an even broader spectrum of modalities (video Christensen et al., 2023, audio Gardner et al., 2023, knowledge graphs Luo et al., 2023). Recent approaches have introduced medical texts in addition to image data (Lu et al., 2023; Huang et al., 2023), which allows associating image data

with medical texts and is thus very suitable for CBHIR applications. We argue that in histopathology and medical imaging in general, there will also be a move towards broader multimodality, especially given the number of different modalities available in the medical domain (WSI, IHC, MRI, CT, EHR, etc.). Graph models of WSIs can also be used as input in these models, encoding the topological information present in WSIs and correlating that with the image data. We also argue that in addition to the patch-based histopathology foundation models currently available, cellular foundation models could prove very useful. Although there exist some SSL-based embedding models specifically designed at the cellular level (Feng et al., 2021; Nakhli et al., 2024), these models are not yet scaled to the training data sizes we see for patch-level foundation models. Since the clinical relevance of cells is largely dependent on the context in which they exist, we argue that cellular foundation models could greatly benefit from GNN-based learning methods, encoding both topological, spatial, and image-based information for each cell. As such, we envision a cellular foundation model that uses a GNN-based encoder or at least uses encodings of spatial/topological features in the training process.

### 5.6. Generalization of GNNs in histopathology

Computational pathology models are plagued by large variations in scanner configurations, staining variability and tissue preparation between medical centers (Leo et al., 2016). Techniques such as stain normalization (Janowczyk et al., 2017), color augmentation (Tellez et al., 2018) and domain adaptation methods (Lafarge et al., 2017) have partly mitigated this issue, but there often remains a considerable performance gap when transferring models to multicenter settings. Graph Neural Networks, especially those based on cell graphs, may be well suited to help improve generalization capabilities of histopathology models, since cellular positions and thus the resulting graph structures themselves are invariant to differences in staining and scanner configuration.<sup>4</sup> No conclusive evidence is available to prove this generalization advantage GNNs might have for computational histopathology, as direct comparisons between CNNs/ViT and GNNs in histopathology generalization performance have not yet been conducted. As such, we argue this potential advantage should be studied in future research. To improve the generalization capabilities of GNNs themselves, many approaches have been proposed, ranging from self-supervised graph contrastive learning approaches to more simple augmentation methods such as edge perturbation or feature masking. We refer to Ding et al. (2022a) for a complete overview of data augmentation in GNNs.

## 6. Conclusion

In this review, we provided a comprehensive overview of the recent developments in the applications of GNNs in histopathology, which can be used for guiding new research in the field. We quantified the growth of different modeling paradigms in the use of GNNs in histopathology. Based on our quantification, we provided a comprehensive overview of several emerging subfields, including hierarchical graph models, adaptive graph structure learning, multimodal modeling with GNNs, and higher-order graph models. We also provided future directions for the field, including the use of topological deep learning, adaptive structure learning for heterophilious graphs, graph-based multimodal fusion with missing modalities, the use of graph transformer models, heterogeneous GNNs, self- and semi-supervised learning using GNNs and graph-based foundation models.

<sup>4</sup> Assuming that the nuclei segmentation is robust to these differences.

## CRediT authorship contribution statement

**Siemen Brussee:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Giorgio Buzzanca:** Writing – review & editing. **Anne M.R. Schrader:** Writing – review & editing, Project administration, Funding acquisition. **Jesper Kers:** Writing – review & editing, Supervision, Project administration.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Siemen Brussee reports financial support was provided by Hanarth Fund Foundation. Giorgio Buzzanca reports financial support was provided by European Commission. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This project has been funded by the Hanarth Foundation for AI in oncology, The Netherlands.

## Appendix A. Supplementary data

As supplementary material, we provide a complete table of all the articles researched with the corresponding categorizations.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.media.2024.103444>.

## References

- Abbas, S.F., Le Vuong, T.T., Kim, K., Song, B., Kwak, J.T., 2023. Multi-cell type and multi-level graph aggregation network for cancer grading in pathology images. *Med. Image Anal.* 90, 102936.
- Abdous, S., Abdollahzadeh, R., Rohban, M.H., 2023. KS-GNNExplainer: Global model interpretation through instance explanations on histopathology images. *arXiv preprint arXiv:2304.08240*.
- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11), 2274–2282.
- Acharya, V., Choi, D., Yener, B., Beamer, G., 2024. Prediction of tuberculosis from lung tissue images of diversity outbred mice using jump knowledge based cell graph neural network. *IEEE Access*.
- Adnan, M., Kalra, S., Tizhoosh, H.R., 2020. Representation learning of histopathology images using graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 988–989.
- Ahmed-Aristizabal, D., Armin, M.A., Denman, S., Fookes, C., Petersson, L., 2022. A survey on graph-based deep learning for computational histopathology. *Comput. Med. Imaging Graph.* 95, 102027.
- Ali, S., Veltri, R., Epstein, J.A., Christudass, C., Madabhushi, A., 2013. Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In: *Medical Imaging 2013: Digital Pathology*. Vol. 8676, SPIE, pp. 164–174.
- Anklan, V., Pati, P., Jaume, G., Bozorgtabar, B., Foncubierta-Rodriguez, A., Thiran, J.-P., Sibony, M., Gabrani, M., Goksel, O., 2021. Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs. *arXiv preprint arXiv:2103.03129*.
- Azadi, P., Suderman, J., Nakhli, R., Rich, K., Asadi, M., Kung, S., Oo, H., Keyes, M., Farahani, H., MacAulay, C., et al., 2023. ALL-IN: AL ocal GL obal graph-based DI stillatio N model for representation learning of gigapixel histopathology images with application in cancer risk assessment. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 765–775.
- Azher, Z.L., Fatemi, M., Lu, Y., Srinivasan, G., Diallo, A.B., Christensen, B.C., Salas, L.A., Kolling IV, F.W., Perreard, L., Palisoul, S.M., et al., 2023. Spatial omics driven crossmodal pretraining applied to graph-based deep learning for cancer pathology analysis. In: *PACIFIC SYMPOSIUM on BIOCOMPUTING 2024*. World Scientific, pp. 464–476.
- Bahade, S.S., Edwards, M., Xie, X., 2023. Cascaded graph convolution approach for nuclei detection in histopathology images. *J. Image Graph.* 11 (1).

- Bai, Y., Mi, Y., Su, Y., Zhang, B., Zhang, Z., Wu, J., Huang, H., Xiong, Y., Gong, X., Wang, W., 2022. A scalable graph-based framework for multi-organ histology image classification. *IEEE J. Biomed. Health Inf.* 26 (11), 5506–5517.
- Bakht, A.B., Javed, S., AlMarzouqi, H., Khandoker, A., Werghi, N., 2021. Colorectal cancer tissue classification using semi-supervised hypergraph convolutional network. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1306–1309.
- Bazargani, R., Fazli, L., Gleave, M., Goldenberg, L., Bashashati, A., Salcudean, S., 2024. Multi-scale relational graph convolutional network for multiple instance learning in histopathology images. *Med. Image Anal.* 96, 103197.
- Bazargani, R., Fazli, L., Goldenberg, L., Gleave, M., Bashashati, A., Salcudean, S., 2022. Multi-scale relational graph convolutional network for multiple instance learning in histopathology images. arXiv preprint arXiv:2212.08781.
- Behzadi, M.M., Madani, M., Wang, H., Bai, J., Bhardwaj, A., Tarakanova, A., Yamase, H., Nam, G.H., Nabavi, S., 2024. Weakly-supervised deep learning model for prostate cancer diagnosis and gleason grading of histopathology images. *Biomed. Signal Process. Control* 95, 106351.
- Bejnordi, B.E., Litjens, G., Hermans, M., Karssemeijer, N., van der Laak, J.A., 2015. A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images. In: Medical Imaging 2015: Digital Pathology. Vol. 9420, SPIE, pp. 99–104.
- Benkirane, H., Vakalopoulou, M., Christodoulidis, S., Garberis, I.-J., Michiels, S., Cournède, P.-H., 2022. Hyper-adac: Adaptive clustering-based hypergraph representation of whole slide images for survival analysis. In: Machine Learning for Health. PMLR, pp. 405–418.
- Bianchi, F.M., Grattarola, D., Alippi, C., 2020. Spectral clustering with graph neural networks for graph pooling. In: International Conference on Machine Learning. PMLR, pp. 874–883.
- Bilgin, C., Demir, C., Nagi, C., Yener, B., 2007. Cell-graph mining for breast tissue modeling and classification. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 5311–5314.
- Bontempo, G., Porrello, A., Bolelli, F., Calderara, S., Ficarra, E., 2023. DAS-MIL: Distilling across scales for MIL classification of histological WSIs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 248–258.
- Breen, J., Allen, K., Zucker, K., Orsi, N.M., Ravikumar, N., 2024. Multi-resolution histopathology patch graphs for ovarian cancer subtyping. arXiv preprint arXiv: 2407.18105.
- Cai, H., Yi, W., Huang, W., Wang, Z., Zhang, Y., Song, J., 2024. A hierarchical hypergraph attention network for survival analysis from pathological images. In: 2024 IEEE International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1–4.
- Chan, T.H., Cendra, F.J., Ma, L., Yin, G., Yu, L., 2023. Histopathology whole slide image analysis with heterogeneous graph representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15661–15670.
- Chandran, P.S., Byju, N.B., Deepak, R., Kumar, R.R., Sudhamony, S., Malm, P., Bengtsson, E., 2012. Cluster detection in cytology images using the cellgraph method. In: 2012 International Symposium on Information Technologies in Medicine and Education. Vol. 2, pp. 923–927. <http://dx.doi.org/10.1109/ITIME.2012.6291454>.
- Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020b. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.
- Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F., 2020a. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans. Med. Imaging* 41 (4), 757–770.
- Chi, C., Shi, H., Zhu, Q., Zhang, D., Shao, W., 2024. Spatially resolved gene expression prediction from histology via multi-view graph contrastive learning with HSIC-bottleneck regularization. arXiv preprint arXiv:2406.12229.
- Christensen, M., Vukadinovic, M., Yuan, N., Ouyang, D., 2023. Multimodal foundation models for echocardiogram interpretation. arXiv preprint arXiv:2308.15670.
- Ciga, O., Xu, T., Martel, A.L., 2022. Self supervised contrastive learning for digital histopathology. *Mach. Learn. Appl.* 7, 100198.
- Cui, Z., Henrickson, K., Ke, R., Wang, Y., 2019. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. Intell. Transp. Syst.* 21 (11), 4883–4894.
- Daneshvar, H., Samavi, R., 2022. Heterogeneous patient graph embedding in readmission prediction. In: AI.
- De, A., Mhatre, R., Tiwari, M., Chowdhury, A.S., 2022. Brain tumor classification from radiology and histopathology using deep features and graph convolutional network. In: 2022 26th International Conference on Pattern Recognition. ICPR, IEEE, pp. 4420–4426.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
- Di, D., Li, S., Zhang, J., Gao, Y., 2020. Ranking-based survival prediction on histopathological whole-slide images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 428–438.
- Di, D., Zhang, J., Lei, F., Tian, Q., Gao, Y., 2022a. Big-hypergraph factorization neural network for survival prediction from whole slide image. *IEEE Trans. Image Process.* 31, 1149–1160.
- Di, D., Zou, C., Feng, Y., Zhou, H., Ji, R., Dai, Q., Gao, Y., 2022b. Generating hypergraph-based high-order representations of whole-slide histopathological images for survival prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (5), 5800–5815.
- di Villaforesta, A.F., Magister, L.C., Barbiero, P., Liò, P., 2023. Digital histopathology with graph neural networks: Concepts and explanations for clinicians. arXiv preprint arXiv:2312.02225.
- Ding, S., Gao, Z., Wang, J., Lu, M., Shi, J., 2023b. Fractal graph convolutional network with MLP-mixer based multi-path feature fusion for classification of histopathological images. *Expert Syst. Appl.* 212, 118793.
- Ding, R., Rodriguez, E., Da Silva, A.C.A.L., Hsu, W., 2023a. Using graph neural networks to capture tumor spatial relationships for lung adenocarcinoma recurrence prediction. In: 2023 IEEE 20th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1–5.
- Ding, K., Wang, J., Li, J., Li, D., Liu, H., 2020. Be more with less: Hypergraph attention networks for inductive text classification. arXiv preprint arXiv:2011.00387.
- Ding, K., Xu, Z., Tong, H., Liu, H., 2022a. Data augmentation for deep graph learning: A survey. *ACM SIGKDD Explor. Newsl.* 24 (2), 61–77.
- Ding, K., Zhou, M., Wang, Z., Liu, Q., Arnold, C.W., Zhang, S., Metaxas, D.N., 2022b. Graph convolutional networks for multi-modality medical imaging: Methods, architectures, and clinical applications. arXiv preprint arXiv:2202.08916.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
- Dwivedi, C., Nofallah, S., Pouryahya, M., Iyer, J., Leidal, K., Chung, C., Watkins, T., Billin, A., Myers, R., Abel, J., et al., 2022. Multi stain graph fusion for multimodal integration in pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1835–1845.
- Ektefaie, Y., Dasoulas, G., Noori, A., Farhat, M., Zitnik, M., 2023. Multimodal learning with graphs. *Nat. Mach. Intell.* 5 (4), 340–350.
- Elforaici, M.E.A., Azzi, F., Trudel, D., Nguyen, B., Montagnon, E., Tang, A., Turcotte, S., Kadoury, S., 2024. Cell-Level GNN-based prediction of tumor regression grade in colorectal liver metastases from histopathology images. In: 2024 IEEE International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1–5.
- Fatemi, M., Feng, E., Sharma, C., Azher, Z., Goel, T., Ramwala, O., Palisoul, S.M., Barney, R.E., Perreard, L., Kolling, F.W., et al., 2023. Inferring spatial transcriptomics markers from whole slide images to characterize metastasis-related spatial heterogeneity of colorectal tumors: A pilot study. *J. Pathol. Inform.* 14, 100308.
- Feng, C., Vanderbilt, C., Fuchs, T., 2021. Nuc2vec: Learning representations of nuclei in histopathology images with contrastive loss. In: Medical Imaging with Deep Learning. PMLR, pp. 179–189.
- Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y., 2019. Hypergraph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33, pp. 3558–3565.
- Gallagher-Syed, A., Rossi, L., Rivellese, F., Pitzalis, C., Lewis, M., Barnes, M., Slabaugh, G., 2023. Multi-stain self-attention graph multiple instance learning pipeline for histopathology whole slide images. arXiv preprint arXiv:2309.10650.
- Gao, Z., Lu, Z., Wang, J., Ying, S., Shi, J., 2022. A convolutional neural network and graph convolutional network based framework for classification of breast histopathological images. *IEEE J. Biomed. Health Inf.* 26 (7), 3163–3173.
- Gao, J., Lyu, T., Xiong, F., Wang, J., Ke, W., Li, Z., 2020. MGNN: A multimodal graph neural network for predicting the survival of cancer patients. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1697–1700.
- Gao, Z., Shi, J., Wang, J., 2021. GQ-GCN: Group quadratic graph convolutional network for classification of histopathological images. In: Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer, pp. 121–131.
- Gao, R., Yuan, X., Ma, Y., Wei, T., Johnston, L., Shao, Y., Lv, W., Zhu, T., Zhang, Y., Zheng, J., et al., 2023. Predicting gene spatial expression and cancer prognosis: An integrated graph and image deep learning approach based on HE slides. bioRxiv.
- Gardner, J., Durand, S., Stoller, D., Bittner, R.M., 2023. Llark: A multimodal foundation model for music. arXiv preprint arXiv:2310.07160.
- Gindra, R.H., Zheng, Y., Green, E.J., Reid, M.E., Mazzilli, S.A., Merrick, D.T., Burks, E.J., Kolachalam, V.B., Beane, J.E., 2024. Graph perceiver network for lung tumor and bronchial premalignant lesion stratification from histopathology. *Am. J. Pathol.*
- Godson, L., Alemi, N., Nsengimana, J., Cook, G.P., Clarke, E.L., Treanor, D., Bishop, D.T., Newton-Bishop, J., Magee, D., 2023. Multi-level graph representations of melanoma whole slide images for identifying immune subgroups. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 85–96.
- Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* 58, 101563.
- Gu, Z., Wang, S., Rong, R., Zhao, Z., Wu, F., Zhou, Q., Wen, Z., Chi, Z., Fang, Y., Peng, Y., et al., 2024. CSGO: A deep learning pipeline for whole-cell segmentation in hematoxylin and eosin stained tissues. *Lab. Invest.* 102184.

- Guan, Y., Zhang, J., Tian, K., Yang, S., Dong, P., Xiang, J., Yang, W., Huang, J., Zhang, Y., Han, X., 2022. 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.
- Gupta, R.K., Kurian, N.C., Jeevan, P., Sethi, A., et al., 2023. Heterogeneous graphs model spatial relationships between biological entities for breast cancer diagnosis. arXiv preprint arXiv:2307.08132.
- Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.* 30.
- Han, M., Zhang, X., Yang, D., Liu, T., Kuang, H., Feng, J., Zhang, L., 2024. Multi-scale heterogeneity-aware hypergraph representation for histopathology whole slide images. arXiv preprint arXiv:2404.19334.
- Hasegawa, T., Arvidsson, H., Tudzarovski, N., Meinke, K., Sugars, R.V., Ashok Nair, A., 2023. Edge-based graph neural networks for cell-graph modeling and prediction. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 265–277.
- He, P., Qu, A., Xiao, S., Ding, M., 2023. A GNN-based network for tissue semantic segmentation in histopathology image. In: Journal of Physics: Conference Series. Vol. 2504, IOP Publishing, 012047.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.
- Hou, W., He, Y., Yao, B., Yu, L., Yu, R., Gao, F., Wang, L., 2023. Multi-scope analysis driven hierarchical graph transformer for whole slide image based cancer survival prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 745–754.
- Hou, W., Huang, H., Peng, Q., Yu, R., Yu, L., Wang, L., 2022a. Spatial-hierarchical graph neural network with dynamic structure learning for histological image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 181–191.
- Hou, W., Yu, L., Lin, C., Huang, H., Yu, R., Qin, J., Wang, L., 2022b. H<sup>2</sup>-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.
- Huang, Z., Bianchi, F., Yuksekogull, M., Montine, T.J., Zou, J., 2023. A visual-language foundation model for pathology image analysis using medical twitter. *Nat. Med.* 29 (9), 2307–2316.
- Ibañez, V., Szostak, P., Wong, Q., Korski, K., Abbasi-Sureshjani, S., Gomariz, A., 2024. Integrating multiscale topology in digital pathology with pyramidal graph convolutional networks. arXiv preprint arXiv:2403.15068.
- Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning. In: International Conference on Machine Learning. PMLR, pp. 2127–2136.
- Janowczyk, A., Basavanhally, A., Madabhushi, A., 2017. Stain normalization using sparse autoencoders (StaNoSA): application to digital pathology. *Comput. Med. Imaging Graph.* 57, 50–61.
- Jaume, G., Pati, P., Foncubierta-Rodriguez, A., Feroce, F., Scognamiglio, G., Anniciello, A.M., Thiran, J.-P., Goksel, O., Gabrani, M., 2020. Towards explainable graph representations in digital pathology. arXiv preprint arXiv:2007.00311.
- Jiang, N., Hou, Y., Zhou, D., Wang, P., Zhang, J., Zhang, Q., 2021. Weakly supervised gleason grading of prostate cancer slides using graph neural network. In: ICPGRAM. pp. 426–434.
- Jiang, Y., Ma, S., Xiao, W., Wang, J., Ding, Y., Zheng, Y., Sui, X., 2023. Predicting EGFR gene mutation status in lung adenocarcinoma based on multifeature fusion. *Biomed. Signal Process. Control* 84, 104786.
- Kim, S.Y., 2023. GNN-surv: Discrete-time survival prediction using graph neural networks. *Bioengineering* 10 (9), 1046.
- Kim, J., Wong, B., Ko, Y., Yi, M., 2024. MicroMIL: Graph-based contextual multiple instance learning for patient diagnosis using microscopy images. arXiv preprint arXiv:2407.21604.
- Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
- Kreuzer, D., Beaini, D., Hamilton, W., Létourneau, V., Tossou, P., 2021. Rethinking graph transformers with spectral attention. *Adv. Neural Inf. Process. Syst.* 34, 21618–21629.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25.
- Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Moeskops, P., Veta, M., 2017. Domain-adversarial neural networks to address the appearance variability of histopathology images. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3. Springer, pp. 83–91.
- Lee, J., Lee, I., Kang, J., 2019. Self-attention graph pooling. In: International Conference on Machine Learning. PMLR, pp. 3734–3743.
- Lee, J., Warner, E., Shaikhouni, S., Bitzer, M., Kretzler, M., Gipson, D., Pennathur, S., Bellovich, K., Bhat, Z., Gadegbeku, C., et al., 2023. Clustering-based spatial analysis (CluSA) framework through graph neural network for chronic kidney disease prediction using histopathology images. *Sci. Rep.* 13 (1), 12701.
- Leo, P., Lee, G., Shih, N.N., Elliott, R., Feldman, M.D., Madabhushi, A., 2016. Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. *J. Med. Imaging* 3 (4), 047502.
- Li, J., Chen, Y., Chu, H., Sun, Q., Guan, T., Han, A., He, Y., 2024b. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11323–11332.
- Li, B., Nelson, M.S., Savari, O., Loeffler, A.G., Eliceiri, K.W., 2022. Differentiation of pancreatic ductal adenocarcinoma and chronic pancreatitis using graph neural networks on histopathology and collagen fiber features. *J. Pathol. Inform.* 13, 100158.
- Li, Y., Shen, Y., Zhang, J., Song, S., Li, Z., Ke, J., Shen, D., 2023. A hierarchical graph V-Net with Semi-supervised Pre-training for histological image based breast cancer classification. *IEEE Trans. Med. Imaging*.
- Li, R., Yao, J., Zhu, X., Li, Y., Huang, J., 2018. Graph CNN for survival analysis on whole slide pathological images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 174–182.
- Li, B., Zhang, Y., Wang, Q., Zhang, C., Li, M., Wang, G., Song, Q., 2024a. Gene expression prediction from histology images via hypergraph neural networks. *Brief. Bioinform.* 25 (6), bbae500.
- Li, S., Zhou, J., Xu, T., Huang, L., Wang, F., Xiong, H., Huang, W., Dou, D., Xiong, H., 2021. Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 975–985.
- Liang, M., Jiang, X., Cao, J., Li, B., Wang, L., Chen, Q., Zhang, C., Zhao, Y., 2024a. CAF-AHGCN: context-aware attention fusion adaptive hypergraph convolutional network for human-interpretable prediction of gigapixel whole-slide image. *Vis. Comput.* 1–19.
- Liang, M., Jiang, X., Cao, J., Zhang, S., Liu, H., Li, B., Wang, L., Zhang, C., Jia, X., 2024b. HSG-MGAF Net: Heterogeneous subgraph-guided multiscale graph attention fusion network for interpretable prediction of whole-slide image. *Comput. Methods Programs Biomed.* 247, 108099.
- Lim, S., Jung, S.-W., 2022. A comparative study on graph construction methods for survival prediction using histopathology images. In: 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). IEEE, pp. 1–4.
- Liu, P., Ji, L., Ye, F., Fu, B., 2023. GraphLSurv: A scalable survival prediction network with adaptive and sparse structure learning for histopathological whole-slide images. *Comput. Methods Programs Biomed.* 231, 107433.
- Liu, Y., Jin, M., Pan, S., Zhou, C., Zheng, Y., Xia, F., Philip, S.Y., 2022. Graph self-supervised learning: A survey. *IEEE Trans. Knowl. Data Eng.* 35 (6), 5879–5900.
- Liu, M., Liu, Y., Xu, P., Cui, H., Ke, J., Ma, J., 2024. Exploiting geometric features via hierarchical graph pyramid transformer for cancer diagnosis using histopathological images. *IEEE Trans. Med. Imaging*.
- Lou, W., Li, G., Wan, X., Li, H., 2024. Cell graph transformer for nuclei classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38, pp. 3873–3881.
- Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Zhang, A., Le, L.P., et al., 2023. Towards a visual-language foundation model for computational pathology. arXiv preprint arXiv:2307.12914.
- Lu, C., Tang, C., Cui, Y., Liu, J., Li, Z., Zhang, X., Yao, S., Lin, H., Yang, D., Liu, Z., et al., 2024. Hypergraph-based multi-instance contrastive reinforcement learning for annotation-free pan-cancer survival prediction on whole slide histology images.
- Luo, Y., Yang, K., Hong, M., Liu, X., Nie, Z., 2023. Molfm: A multimodal molecular foundation model. arXiv preprint arXiv:2307.09484.
- Ma, M., Ren, J., Zhao, L., Tulyakov, S., Wu, C., Peng, X., 2021. Smil: Multimodal learning with severely missing modality. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35, pp. 2302–2310.
- Magister, L.C., Kazhdan, D., Singh, V., Liò, P., 2021. Gcexplainer: Human-in-the-loop concept-based explanations for graph neural networks. arXiv preprint arXiv: 2107.11889.
- Meng, X., Zou, T., 2023. Clinical applications of graph neural networks in computational histopathology: A review. *Comput. Biol. Med.* 164, 107201.
- Min, E., Chen, R., Bian, Y., Xu, T., Zhao, K., Huang, W., Zhao, P., Huang, J., Ananidou, S., Rong, Y., 2022. Transformer for graphs: An overview from architecture perspective. arXiv preprint arXiv:2202.08455.
- Mirabadi, A.K., Archibald, G., Darbandsari, A., Contreras-Sanz, A., Nakhli, R.E., Asadi, M., Zhang, A., Gilks, C.B., Black, P., Wang, G., et al., 2024. GRASP: Graph-structured pyramidal whole slide image representation. arXiv preprint arXiv: 2402.03592.
- Nair, A., Arvidsson, H., Gatica, J.E., Tudzarovski, N., Meinke, K., Sugars, R.V., 2022. A graph neural network framework for mapping histological topology in oral mucosal tissue. *BMC Bioinform.* 23 (1), 506.
- Nakhli, R., Rich, K., Zhang, A., Darbandsari, A., Shenasa, E., Hadjifaradji, A., Thiessen, S., Milne, K., Jones, S.J., McAlpine, J.N., et al., 2024. Volta: an environment-aware contrastive cell representation learning for histopathology. *Nature Commun.* 15 (1), 3942.
- Nakhli, R., Zhang, A., Mirabadi, A., Rich, K., Asadi, M., Gilks, B., Farahani, H., Bashashati, A., 2023. CO-PILOT: Dynamic top-down point cloud with conditional neighborhood aggregation for multi-gigapixel histopathology image representation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21063–21073.
- Ochoa, J.G.D., Mustafa, F.E., 2022. Graph neural network modelling as a potentially effective method for predicting and analyzing procedures based on patients' diagnoses. *Artif. Intell. Med.* 131, 102359.

- Palmal, S., Saha, S., Arya, N., Tripathy, S., 2024. CAGCL: Predicting short-and long-term breast cancer survival with cross-modal attention and graph contrastive learning. *IEEE J. Biomed. Health Inf.*
- Pan, L., Peng, Y., Li, Y., Wang, X., Liu, W., Xu, L., Liang, Q., Peng, S., 2024. SELECTOR: Heterogeneous graph network with convolutional masked autoencoder for multimodal robust prediction of cancer survival. *Comput. Biol. Med.* 172, 108301.
- Papillon, M., Sanborn, S., Hajij, M., Miolane, N., 2023. Architectures of topological deep learning: A survey on topological neural networks. arXiv preprint arXiv: 2304.10031.
- Pati, P., Jaume, G., Fernandes, L.A., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Riccio, D., Di Bonito, M., et al., 2020. Hact-net: 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. Springer, pp. 208–219.
- Pati, P., Jaume, G., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Fiche, M., Dubruc, E., Riccio, D., et al., 2022. Hierarchical graph representations in digital pathology. *Med. Image Anal.* 75, 102264.
- Pati, P., Karkampouna, S., Bonollo, F., Comperat, E., Radic, M., Spahn, M., Martinelli, A., Wartenberg, M., Kruithof-de Julio, M., Rapsomaniki, M.A., 2023. Multiplexed tumor profiling with generative AI accelerates histopathology workflows and improves clinical predictions. bioRxiv.
- Paul, S., Yener, B., Lund, A.W., 2024. C2P-GCN: Cell-to-patch graph convolutional network for colorectal cancer grading. arXiv preprint arXiv:2403.04962.
- Pina, O., Vilaplana, V., 2022. Self-supervised graph representations of wsis. In: Geometric Deep Learning in Medical Image Analysis. PMLR, pp. 107–117.
- Qiu, L., Kang, D., Wang, C., Guo, W., Fu, F., Wu, Q., Xi, G., He, J., Zheng, L., Zhang, Q., et al., 2022. Intratumor graph neural network recovers hidden prognostic value of multi-biomarker spatial heterogeneity. *Nat. Commun.* 13 (1), 4250.
- Rampášek, L., Galkin, M., Dwivedi, V.P., Luu, A.T., Wolf, G., Beaini, D., 2022. Recipe for a general, powerful, scalable graph transformer. *Adv. Neural Inf. Process. Syst.* 35, 14501–14515.
- Reynolds, H.M., Williams, S., Zhang, A.M., Ong, C.S., Rawlinson, D., Chakravorty, R., Mitchell, C., Haworth, A., 2014. Cell density in prostate histopathology images as a measure of tumor distribution. In: Medical Imaging 2014: Digital Pathology. Vol. 9041, SPIE, pp. 180–187.
- Rocheteau, E., Tong, C., Veličković, P., Lane, N., Liò, P., 2021. Predicting patient outcomes with graph representation learning. arXiv preprint arXiv:2101.03940.
- Saeidi, N., Karshenas, H., Shouhtarian, B., Hatamikia, S., Woitek, R., Mahbod, A., 2024. Breast histopathology image retrieval by attention-based adversarially regularized variational graph autoencoder with contrastive learning-based feature extraction. arXiv preprint arXiv:2405.04211.
- Santoiemma, P.P., Powell, Jr., D.J., 2015. Tumor infiltrating lymphocytes in ovarian cancer. *Cancer Biol. Ther.* 16 (6), 807–820.
- Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G., 2008. The graph neural network model. *IEEE Trans. Neural Netw.* 20 (1), 61–80.
- Schütt, K.T., Sauceda, H.E., Kindermans, P.-J., Tkatchenko, A., Müller, K.-R., 2018. Schnet—a deep learning architecture for molecules and materials. *J. Chem. Phys.* 148 (24).
- Shao, W., Shi, Y., Zhang, D., Zhou, J., Wan, P., 2024. Tumor micro-environment interactions guided graph learning for survival analysis of human cancers from whole-slide pathological images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11694–11703.
- Sharma, H., Zerbe, N., Lohmann, S., Kayser, K., Hellwich, O., Hufnagl, P., 2015. A review of graph-based methods for image analysis in digital histopathology. *Diagn. Pathol.* 1 (1).
- Shi, J., Shu, T., Wu, K., Jiang, Z., Zheng, L., Wang, W., Wu, H., Zheng, Y., 2024. Masked hypergraph learning for weakly supervised histopathology whole slide image classification. *Comput. Methods Programs Biomed.* 253, 108237.
- Shi, J., Tang, L., Gao, Z., Li, Y., Wang, C., Gong, T., Li, C., Fu, H., 2023a. MG-Trans: Multi-scale graph transformer with information bottleneck for whole slide image classification. *IEEE Trans. Med. Imaging*.
- Shi, J., Tang, L., Li, Y., Zhang, X., Gao, Z., Zheng, Y., Wang, C., Gong, T., Li, C., 2023b. A structure-aware hierarchical graph-based multiple instance learning framework for pT staging in histopathological image. *IEEE Trans. Med. Imaging*.
- Shirzad, H., Velingker, A., Venkatachalam, B., Sutherland, D.J., Sinop, A.K., 2023. Exphormer: Sparse transformers for graphs. arXiv preprint arXiv:2303.06147.
- Shu, T., Shi, J., Sun, D., Jiang, Z., Zheng, Y., 2024. SlideGCD: Slide-based graph collaborative training with knowledge distillation for whole slide image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 470–480.
- Sims, J., Grabsch, H.I., Magee, D., 2022. Using hierarchically connected nodes and multiple GNN message passing steps to increase the contextual information in cell-graph classification. In: MICCAI Workshop on Imaging Systems for GI Endoscopy. Springer, pp. 99–107.
- Song, Z., Yang, X., Xu, Z., King, I., 2022. Graph-based semi-supervised learning: A comprehensive review. *IEEE Trans. Neural Netw. Learn. Syst.* 34 (11), 8174–8194.
- Su, R., He, H., Sun, C., Wang, X., Liu, X., 2023. Prediction of drug-induced hepatotoxicity based on histopathological whole slide images. *Methods* 212, 31–38.
- Sudharshan, P., Petitjean, C., Spanhol, F., Oliveira, L.E., Heutte, L., Honeine, P., 2019. Multiple instance learning for histopathological breast cancer image classification. *Expert Syst. Appl.* 117, 103–111.
- Sureka, M., Patil, A., Anand, D., Sethi, A., 2020. Visualization for histopathology images using graph convolutional neural networks. In: 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering. BIBE, IEEE, pp. 331–335.
- Tang, S., Li, B., Yu, H., 2019. ChebNet: Efficient and stable constructions of deep neural networks with rectified power units using chebyshev approximations. arXiv preprint arXiv:1911.05467.
- Tellez, D., Balkenhol, M., Karssemeijer, N., Litjens, G., van der Laak, J., Ciompi, F., 2018. H and E stain augmentation improves generalization of convolutional networks for histopathological mitosis detection. In: Medical Imaging 2018: Digital Pathology. Vol. 10581, SPIE, pp. 264–270.
- Tendle, A., Hasan, M.R., 2021. A study of the generalizability of self-supervised representations. *Mach. Learn. Appl.* 6, 100124.
- Thang, D.C., Dat, H.T., Tam, N.T., Jo, J., Hung, N.Q.V., Aberer, K., 2022. Nature vs. nurture: Feature vs. structure for graph neural networks. *Pattern Recognit. Lett.* 159, 46–53.
- Turkki, R., Linder, N., Kovanen, P.E., Pellinen, T., Lundin, J., 2016. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *J. Pathol. Inform.* 7 (1), 38.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903.
- Veličković, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D., 2018. Deep graph infomax. arXiv preprint arXiv:1809.10341.
- Verontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Liu, S., Mathieu, P., van Eck, A., Lee, D., Viret, J., et al., 2023. Virchow: A million-slide digital pathology foundation model. arXiv preprint arXiv:2309.07778.
- Wang, Z., Fan, K., Zhu, X., Liu, H., Meng, G., Wang, M., Li, A., 2023c. Cross-domain nuclei detection in histopathology images using graph-based nuclei feature alignment. *IEEE J. Biomed. Health Inf.*
- Wang, H., Huang, G., Zhao, Z., Cheng, L., Juncker-Jensen, A., Nagy, M.L., Lu, X., Zhang, X., Chen, D.Z., 2023a. CCF-GNN: A unified model aggregating appearance, microenvironment, and topology for pathology image classification. *IEEE Trans. Med. Imaging*.
- Wang, Z., Li, J., Pan, Z., Li, W., Sisk, A., Ye, H., Speier, W., Arnold, C.W., 2021b. Hierarchical graph pathomic network for progression free survival prediction. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer, pp. 227–237.
- Wang, S., Rong, R., Zhou, Q., Yang, D.M., Zhang, X., Bishop, J., Chi, Z., Wilhelm, C.J., Zhang, S., et al., 2023b. Deep learning of cell spatial organizations identifies clinically relevant insights in tissue images. *Nat. Commun.* 14 (1), 7872.
- Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Huang, J., Yang, W., Han, X., 2021a. Transpath: Transformer-based self-supervised learning for histopathological image classification. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer, pp. 186–195.
- Wang, K., Zheng, F., Cheng, L., Dai, H.-N., Dou, Q., Qin, J., 2024. Breast cancer classification from digital pathology images via connectivity-aware graph transformer. *IEEE Trans. Med. Imaging*.
- Wojciechowska, M., Malacrina, S., Garcia Martin, N., Fehri, H., Rittscher, J., 2021. Early detection of liver fibrosis using graph convolutional networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 217–226.
- Wu, Q., Zhao, W., Yang, C., Zhang, H., Nie, F., Jiang, H., Bian, Y., Yan, J., 2024. Simplifying and empowering transformers for large-graph representations. *Adv. Neural Inf. Process. Syst.* 36.
- Wu, Y., Zuo, Y., Zhu, Q., Sheng, J., Zhang, D., Shao, W., 2023. Transfer learning-assisted survival analysis of breast cancer relying on the spatial interaction between tumor-infiltrating lymphocytes and tumors. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 612–621.
- Xiang, X., Wu, X., 2021. Multiple instance classification for gastric cancer pathological images based on implicit spatial topological structure representation. *Appl. Sci.* 11 (21), 10368.
- Xiao, W., Jiang, Y., Yao, Z., Zhou, X., Sui, X., Zheng, Y., 2022. LAD-GCN: Automatic diagnostic framework for quantitative estimation of growth patterns during clinical evaluation of lung adenocarcinoma. *Front. Physiol.* 13, 946099.
- Xie, Y., Niu, G., Da, Q., Dai, W., Yang, Y., 2022b. Survival prediction for gastric cancer via multimodal learning of whole slide images and gene expression. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine. BIBM, IEEE, pp. 1311–1316.
- Xie, C., Vanderbilt, C., Feng, C., Ho, D., Campanella, G., Egger, J., Plodkowski, A., Girshman, J., Sawan, P., Arbour, K., et al., 2022a. Computational biomarker predicts lung ICI response via deep learning-driven hierarchical spatial modelling from H&E.

- Xing, X., Ma, Y., Jin, L., Sun, T., Xue, Z., Shi, F., Wu, J., Shen, D., 2021. A multi-scale graph network with multi-head attention for histopathology image diagnosis. In: COMPAY 2021: The Third MICCAI Workshop on Computational Pathology.
- Xu, K., Hu, W., Leskovec, J., Jegelka, S., 2018. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826.
- Xu, J., Xin, J., Shi, P., Wu, J., Cao, Z., Feng, X., Zheng, N., 2023. Lymphoma recognition in histology image of gastric mucosal biopsy with prototype learning. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, pp. 1–4.
- Yang, Z., Zhang, Y., Zhuo, L., Sun, K., Meng, F., Zhou, M., Sun, J., 2024. Prediction of prognosis and treatment response in ovarian cancer patients from histopathology images using graph deep learning: a multicenter retrospective study. Eur. J. Cancer 199, 113532.
- Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J., 2019. Gnnexplainer: Generating explanations for graph neural networks. Adv. Neural Inf. Process. Syst. 32.
- Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W.L., Leskovec, J., 2018a. Graph convolutional neural networks for web-scale recommender systems. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 974–983.
- Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J., 2018b. Hierarchical graph representation learning with differentiable pooling. Adv. Neural Inf. Process. Syst. 31.
- Yu, J., Xu, T., He, R., 2021. Towards the explanation of graph neural networks in digital pathology with information flows. arXiv preprint arXiv:2112.09895.
- Zhang, M., Dong, B., Li, Q., 2022. MS-GWNN: multi-scale graph wavelet neural network for breast cancer diagnosis. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1–5.
- Zhang, J., Hua, Z., Yan, K., Tian, K., Yao, J., Liu, E., Liu, M., Han, X., 2021. Joint fully convolutional and graph convolutional networks for weakly-supervised segmentation of pathology images. Med. Image Anal. 73, 102183.
- Zhang, Z., Zhao, Y., Duan, J., Liu, Y., Zheng, H., Liang, D., Zhang, Z., Li, Z.-C., 2024. Pathology-genomic fusion via biologically informed cross-modality graph learning for survival analysis. arXiv preprint arXiv:2404.08023.
- Zhao, J., Wang, X., Shi, C., Hu, B., Song, G., Ye, Y., 2021. Heterogeneous graph structure learning for graph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35, pp. 4697–4705.
- Zhao, W., Wang, S., Yeung, M., Niu, T., Yu, L., 2023. MuLT: Multi-task graph-transformer with task-aware knowledge injection and domain knowledge-driven pooling for whole slide image analysis. arXiv preprint arXiv:2302.10574.
- Zheng, Y., Conrad, R.D., Green, E.J., Burks, E.J., Betke, M., Beane, J.E., Kolachalam, V.B., 2023. Graph attention-based fusion of pathology images and gene expression for prediction of cancer survival. bioRxiv.
- Zheng, Y., Jiang, B., Shi, J., Zhang, H., Xie, F., 2019. Encoding histopathological wsis using gnn for scalable diagnostically relevant regions retrieval. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. Springer, pp. 550–558.
- Zheng, Y., Jiang, Z., Xie, F., Shi, J., Zhang, H., Huai, J., Cao, M., Yang, X., 2020. Diagnostic regions attention network (dra-net) for histopathology wsi recommendation and retrieval. IEEE Trans. Med. Imaging 40 (3), 1090–1103.
- Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Shi, J., Xue, C., 2021. Histopathology wsi encoding based on gcns for scalable and efficient retrieval of diagnostically relevant regions. arXiv preprint arXiv:2104.07878.
- Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., Yang, Y., Zhao, Y., 2022. Multi-modal graph learning for disease prediction. IEEE Trans. Med. Imaging 41 (9), 2207–2216.
- Zhong, Z., Li, C.-T., Pang, J., 2023. Hierarchical message-passing graph neural networks. Data Min. Knowl. Discov. 37 (1), 381–408.
- Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.-A., Rajpoot, N., 2019. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. p. 0.
- Zhou, H., Tian, Z., Han, X., Du, S., Gao, Y., 2024. Crcc metastasis prediction via exploring high-order correlations on multiple WSIs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 145–154.
- Zhu, Y., Xu, W., Zhang, J., Du, Y., Zhang, J., Liu, Q., Yang, C., Wu, S., 2021a. A survey on graph structure learning: Progress and opportunities. arXiv preprint arXiv:2103.03036.
- Zhu, Y., Xu, W., Zhang, J., Liu, Q., Wu, S., Wang, L., 2021b. Deep graph structure learning for robust representations: A survey. arXiv preprint arXiv:2103.03036 14.
- Zhu, D., Zhang, F., Wang, S., Wang, Y., Cheng, X., Huang, Z., Liu, Y., 2020. Understanding place characteristics in geographic contexts through graph convolutional neural networks. Ann. Am. Assoc. Geogr. 110 (2), 408–420.
- Zitnik, M., Agrawal, M., Leskovec, J., 2018. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics 34 (13), i457–i466.
- Zuo, Y., Wu, Y., Lu, Z., Zhu, Q., Huang, K., Zhang, D., Shao, W., 2022. Identify consistent imaging genomic biomarkers for characterizing the survival-associated interactions between tumor-infiltrating lymphocytes and tumors. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 222–231.