

A Survey on Graph-Based Deep Learning for Computational Histopathology

David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, Lars Petersson

Abstract—With the remarkable success of representation learning for prediction problems, we have witnessed a rapid expansion of the use of machine learning and deep learning for the analysis of digital pathology and biopsy image patches. However, learning over patch-wise features using convolutional neural networks limits the ability of the model to capture global contextual information and comprehensively model tissue composition. The phenotypical and topological distribution of constituent histological entities play a critical role in tissue diagnosis. As such, graph data representations and deep learning have attracted significant attention for encoding tissue representations, and capturing intra- and inter- entity level interactions. In this review, we provide a conceptual grounding for graph analytics in digital pathology, including entity-graph construction and graph architectures, and present their current success for tumor localization and classification, tumor invasion and staging, image retrieval, and survival prediction. We provide an overview of these methods in a systematic manner organized by the graph representation of the input image, scale, and organ on which they operate. We also outline the limitations of existing techniques, and suggest potential future research directions in this domain.

Index Terms—Digital pathology, Cancer classification, Cell-graph, Tissue-graph, Hierarchical graph representation, Graph Convolutional Networks, Deep learning.

I. INTRODUCTION

RECENT advances in deep learning techniques have rapidly transformed these approaches into the methodology of choice for analyzing medical images, and in particular for histology image classification problems [1]. Because of the increasing availability of large scale high-resolution whole-slide images (WSI) of tissue specimens, digital pathology and microscopy have become appealing application areas for deep learning algorithms. Given wide variations in pathology and the often time-consuming diagnosis process, clinical experts have begun to benefit from computer-aided detection and diagnosis methods capable of learning features that optimally represent the data [2]. This thorough survey serves as an accurate guide to biomedical engineering and clinical research communities interested in discovering the tissue composition-to-functionality relationship using image-to-graph translation and deep learning.

There are several review papers available that analyse the benefits of deep learning for providing reliable support for microscopic and digital pathology diagnosis and treatment decisions [1], [3]–[6], and specifically for cancer diagnosis [7]. Compared to other medical fields such as dermatology, ophthalmology, neurology, cardiology, and radiology,

D. Ahmedt-Aristizabal, A. Armin and L. Petersson are with the Imaging and Computer Vision group, CSIRO Data61, Canberra, Australia. (Corresponding author: david.ahmedtaristizabal@data61.csiro.au)

D. Ahmedt-Aristizabal, S. Denman and C. Fookes are with SAI/T, Queensland University of Technology, Brisbane, Australia.

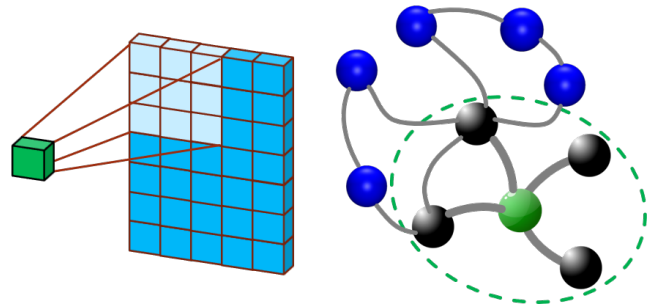


Fig. 1: Traditional CNNs excel at modelling local relations in grid representation, where the topology of the neighborhood is constant (Left). GCNs can take into account different neighbouring relations (global relation) by going beyond the local pixel neighbourhoods used by convolutions. On a graph, the neighbours of a node are unordered and variable in size (Right).

digital pathology and microscopy is one of the most dominant medical applications of deep learning. One driving force behind innovation in computational pathology has been the introduction of grand challenges (*e.g.* NuCLS [8], BACH [9], MoNuSeg [10]). Developed techniques that offer decision support to human pathologists have shown bright prospects for detecting, segmenting, and classifying the cell and nucleus; and detecting and classifying diseases such as cancer.

Deep learning techniques such as convolutional neural networks (CNNs) have demonstrated success in extracting image-level representations, however, they are inefficient when dealing with relation-aware representations. Modern deep learning variations of graph neural networks (GNNs) have made a significant impact in many technological domains for describing relationships. Graphs, by definition, capture relationships between entities and can thus be used to encode relational information between variables [11]. As a result, special emphasis has been placed on the generalisation of GNNs into non-structured and structured scenarios. Traditional CNNs analyse local areas based on fixed connectivity (determined by the convolutional kernel), leading to limited performance, and difficulty in interpreting the structures being modeled. Graphs, on the other hand, offer more flexibility to analyse unordered data by preserving neighboring relations. This difference is illustrated in Fig. 1.

The adaptation of deep learning from images to graphs has received increased attention, leading to a new cross-domain field of graph-based deep learning which seeks to learn informative representations of graphs in an end-to-end manner. This field has exhibited remarkable success for various tasks as discussed by recent surveys on graph deep learning frameworks

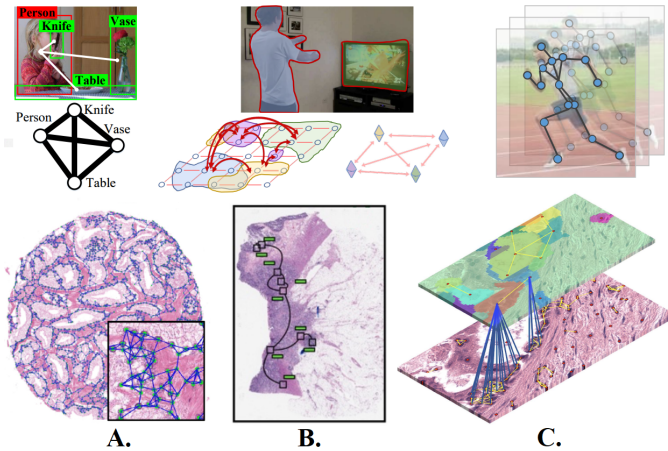


Fig. 2: **Top:** Graph-based representation of images for relation-aware human-object interaction, image segmentation, and human pose estimation (left-to-right). Images adapted from [12]–[14]. **Bottom:** **A.** Cell-graph representation for prostate cancer. **B.** Tissue-graph representation for colorectal cancer. **C.** Hierarchical cell-to-tissue graph representation for breast cancer. Images adapted from [15]–[17].

and their applications [11], [18]–[20]. Graph embeddings have appeared in computer vision tasks where graphs can efficiently define relationships between objects, or for the purpose of graph-structured image analysis. Interesting results have been obtained for object detection, semantic segmentation, skeleton-based action recognition, image classification and human-object interaction tasks as illustrated in Fig. 2 (Top).

Medical applications have benefited from rapid progress in the field of computer vision and GNNs. The development of GNNs has seen the application of deep learning methods to GNNs, such as graph convolutional networks (GCNs). These models have been proposed as a powerful tool to model functional and anatomical structures, brain electrical activity, and segmentation of the vasculature system and organs [21].

Histological images depict the micro-anatomy of a tissue sample, and pathologists use histological images to make diagnoses based on morphological changes in tissues, the spatial relationship between cells, cell density, and other factors. Graph-based methods, which can capture geometrical and topological properties, are able to model cell-level information and overall tissue micro-architecture. Prior to the advent of deep learning, numerous approaches for processing histopathological images as graphs were investigated [22]. These methods used classical machine learning approaches, which are less accurate for graph classification compared to GCNs. The capabilities of graph-based deep learning, which bridges the gap between deep learning methods and traditional cell graphs for disease diagnosis, are yet to be sufficiently investigated.

In this survey, we analyse how graph embeddings are employed in histopathology diagnosis and analysis. While graphs are not directly expressed within this data, they can efficiently describe relationships between tissue regions and cells. This setting offers a very different task for GNNs in comparison to analysis of unstructured data such as electro-

physiological and neuroimaging recordings where the data can be directly mapped to a graph [21]. Selected samples of graph representations in digital pathology (cell-graph, patch-graph, tissue-graph and cell-tissue representation) used to capture and learn relevant morphological regions that will be covered in this review are illustrated in Fig. 2 (Bottom).

This survey offers a comprehensive overview of preprocessing, graph models and explainability tools used in computational pathology, highlighting the capability of GNNs to detect and associate key tissue architectures, regions of interest, and their interdependence. Although some papers have surveyed conventional cell graphs with handcrafted features to characterize the entities [22], [23], and others have briefly touched upon the benefits of GCNs in biology and medicine [24], to the best of our knowledge, no systematic review exists that presents and discusses all relevant works concerning graph-based representations and deep learning models for computational pathology.

A. Why graph-based deep learning for characterizing diseases through histopathology slides?

Deep learning has increased the potential of medical image analysis by enabling the discovery of morphological and textural representations in images solely from the data. Although CNNs have shown impressive performance in the field of histopathology analysis, they are unable to capture complex neighborhood information as they analyse local areas determined by the convolutional kernel. To extract interaction information between objects, a CNN needs to reach sufficient depth by stacking multiple convolutional layers, which is inefficient. This leads to limitations in the performance and interpretability of the analysis of anatomical structures and microscopic samples.

Graph convolutional networks (GCNs) are a deep learning-based method that operate over graphs, and are becoming increasingly useful for medical diagnosis and analysis [21]. GCNs can better exploit irregular relationships and preserve neighboring relations compared with CNN-based models [11]. Below we outline the reasons why current research in histopathology has shifted the analytical paradigm from pixel to entity-graph processing:

- 1) The potential correlations among images are ignored during traditional CNN feature learning, however, a GCN can be introduced to estimate the dependencies between images and enhance the discriminative ability of CNN features [25].
- 2) CNNs have been commonly used for the analysis of whole slide images (WSI) by classifying fixed-sized biopsy image patches using fixed fusion rules such as averaging features or class scores, or weighted averaging with learnable weights to obtain an image-level classification score. Aggregation using a CNN also includes excessive whitespace, putting undue reliance on the orientation and location of the tissue segment. Even though CNN-based models have practical merits through considering important patches for prediction, they dismiss the spatial relationships between patches, or global contextual information. Architectures are required

to be capable of dealing with size and shape variation in region-of-interests (ROIs), and must encode the spatial context of individual patches and their collective contribution to the diagnosis, which can be addressed with graph-based representations [26], [27].

- 3) A robust computer-aided detection system should be able to capture multi-scale contextual features in tissues, which can be difficult with traditional CNN-based models. A pathological image can be transformed into a graph representation to capture the cellular morphology and topology (cell-graph) [28], and the attributes of the tissue parts and their spatial relationships (tissue-graph) [17], [29].
- 4) Graph representations can enhance the interpretation of the final representation by modeling relations among different regions of interest. Graph-based models offer a new way to verify existing observations in pathology. Attention mechanisms with GCNs, for example, highlight informative nuclei and inter-nuclear interactions, allowing the production of interpretable maps of tissue images displaying the contribution of each nucleus and its surroundings to the final diagnosis [30].
- 5) By incorporating any task-specific prior pathological information, an entity-graph can be customized in various ways. As a result, pathology-specific interpretability and human-machine co-learning are enabled by the graph format [31].
- 6) GCNs are a complimentary method to CNNs for morphological feature extraction, and they can be employed instead of, or in addition to CNNs during multimodal fusion for fine-grained patient stratification [32].

B. Contribution and organisation

Compared to other recent reviews on traditional deep learning in histopathology slides, our manuscript captures the current efforts relating to entity-graphs and recent advancements in GCNs for characterizing diseases and pathology tasks.

Papers included in the survey are obtained from various journals, conference proceedings and open-access repositories. Table I outlines the applications that were addressed across all reviewed publications. It is noted that breast cancer analysis constitutes the major application in digital pathology that has been analyzed using graph-based deep learning techniques.

This review is divided into three major sections. In Section II we provide a technical overview of the prevailing tools for entity-graph representation and graph architectures used in accelerating digital pathology research. In Section III we introduce the current applications of deep graph representation learning and cluster these proposals based on the graph construction (cell-graph, patch-graph, tissue-graph, hierarchical graph) and feature level fusion methods followed by the task or organ on which they operate. Finally, Section IV highlights open problems and perspectives regarding the shifting analytical paradigm from pixel to entity-based processing. Specifically, we discuss the topics of graph construction, embedding expert knowledge, complexity of graph models, training paradigms, and graph model interpretability.

TABLE I: Summary of applications of graph-based deep learning in histopathology covered in this survey.

Application	#Applications	Reference
Breast cancer	11	[17], [26], [28]–[31], [33]–[37]
Colorectal cancer	6	[16], [27], [38]–[41]
Prostate cancer	3	[15], [30], [42]
Lung cancer	3	[43]–[45]
Cervical cancer	2	[25], [46]
Lymphoma	1	[16]
Skin cancer	1	[47]
Renal cancer	1	[32]
Total	28	

II. GRAPH REPRESENTATION LEARNING IN DIGITAL PATHOLOGY: BACKGROUND

Translating patient histopathological images into graphs to encode the spatial context of cells and tissues for a given patient has been used to improve prediction accuracy of various pathology tasks. Graph representations followed by GNN-based models and interpretability approaches allows pathologists to directly comprehend and reason for the outcomes. GNNs can also serve a variety of prediction purposes by adapting different designs, such as performing node-level and graph-level predictions.

A standard entity-graph based pathological workflow requires several phases, such as node and graph topology definition, as well as the choice of GNN architecture. In this section, we provide technical insights of these phases that are required for graph analytics in computational pathology: (1) Graph representation (entity, embeddings and edges definition); (2) Graph models (graph structures for processing graph-structured); and (3) Explainability (a set of interpretation methodologies such as model-based and post-hoc interpretability). A traditional framework with aforementioned phases is illustrated in Fig. 3. A deep analysis of each GNN model can be found in survey papers that deal with graph architectures [11], [20].

A. Histopathology graph representation

1) *Preliminaries:* A graph can be represented by $G = (\mathcal{V}, \mathcal{E}, W)$, where V is a vertex set with $|\mathcal{V}| = n$ nodes and \mathcal{E} denotes the set of edges connecting these nodes. Data in \mathcal{V} can be represented by a feature matrix $X \in \mathbb{R}^{n \times d}$, where n and d denote the input feature dimensions. $W \in \mathbb{R}^{n \times n}$ is a binary or weighted adjacency matrix describing the connections between any two nodes in \mathcal{V} , in which the importance of the connections between the i -th and the j -th nodes is measured by the entry W in the i -th row and j -th column, and denoted w_{ij} . Commonly used methods to determine the entries, w_{ij} , of W include Pearson correlation-based graph, the K-nearest neighbor (KNN) method, and the distance-based graph [48]. In general, GNNs learn a feature transformation function for X and produce output $Z \in \mathbb{R}^{n \times d'}$, where d' denotes the output feature dimension.

Presented graph methods in digital pathology typically use data in one of two forms. Whole slide images (WSI), also known as virtual microscopy, are high-resolution images generated by combining many smaller image tiles or strips and tiling them to form a single image. Tissue microarrays (TMAs)

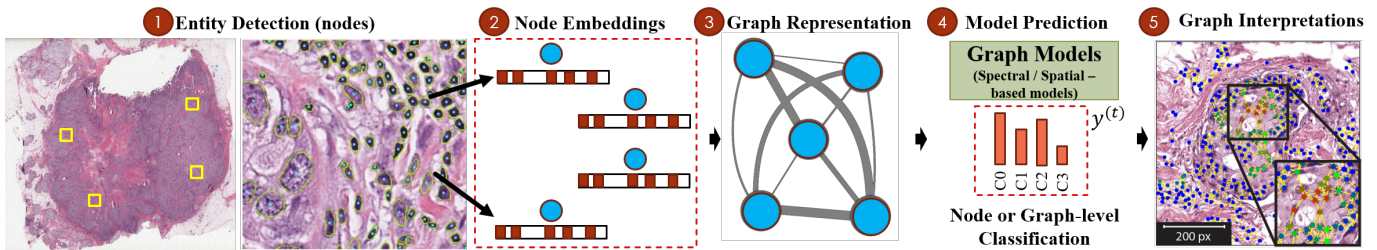


Fig. 3: Overview of a standard graph-based workflow in computational pathology. The WSI image is first transformed into one or more graphs. **1.** The entities can be nuclei, patches or tissue regions. **2.** Node features comprise handcrafted or deep learning features to characterize the entities. **3.** The edges encode intrinsic relationships (spatial or semantic) among the entities. **4.** Graph encoding and classification (node-level or graph-level prediction): the graph representation is processed using GNNs and its variants such as ChebNet, GCN, GraphSAGE, GAT, and GIN, including different graph pooling strategies (global or hierarchical pooling). **5.** Graph interpretations: a set of GNN model interpretability tools such as graph attentions or post-hoc graph explainers (*e.g.* GNNExplainer and GraphGrad-CAM.)

consist of paraffin blocks produced by extracting cylindrical tissue cores and inserting them into a single recipient block (microarray) in a precisely spaced pattern. With this technique, up to 1000 tissue cores can be assembled in a single paraffin block to allow multiplex histological analysis.

2) *Graph construction:* Graph representations have been used in digital pathology for multiple tasks where a histology image is described as an entity-graph, and nodes and edges of a graph denote biological entities and inter-entity interactions respectively. The entities can be biologically-defined such as nuclei and tissue regions, or can be defined patch-wise. Therefore, constructing an entity-graph for graph analytics in computational pathology demands the following pre-processing steps.

a) *Node definition:* WSI usually includes significant non-tissue regions. To identify tissue regions the foreground is segmented with Gaussian smoothing and OTSU thresholding [49].

One of the most common graph representation, cell-graphs, requires model training and fine-tuning for cell detection or segmentation. To detect nuclei several methods have been used such as Hover-Net [50], CIA-Net [51], UNet [52] and cGANs [53], that are trained on multi-organ nuclei segmentation datasets (MoNuSeg [54], PanNuke [55], CoNSep [50]). The entities can also be calculated using agglomerative clustering [56] of detected cells.

The nodes in a graph can also be represented by fixed-sized patches (patch-graphs) randomly sampled from the raw WSI or by using a patch selection method where non-tissue regions are removed [57]. Important patches can be sampled from segmented tissues using color thresholds where patches with similar features (tissue cluster) are modeled as a node. Pre-trained deep learning models on tissue datasets (*e.g.* NCT-CRC-HE-100 [58]) have also been used to detect the tumor region of the specific pathological task.

Meaningful tissue regions have been also used as nodes to capture the tissue distribution (tissue-graphs). To separate tissue structures, superpixels [59] obtained using unsupervised algorithms such as simple linear iterative clustering (SLIC) [60] become nodes.

b) *Node embeddings:* Node features can comprise handcrafted features including morphological and topological prop-

erties (*e.g.* shape, size, orientation, nuclei intensity, and the chromaticity using the gray-level co-occurrence matrix). For cell-graph representations, some works include learned features extracted from the trained model used to localise the nuclei.

In patch-graph methods, deep neural networks are used to automatically learn a feature representation from patches around the centroids of the nuclei and tissue regions. If the entity is larger than the specified patch size, multiple patches inside the entity are processed, and the final feature is computed as the mean of the patch-level deep features. Some works have aggregated features from neighboring patches and combined them to obtain a central node representation to increase feature learning performance. Authors have adopted CNNs (MobileNetV2, DenseNet, ResNet-18 or ResNet-50 [61]), and encoder-decoder segmentation models (UNet [52]) for the purpose of deep feature extraction. To generate patch-level embeddings, ImageNet-pretrained CNN as well as a CNN pretrained for tissue sub-compartment classification task have been used.

c) *Edge definition:* The edge configuration encodes the cellular or tissue interactions, *i.e.* how likely two nearby entities will interact and consequently form an edge. This topology is often defined heuristically using a pre-defined proximity threshold, a nearest neighbor rule, a probabilistic model, or a Waxman model [22]. The graph topology can also be computed by constructing a region adjacency graph (RAG) [62] by using the spatial centroids of superpixels.

3) *Training paradigms:* From the perspective of supervision, we can categorize graph learning tasks into different training settings. Such approaches have also been used to extract effective representations from data.

- The *Supervised* learning setting provides labeled data for training.
- *Weakly or partially supervised learning* refers to models that are trained using examples that are only partially annotated.
- *Semi-supervised learning* trains a model using a small set of annotated samples, then generates pseudo-labels for a large set of samples without annotations, and learns a final model by mixing both sets of samples.

- *Self-supervised learning* is a form of unsupervised learning in which the data provides supervisory signals when learning a representation via a proxy task. Annotated data is used to fine-tune the representation once it has been learned. Some self-supervised approaches adopted as feature extractors include contrastive predictive coding (CPC) [63], texture auto encoder (Deep Ten) [64], and variational autoencoders (VAE) [65].

B. Graph neural networks models

Following graph building, the entity graph is processed using a graph-based deep learning model that works with graph-structured data to perform analysis.

GCNs can be broadly categorised as spectral-based [66], [67] and spatial-based [68]. Spectral-based GCNs use spectral convolutional neural networks, that build upon the graph Fourier transform and the normalized Laplacian matrix of the graph. Spatial-based GCNs define a graph convolution operation based on spatial relationships that exist among graph nodes.

Graph convolutional networks, similar to CNNs, learn abstract feature representations for each feature at a node via message passing, in which nodes successively aggregate feature vectors from their neighborhood to compute a new feature vector at the next hidden layer in the network.

A basic GNN consists of two components: The *AGGREGATE* operation can aggregate neighboring node representations of the center node, whereas the *COMBINE* operation combines the neighborhood node representation with the center node representation to generate the updated center node representation. The Aggregate and Combine at each l -th layer of the GNN can be defined as follows:

$$h_{\mathcal{N}_v}^{(t)} = \text{AGGREGATE}^{(l)}(\{h_u^{t-1}, \forall u \in \mathcal{N}_v\}), \quad (1)$$

where $h_{\mathcal{N}_v}^{(t)}$ is the aggregated node feature of the neighbourhood, h_u^{t-1} is the node feature in neighbourhood $\mathcal{N}(\cdot)$ of node v .

$$h_v^{(t)} = \text{COMBINE}^{(l)}(h_v^{t-1}, h_{\mathcal{N}_v}^{(t)}) = \sigma(W^t \cdot [h_v^{t-1} || h_{\mathcal{N}_v}^{(t)}]), \quad (2)$$

where $h_v^{(t)}$ is the node representation at the l -th iteration. $h_v^{(0)} = x_v$ where x_v is the initial feature vector for the node, σ denotes the logistic sigmoid function, and $||$ denotes vector concatenation.

With the network structure and node content information as inputs, the outputs of GNNs can focus on various graph analytic tasks using one of the processes listed below:

- *Node-level prediction*: A GNN operating at the node-level computes values for each node in the graph and is thus useful for node classification and regression purposes. In node classification, the task is to predict the node label for every node in a graph. To compute the node-level predictions, the node embedding is input to a Multi-Layer Perceptron (MLP) (See Fig. 4).
- *Graph-level prediction*: Refers to GNNs that predict a single value for an entire graph. This is mostly used to classify entire graphs, or compute similarities between graphs. To compute graph-level predictions, the same

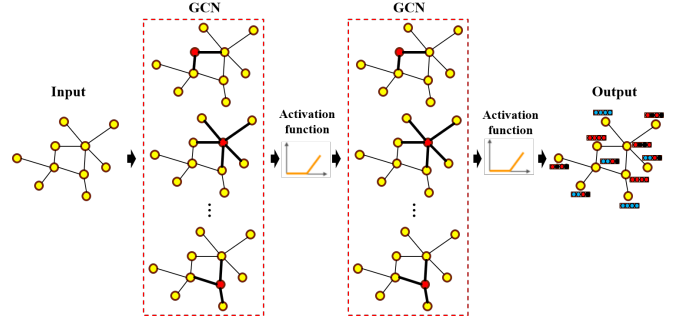


Fig. 4: Representation of graph architectures for node-level classification. Recreated from [11].

node embedding used in node-level prediction is input to a pooling process followed by a separate MLP (See Fig. 5).

In the following subsections, we describe in more detail the GNN architectures considered in digital pathology analysis methods. Different GNN variants employ different aggregators to acquire information from each node's neighbors, as well as different techniques to update the nodes' hidden states. In GNNs, the number of parameters is dependent on the number of node and edge features, as their aggregation is learned.

1) *ChebNet*: The convolution operation for spectral-based GCNs is defined in the Fourier domain by determining the eigen decomposition of the graph Laplacian [69]. The normalized graph Laplacian is defined as $L = I_N - D^{-1/2}AD^{-1/2} = U\Lambda U^T$ (D is the degree matrix and A is the adjacency matrix of the graph), where the columns of U are the matrix of eigenvectors and Λ is a diagonal matrix of its eigenvalues. The operation can be defined as the multiplication of a signal $x \in \mathbb{R}^N$ (a scalar for each node) with a filter $g_\theta = \text{diag}(\theta)$, parameterized by $\theta \in \mathbb{R}^N$,

$$g_\theta \star x = U g_\theta(\Lambda) U^T x. \quad (3)$$

Defferrard et al. [66] proposed a Chebyshev spectral CNN (ChebNet), which approximates the spectral filters by truncated Chebyshev polynomials, avoiding the calculation of the eigenvectors of the Laplacian matrix, and thus reducing the computational cost. A Chebyshev polynomial $T_m(x)$ of order m evaluated at \tilde{L} is used. Thus the operation is defined as,

$$g_\theta \star x \approx \sum_{m=0}^{M-1} \theta_m T_m(\tilde{L})x, \quad (4)$$

where \tilde{L} is a diagonal matrix of scaled eigenvalues defined as $\tilde{L} = 2L/\lambda_{\max} - I_N$. λ_{\max} denotes the largest eigenvalue of L . The Chebyshev polynomials are defined as $T_m(x) = 2xT_{m-1}(x) - T_{m-2}(x)$ with $T_0(x) = 1$ and $T_1(x) = x$.

2) *GCN*: A GCN is a spectral-based GNN with mean pooling aggregation. Kipf and Welling [67] presented the GCN using a localized first-order approximation of ChebNet. It limits the layer-wise convolution filter to $K = 1$ and uses a further approximation of $\lambda \approx 2$, to avoid overfitting and limit the number of parameters. Thus, Equation 4 can be simplified to,

$$g_\theta \star x \approx \theta'_0 x + \theta'_1 x(L - I_N)x = \theta'_0 x + \theta'_1 D^{-1/2}AD^{-1/2}x. \quad (5)$$

Here, θ'_0, θ'_1 are two unconstrained variables. A GCN further assumes that $\theta = \theta'_0 = -\theta'_1$, leading to the following definition of a graph convolution:

$$g_\theta \star x \approx \theta(I_N + D^{-1/2}AD^{-1/2})x \quad (6)$$

The definition to a signal $X \in \mathbb{R}^{N \times C}$ with C input channels and F filters for feature maps is generalized as follows,

$$Z = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} X \Theta, \quad (7)$$

where $\Theta \in \mathbb{R}^{C \times F}$ is the matrix formed by the filter bank parameters, and $Z \in \mathbb{R}^{N \times F}$ is the signal matrix obtained by convolution. From a spatial-based perspective, Equation 7 is reformulated in [70] as a message passing layer which updates the node's representation x_i^k as follows:

$$m_i^{k+1} = \sum_{j \in N(i) \cup i} \frac{x_j^k}{\sqrt{|N(j)||N(i)|}}, \quad (8)$$

$$x_i^{k+1} = \sigma(W^k m_i^{k+1}),$$

where m_i^k is the output of a message passing iteration, $|N(j)|$ and $|N(i)|$ denote the node degree of node j and i respectively, W^k denotes a layer-specific trainable weight matrix and σ is a non-linearity function.

3) *GraphSAGE*: GraphSAGE is a spatial-GCN which uses a node embedding with max-pooling aggregation. Hamilton et al. [68] offer an extension of GCNs for inductive unsupervised representation learning with trainable aggregation functions instead of simple convolutions applied to neighborhoods as in a GCN. The authors propose a batch-training algorithm for GCNs to save memory at the cost of sacrificing time efficiency. In [68] three aggregating functions are proposed: the element-wise mean, an LSTM, and max-pooling. The mean aggregator is an approximation of the convolutional operation from the transductive GCN framework [67]. An LSTM is adapted to operate on an unordered set by permuting the neighbors of the node. In the pooling aggregator, each neighbor's hidden state is fed through a fully-connected layer, and then a max-pooling operation is applied to the set of the node's neighbors. These aggregator functions are denoted as,

$$h_{\mathcal{N}_v}^t = \max\{\sigma(W_{\text{pool}} h_u^{t-1} + b_{\text{pool}}), \forall u \in \mathcal{N}_v\}, \quad (9)$$

where \mathcal{N}_v is the neighborhood set of node v , W_{pool} and b_{pool} are the parameters to be learned, and $\max\{\cdot\}$ is the element-wise maximum. Hence, following the message passing formulation in Equation 8, the node representation is updated according to,

$$m_i^{k+1} = \text{MEAN}_{j \in N(i) \cup i}(x_j^k), \quad (10)$$

$$x_i^{k+1} = \sigma(W^k m_i^{k+1}),$$

4) *GAT*: Inspired by the self-attention mechanism [71], graph attention networks (GAT) [72] incorporate the attention mechanism into the propagation steps by modifying the convolution operation. GAT is a spatial-GCN model that incorporates masked self-attention layers into graph convolutions and uses a neural network architecture to learn neighbor-specific weights. Veličković et al. [72] constructed a graph attention network by stacking a single graph attention layer, a , which is a single-layer feed-forward neural network, parametrized by a

weight vector $\vec{a} \in \mathbb{R}^{2F^i}$. The layer computes the coefficients in the attention mechanisms of the node pair (i, j) by,

$$\alpha_{i,j} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i \parallel W\vec{h}_j]))}{\sum_{k \in N_i \setminus N} \exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i \parallel W\vec{h}_k]))}, \quad (11)$$

where \parallel represents the concatenation operation. The attention layer takes as input a set of node features $h = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$, $\vec{h}_i \in \mathbb{R}^F$, where N is the number of nodes of the input graph and F the number of features for each node, and produces a new set of node features $h' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}$, $\vec{h}'_i \in \mathbb{R}^F$ as its output. To generate higher-level features, as an initial step a shared linear transformation, parametrized by a weight matrix $W \in \mathbb{R}^{F' \times F}$, is applied to every node and subsequently a masked attention mechanism is applied to every node, resulting in the following scores,

$$e_{ij} = a(W\vec{h}_i, W\vec{h}_j), \quad (12)$$

that indicates the importance of node j 's features to node i . The final output feature of each node can be obtained by applying a non-linearity, σ ,

$$h'_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W h_j\right). \quad (13)$$

The layer also uses multi-head attention to stabilise the learning process. K different attention heads are applied to compute mutually independent features in parallel, and then their features are concatenated.

The attention coefficients are used to update the node representation according to the following message passing formulation,

$$m_i^{k+1} = \sum_{j \in N(i)} \alpha_{i,j}^k W^k x_j^k, \quad (14)$$

$$x_i^{k+1} = \sigma(\alpha_{i,j}^k W^k x_j^k + m_i^{k+1}),$$

5) *GIN*: The graph isomorphism network (GIN) [73] is a spatial-GCN that aggregates neighborhood information by summing the representations of neighboring nodes. Isomorphism graph-based models are designed to interpret graphs with different nodes and edges. The representation of node i itself is then updated using a MLP,

$$m_i^{k+1} = \sum_{j \in N(i)} x_j^k, \quad (15)$$

$$x_i^{k+1} = F((1 + \epsilon) \cdot x_i^k + m_i^{k+1}),$$

where F is the MLP and ϵ is either a learnable parameter or fixed. GIN's aggregation and readout functions are injective, and thus are designed to achieve maximum discriminative power [73].

6) *Other GNN architectures in histopathology*: Other GNN architectures considered for entity-graph evaluation in digital pathology that were proposed by the surveyed works include:

- *Edge graph neural network (EGNN)* [38], [74]: Edge features are included when leveraging the graph structure in the network.
- *Robust spatial filtering (RSF)* [28], [30], [75]: These spatial-based models are more flexible when dealing with

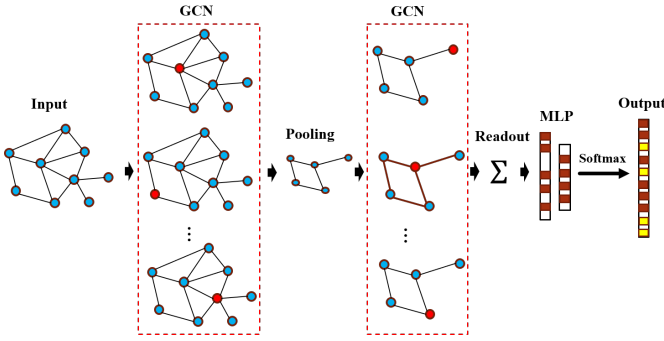


Fig. 5: Representation of graph models for graph-level classification. Recreated from [11].

heterogeneous graphs as the graph inputs can be easily incorporated into the aggregation function.

- *Adaptive GraphSAGE* [27], [39]: Graph networks with the ability to more effectively learn the embedding feature between nodes, by using a learnable pattern to adaptively aggregate multi-level embedding features for each node. 3
- *Jumping Knowledge Network (JK-Net)* Xu et al. [76] proposed the Jumping Knowledge (JK) approach to adaptively leverage, for each node, different neighborhood ranges to better represent feature.
- *Feature-enhanced spatial-GCN (FENet)* [41], [73]: This model is proposed to analyse non-isomorphic graphs, distinct from isomorphic graphs which strictly share the same adjacency neighborhood matrix. The feature-enhance mechanism adaptively selects the node representation from different graph convolution layers. The model adopts sum-pooling to capture the full structural information of the entire graph representation.
- *Multi-scale graph wavelet neural network (MS-GWNN)* [29], [77]: This spectral model leverages the localization property of graph wavelets to perform multi-scale analysis with a variety of scaling parameters in parallel, offering high efficiency and good interpretability for graph convolution.

C. Graph pooling

Different graph pooling strategies have been developed to minimise the graph size in order to learn hierarchical features for improved graph-level classification, and reduce computational complexity.

a) Global pooling: The most fundamental type of signal pooling on a graph is global pooling. It is also referred to as a readout layer in the literature. Similar to CNNs, mean, max, and sum functions are often utilized as basic pooling methods. Other approaches, instead of employing these simple aggregators, transform the vertex representation to a permutation invariant graph-level representation or embedding. In particular, Li et al. [78] proposed a *global attention pooling* system that uses a soft attention mechanism to determine which nodes are relevant to the present graph-level task and returns the pooled feature vector from all nodes.

b) Hierarchical pooling: A graph pooling layer in the GCN pools information from multiple vertices to one vertex, to reduce graph size and expand the receptive field of the graph filters. Many graph classification methods use hierarchical pooling in conjunction with a final global pooling or readout layer to represent the graph as illustrated in Fig. 5 Below we outline the most common hierarchical pooling techniques used in digital pathology.

- *DiffPool:* Ying et al. [79] introduced the differentiable graph pooling operator (DiffPool) which uses another graph convolution layer to generate the assignment matrix for each node (*i.e.* DiffPool does not simply cluster the nodes in a graph, but learns a cluster assignment matrix).
- *SAGPool* The self-attention graph pooling (SAGPool) introduced by Lee et al. [80] is a hierarchical pooling method that performs local pooling operations over node embeddings in a graph. The pooling module considers both node features and graph topology and learns to pool features via a self-attention mechanism, which can reduce computational complexity.

D. Graph interpretations

Graph representations embed biological entities and their interactions, but their explainability for digital pathology is less explored. While cells and their spatial interactions are visible in great detail, identifying relevant visual features is difficult. To undertake due diligence on model outputs and improve understanding of disease mechanisms and therapies, the medical community requires interpretable models.

The two most popular types of interpretation methodologies are model-based and post-hoc interpretability. The former constrains the model so that it can quickly deliver meaningful details about the relationships that have been discovered (such as sparsity, modularity, etc). Here, internal model information such as weights or structural information can be accessed and used to infer group-level patterns across training instances. The latter seeks to extract information about the learnt relationships in the model. These post-hoc methods are typically used to analyze individual feature input and output pairs, limiting their explainability to the individual sample level.

1) Attention mechanisms: Graph-structured data can be both massive and noisy, and not all portions of the graph are equally important. As such, attention mechanisms can direct a network to focus on the most relevant parts of the input, suppressing uninformative features, reducing computational cost and enhancing accuracy. A gate-based attention mechanism [81] controls, for example, the expressiveness of each feature. Attention has also been used as an explanation technique where the attention weights highlight the nodes and edges in their relative order of importance, and can be used for discovering the underlying dependencies that have been learnt. The activation map and gradient sensitivity of GAT models are used to interpret the salient input features at both the group and individual levels.

In a graph model with attention, selected layers of the graph are connected to an attention layer, and all attention layers are jointly trained with the network. A traditional attention

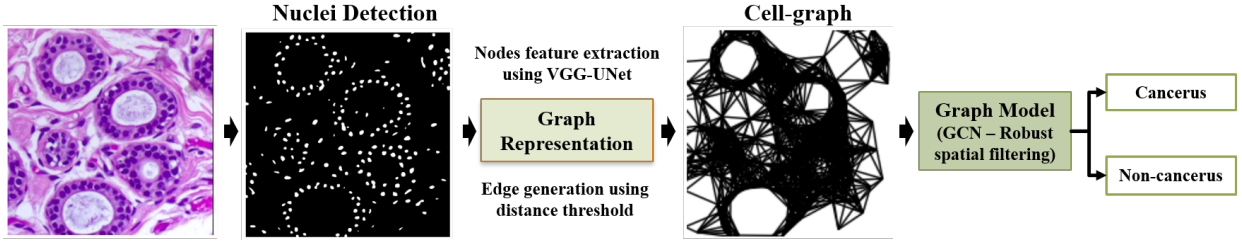


Fig. 6: Cell-graph based representation. Nucleus detection is conducted using fully convolutional networks. Then, edge and vertex features are computed to obtain an entity-graph representation as input to a GCN for cancer classification. Recreated from [28].

mechanism that can be learned by gradient-based methods [82] can be formulated as,

$$\begin{aligned}
 u_t &= \tanh(W h_t + b), \\
 \alpha_t &= \frac{\exp(u_t^T u_w)}{\sum_{j=1}^n \exp(u_t^T u_w)}, \\
 s_t &= \sum_t \alpha_t h_t,
 \end{aligned} \tag{16}$$

where h_t is the output of a layer; and W , u_w and b are trainable weights and bias. The importance of each element in h_t is measured by estimating the similarity between u_t and h_t , which is randomly initialized. α_t is a softmax function. The scores are multiplied by the hidden states to calculate the weighted combination, s_t (the attention-weighted final output).

2) *Graph explainers*: Several post-hoc feature attribution graph explainers have been presented in the literature including excitation backpropagation [83], a node pruning-based explainer (GNNExplainer) [84], gradient-based explainers (GraphGrad-CAM [85] and GraphGrad-CAM++ [86]), a layerwise relevance propagation explainer (GraphLRP) [87], [88], and deep graph mapper [89].

III. APPLICATIONS OF GRAPH DEEP LEARNING IN DIGITAL PATHOLOGY

The case studies presented in this section are organised according to the methodology adopted for the graph representation and the clinical application. The graph model, training paradigm, and datasets used in all applications are detailed in Table II. Rather than providing an exhaustive review of the literature, we present prominent highlights concerning the pre-processing, graph construction and graph models adopted, and their benefits in addressing various pathology tasks.

With the development of TMAs and WSI scanning techniques, as well as access to massive digital datasets of tissue images, deep learning methods for tumor localization, survival prediction and cancer recurrence prediction have made substantial progress [104]. Both the spatial arrangement of cells of various types (macro features), and the details of specific cells (micro features) are important for detecting and characterizing cancers. Thus, a valuable representation of histopathology data must capture micro features and macro spatial relationships. Graphs are powerful representational data structures, and have attracted significant attention in analysis of histopathological images [105] due to their ability to represent tissue architectures. The paradigm change from pixel-based to entity-based

research has the potential to improve deep learning techniques' interpretability in digital pathology, which is relevant for diagnostics.

A. Cell-graph representation

Most of these works follow a similar framework where a cell-graphs is introduced using cells as the entities to capture the cell micro-environment. The image is converted into a graph representation with the locations of identified cells serving as graph vertices and edges constructed depending on spatial distance. Cell-level features are extracted as the initial node embedding. The cell-graph is fed to a GCN to perform image-wise classification.

1) *Breast cancer*: Breast cancer is the most commonly diagnosed cancer and registers the highest number of cancer deaths among women. A majority of breast lesions are diagnosed along a spectrum of cancer classes that ranges from benign to invasive. Cancer diagnosis and the detection of breast cancer is one of the most common applications of machine learning and computer vision within digital pathology analysis. CNNs have been used for various digital pathology tasks in breast cancer diagnosis such as nucleus segmentation and classification, and tumor detection and staging. However, these patch-wise approaches do not explicitly capture the inter-nuclear relationships and limit access to global information.

Anand et al. [28] proposed the use of GCNs to classify WSIs represented by graphs of their constituent cells. Micro-level features (nuclear morphology) were incorporated as vertex features using local image descriptors, while macro-level features (gland formation) were included as edge attributes based on a mapping of Euclidean distances between nearby nuclei. The vertex features are represented by the average RGB intensity, morphological features and learned features extracted from a pre-trained CNN applied to a window around the nuclei centroid. Finally, each tissue image is classified by giving its cell-graph as an input to the GCN which is trained in a supervised manner. The authors adopted a spatial GCN known as robust spatial filtering (RSF) [75], which can take heterogeneous graphs as input. This framework is depicted in Fig. 6. The authors demonstrate competitive performance compared to conventional patch-based CNN approaches to classify patients into cancerous or non-cancerous groups using the Breast Cancer Histology Challenge (BACH) dataset [9].

Sureka et al. [30] modeled histology tissue as a graph of nuclei and employed the RSF with a GCN [75] with attention

TABLE II: Summary of applications and graphs models in computational pathology.

Authors	Topic	Application	Entity-graph	GNN Model + Explainer	Input; Training (Node detection/embeddings); Training (GNN model/pathology task); Datasets; Additional remarks
Jaume et al. (2021) [33]	Classification	Breast cancer	CG	GIN + Post-hoc explainers	WSI; Supervised; Supervised; BRACS [17] (5 classes); Post-hoc explainers: GNNExplainer, GraphGrad-CAM, GraphGrad-CAM++, GraphLRP.
Jaume et al. (2020) [34]	Classification	Breast cancer	CG	GIN + CGExplainer	WSI; Supervised; Supervised; BRACS [17] (5 classes); Customized cell-graph explainer based on GNNExplainer.
Sureka et al. (2020) [30]	Classification	Breast cancer / Prostate cancer	CG	GCN, RSF + Attention/Node occlusion	WSI, TMAs; Supervised; Supervised; Breast cancer: BACH [9] (2 classes), Prostate cancer: TM [90] (2 classes); Gleason grade.
Anand et al. (2020) [28]	Classification	Breast cancer	CG	GCN, RSF	WSI; Supervised; Supervised; BACH [9] (4 classes).
Studer et al. (2021) [38]	Classification	Colorectal cancer	CG	GCN, GraphSAGE, GAT, GIN, ENN, JK-Net	WSI; Supervised; Supervised; pT1-Gland Graph [91] (2 classes); Graph-level output. Concatenation of global add, mean and max pooling). Dysplasia of intestinal glands.
Zhou et al. (2019) [39]	Classification	Colorectal cancer	CG	Adaptive GraphSAGE, JK-Net, Graph clustering	WSI; Supervised; Supervised; CRC dataset [92] (3 classes); Graph-level output. Hierarchical representation of cells based on graph clustering method from DiffPool).
Wang et al. (2020) [15]	Classification	Prostate cancer	CG	GraphSAGE, SAGPool	TMA; Self-supervised; Weakly-supervised; UZH prostate TMAs [93] (2 classes); Graph-level output. Grade classification (low and high-risk).
Ozen et al. (2020) [35]	ROI Retrieval	Breast cancer	PG	GCN, DiffPool	WSI; Supervised; Self-Supervised; Department of Pathology at Hacettepe University (private) (4 classes); Histopathological image retrieval (slide-level and ROI-level).
Lu et al. (2020) [36]	Classification	Breast cancer (HER2, PR)	TG	GIN	WSI; Supervised; Supervised; TCGA-BRCA [94] (2 classes); Graph-level. Status of Human epidermal growth factor receptor 2 (HER2) and Progesterone receptor (PR).
Aygüneş et al. (2020) [26]	Classification	Breast cancer	PG	GCN	WSI; Supervised; Weakly-supervised; Department of Pathology at Hacettepe University (private) (4 classes). ROI-level classification.
Ye et al. (2019) [37]	Classification	Breast cancer	PG	GCN	WSI; Supervised; Supervised; BACH [9] (4 classes); Graph construction based on the ROI segmentation map.
Zhao et al. (2020) [40]	Classification	Colorectal cancer	PG	ChebNet, SAGPool	WSI; Self-Supervised; Weakly-supervised; TCGA-COAD [95] (2 classes); Multiple instance learning. Graph-level output.
Raju et al. (2020) [27]	Classification	Colorectal cancer	TG	Adaptive GraphSage + Attention	WSI; Self-Supervised; Weakly-supervised; MCO [96] (4 classes); Multiple instance learning. Cluster embedding (Siamese architecture); Tumor node metastasis staging.
Ding et al. (2020) [41]	Classification	Colorectal cancer	PG	Spatial-GCN (FENet)	WSI; Supervised; Supervised; TCGA-COAD and TCGA-READ [97] (2 classes); Genetic mutational prediction.
Adnan et al. (2020) [43]	Classification	Lung cancer	PG	ChebNet, GraphSAGE + Global attention pooling	WSI; Supervised; Supervised; TCGA-LUSC [98] (2 classes), MUSK1 [99]; Adjacency learning layer. Multiple instance learning.
Zheng et al. (2019) [44]	Retrieval	Lung cancer	PG	GNN, DiffPool (GNN-Hash)	WSI; Supervised; Similarity (Hamming distance); ACDC-LungHP [98]; Hashing methods and binary encoding. Histopathological image retrieval.
Li et al. (2018) [45]	Classification	Lung cancer	PG	ChebNet + Attention	WSI; Self-Supervised; Supervised; TCGA-LUSC [98] (2 classes), NLST [100] (2 classes); Survival prediction.
Wu et al. (2019) [47]	Classification	Skin cancer	PG	GCN	WSI; Supervised; Weakly- and Semi-supervised; BCC data collected from 2 different hospitals (private) (4 classes).
Anklin et al. (2021) [42]	Segmentation / Classification	Prostate cancer	TG	GIN (SegGini) + GraphGrad-CAM	TMA, WSI; Supervised; Weakly-supervised; UZH prostate TMAs [93] (4 classes), SICAPv2 [101] (4 classes); Gleason grade, Post-hoc interpretability.
Pati et al. (2021) [31]	Classification	Breast cancer	CG, TG, HR	GIN-PNA (HACT-Net) + GraphGrad-CAM	WSI; Supervised; Supervised; BRACS [17] (7 classes), BACH [9] (4 classes); Cell-to-Tissue Hierarchies.
Pati et al. (2020) [17]	Classification	Breast cancer	CG, TG, HR	GIN (HACT-Net)	WSI; Supervised; Supervised; BRACS [17] (5 classes); Cell-to-Tissue Hierarchies.
Zhang and Li (2020) [29]	Classification	Breast cancer	PG, HR	MS-GWNN	WSI; Supervised; Supervised; BACH [9] (4 classes), BreakHis [102] (2 classes); Multi-scale graph feature learning (node-level and graph-level prediction).
Levy et al. (2021) [16]	Regression	Colorectal cancer / lymphoma	PG, HR	GAT, TDA + Graph Mapper	WSI; Supervised; Supervised; Dartmouth Hitchcock Medical Center (private); colon (9 classes), lymph (4 classes); Hierarchical representation. Tumor invasion score and staging.
Shi et al. (2020) [46]	Classification	Cervical cancer	CCG	Fusion CNN-GCN	RGB; Supervised; Semi-supervised; SIPaKMed [103] (5 classes), MotiC [25] (7 classes); Population analysis of isolated cell images.
Shi et al. (2019) [25]	Classification	Cervical cancer	CCG	Fusion CNN-GCN	RGB; Supervised; Supervised; SIPaKMed [103] (5 classes), MotiC [25] (7 classes); Population analysis of isolated cell images.
Chen et al. (2020) [32]	Classification	Renal Cancer	CG	GraphSAGE, SAGPool + Attention	Fusion: WSI+Genome; Self-Supervised; Self-Supervised; TCGA-GBMLGG, TCGA-KIRC [98]; Survival outcome, Integrated gradient method.

Graph representation: Cell-Graph (CG); Patch-Graph (PG); Tissue-Graph (TG); Hierarchical Representation (HR); Cluster-Centroids-Graph (CCG)

mechanisms and node occlusion to highlight the relative cell contributions in the image, which fits the mental model used by pathologists. In the first approach, the authors occluded nuclei clusters to assess the drop in the probability of the correct class, while also including a method based on [106] to learn enhanced vertex and edge features. In a second approach, an attention layer is introduced before the first pooling operation for visualization of important nuclei for the binary classification of breast cancer on the BACH dataset and Gleason grade classification on a prostate cancer [90] dataset.

Several explainers have been applied in digital pathology, inspired by explainability techniques for CNN model predictions on images. However, pixel-level explanations fail to encode tumor macro-environment information, and result in ill-defined visual heatmaps of important locations as illustrated in Fig. 7. Thus, graph representations are relevant for both diagnostics and interpretation. Generating intuitive explanations for pathologists is critical to quantify the quality of the explanation. To address this, Jaume et al. [33] introduced a framework using entity-based graph analysis to provide pathologically-understandable concepts (*i.e.* to make the graph decisions understandable to pathologists). The authors pro-

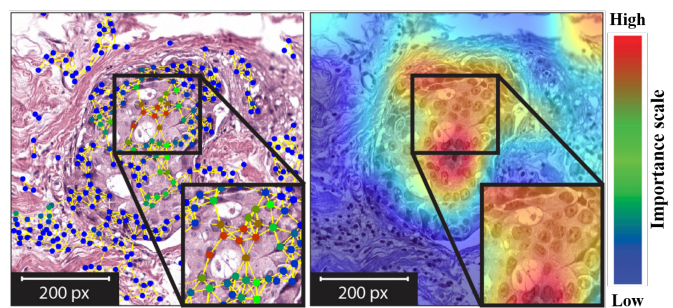


Fig. 7: For a ductal carcinoma, examples of explanations given by graph-based (Left) and pixel-based (Right) explainability algorithms. Recreated from [33].

posed a set of quantitative metrics based on pathologically measurable cellular properties to characterize explainability techniques in cell-graph representations for breast cancer subtyping.

In [33], the authors first transform the histology image into a cell-graph, and a GIN model is used to map the corresponding class level. Then, a post-hoc graph explainer

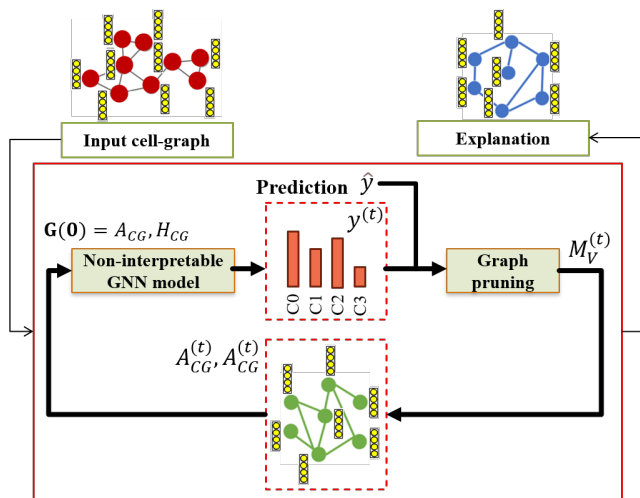


Fig. 8: Cell-graph explainer (CGExplainer): a customized post-hoc graph explainer based on graph pruning optimization. Recreated from [34].

generates an explanation per entity graph. Finally, the proposed metrics are used to assess explanation quality in identifying the nuclei driving the prediction (nuclei importance maps). Four graph explainers were considered in this analysis: GNNExplainer [84], GraphGrad-CAM [85], GraphGrad-CAM++ [86], and GraphLRP [87]. The results on the Breast Carcinoma Subtyping (BRACS) dataset [17] confirm that GraphGrad-CAM++ produces the best overall agreement with pathologists. The proposed metrics, which include domain-specific user-understandable terminology, could be useful for quantitative evaluation of graph explainability.

Jaume et al. [34] focused on the analysis of cells and cellular interactions in breast cancer sub-typing classification, and introduced an instance-level post-hoc graph-pruning explainer to identify decisive cells and interactions from the input graph in the BRACS dataset [17]. To create the cell-graph, nuclei are detected with segmentation algorithms and hand-crafted features including shape, texture and color attributes are extracted to represent each nucleus. The cell-graph topology uses the KNN algorithm and is based on the assumption that spatially close cells encode biological relationships and, as a result, should create an edge. The cell-graph is processed by a GIN model, followed by a MLP to predict the cancer stages.

Jaume et al. [34] designed a cell-graph explainer (CGExplainer), based on the GNNExplainer, to remove redundant and uninformative graph components, and the resulting sub-graph will be responsible for class-specific patterns that will aid disease comprehension. This module aims to learn a mask at the node-level that activates or deactivates parts of the graph. Fig. 8 provides an overview of the explainer module. The proposed explainer was shown to prune a substantial percentage of nodes and edges to extract valuable information while retaining prediction accuracy (*e.g.* the explanations retain relevant tumor epithelial nuclei for cancer diagnosis).

2) *Colorectal cancer*: Colorectal cancer (CRC) grading is a critical task since it plays a key role in determining the appropriate follow-up treatment, and is also indicative of overall patient outcome. The grade of a cancer is determined,

for example, by assessing the degree of glandular formation in the tumour. Nevertheless, automatic CNN-based methods for grading CRC typically use image patches which fail to include information on the micro-architecture of the entire tissue sample, and do not capture correspondence between the tissue morphology and glandular structure. To model nuclear features along with their cellular interactions, Zhou et al. [39] proposed a cell-graph model for grading CRC, in which each node is represented by a nucleus within the original image, and cellular interactions are captured as graph edges based on node similarity. A nuclear instance segmentation model is used to detect the nucleus and to extract accurate node features including nucleus shape and appearance features. Spatial features such as centroid coordinates, nuclei intensity and dissimilarity extracted from the grey level co-occurrence matrix were used as descriptors for predicting the grade of cancer. To reduce the number of nodes and edges based on the relative inter-node distance, an additional sampling strategy was used.

To conduct the graph-level classification, the authors in [39] proposed the Adaptive GraphSAGE model, which is inspired by GraphSAGE [68] and JK-Net [76], to obtain multi-level features (*i.e.* capturing the gland structure at various scales). To achieve multi-scale feature fusion, Adaptive GraphSAGE employs an attention technique which allows the network to adaptively generate an effective node representation.

A graph clustering operation, which can be considered as an extension of DiffPool [79], is used to group cells according to their appearance and tissue type, and to extract more abstract features for hierarchical representation. However, since the tissue hierarchy is inaccessible via this approach, the representation does not include high-level tissue features. Based on the degree of gland differentiation, the graph model categorises each image as normal, low-grade, or high-grade. In comparison with a traditional CNN, the proposed model achieves better accuracy by incorporating both nuclear and graph-level features.

Dysplasia of intestinal glands is especially important in pT1 colorectal cancer, the earliest stage of invasive colorectal cancer. Studer et al. [91] introduced the pT1 Gland graph (pT1-GG) dataset that consists of cell-graphs of healthy and dysplastic intestinal glands. In this work, the authors established a baseline for gland classification using labelled cell-graphs and the graph edit distance (GED), which is an error-tolerant measurement of similarity between two graphs. This technique is an improved version of the bipartite graph-matching method (BP2) [107] combined with a KNN algorithm to perform classification.

Later, the same authors investigated different graph-based architectures [38] to classify healthy gland tissue and dysplastic glandular areas on the pT1-GG dataset. The GNN architectures evaluated for cell-graph classification are GCN [67], GraphSAGE [68], GAT [72], GIN [73], EGNN [74] and a 1-dimensional GNN [108]. All models are trained using three graph convolution layers where GraphSAGE and GCN are also trained with jumping knowledge (JK) [76] to allow for an adaptive neighborhood range by aggregating representations across different layers. A concatenation of global sum-

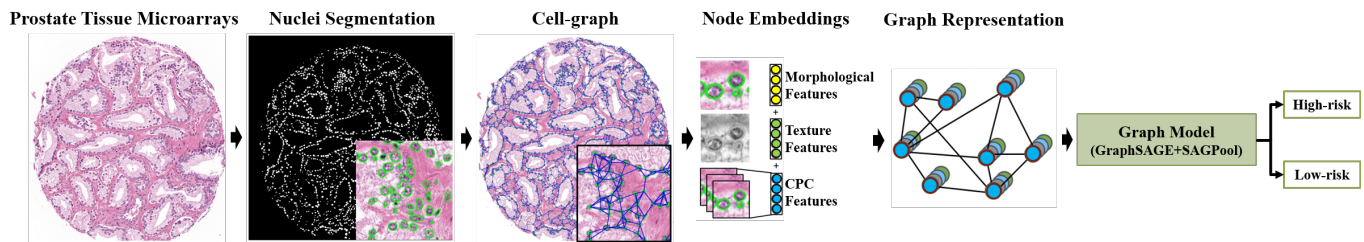


Fig. 9: The nuclei that have been detected are segmented, and a graph is constructed using the centroid of each nuclei. For each node, morphological, texture and contrastive predictive coding features are extracted, and GCNs are used as the graph representation. Recreated from [15].

pooling, global mean-pooling and global max-pooling is used to get the graph-level output, followed by a MLP to classify an input graph. The results demonstrated that graph-based deep learning methods outperformed classical graph-based and CNN-based methods. It should be emphasised, however, that each node is only linked to its two spatially closest neighbors, resulting in very restricted information sharing during message passing.

3) *Prostate cancer*: The commonly used Gleason score, which is based on the architectural pattern of tumor tissues and the distribution of glands, determines the aggressiveness of prostate cancer. CNNs have been used for histology image classification including Gleason score assignment, but CNNs are unable to capture the dense spatial relationships between cells and require detailed pixel level annotations for training.

To analyse the spatial distribution of the glands in prostate TMAs, Wang et al. [15] proposed a weakly-supervised approach for grade classification and to stratify low and high-risk cases (Gleason score < 6 is normal tissue; Gleason score ≥ 6 is abnormal tissue or high-risk). The authors segmented the nuclei and construct a cell-graph for each image with nuclei as the nodes, and the distance between neighboring nuclei as the edges, as illustrated in Fig. 9. Using prostate TMAs with only image-level labels rather than pixel-level labels, a GCN is used to identify high-risk patients via a self-supervised technique known as contrastive predictive coding (CPC) [63]. Features for each node are generated by extracting morphological (area, roundness) and texture features (dissimilarity, homogeneity) as well as features from CPC-based learning. A GraphSAGE convolution and a self-attention graph pooling (SAGPool) [80] are applied to the graph representation to learn from the global distribution of cell nuclei, cell morphology and spatial features. The proposed method can calculate attention scores, focus on the more significant node attributes, and aggregate information at different levels.

B. Patch-graphs and Tissue-graphs representations

The majority of the following works transform pathological images into patch-graphs, where nodes are important patches, and edges encode the intrinsic relationships between these patches. These patches are sampled using methods such as color-based, cell density or attention mechanisms. Then, CNNs are used to extract features from these patches to generate a feature vector for the node embedding of the graph representation. Given the constructed graph, a graph deep learning

model is used to conduct node or graph classification. It is important to make the distinction between tissue-graphs, which are biologically-defined and capture relevant morphological regions; while patch-graphs connect patches of interest, where each patch can contain multiple biological entities, with each other.

1) *Breast cancer*: Multi-class classification of arbitrarily sized ROIs is an important problem that serves as a necessary step in the diagnostic process for breast cancer. Aygüneş et al. [26] proposed to incorporate local context through a graph-based ROI representation over a variable number of patches (nodes) and their spatial proximity relationships (edges). A CNN is used to extract a feature vector for each node represented by fixed-sized patches of the ROI. Then, to propagate information across patches and incorporate local contextual information, two consecutive GCNs are used, which also aggregate the patch representation to classify the whole ROI into a diagnostic class. The classification is conducted in a weakly-supervised manner over the patches and ROI-level annotations, without having access to patch-level labels. Results on a private data collected from the Department of Pathology at Hacettepe University outperformed CNN-based models that incorporated majority-voting, learned-fusion and base-penultimate methods.

Some traditional CNN-based models have proposed to jointly segment a ROI of an image and classify WSIs and that enabled the classifier to better predict the image class [109]. Ye et al. [37] captured the topological structure of a ROI image through a GCN where a graph is constructed with segmentation masks of image patches that contain high levels of semantic information. The segmentation mask for each image patch is obtained using an encoder-decoder semantic segmentation framework where each pixel is classified as one of the four classes of tissue samples (normal, benign, in situ, and invasive) of the BACH [9] dataset. The combined segmentation masks of the image patches yield the total ROI segmentation mask. The area ratio of each lesion is calculated as the value of the unit node in each picture patch. Then, a graph is constructed to capture the spatial dependencies using the features of the image patch segmentation masks. Finally, the ROI image is classified based on the features learned by the GCNs.

One limitation of previous works is that they construct graphs using small patches of the WSI. Lu et al. [36] overcome this challenge by introducing a pipeline to construct a

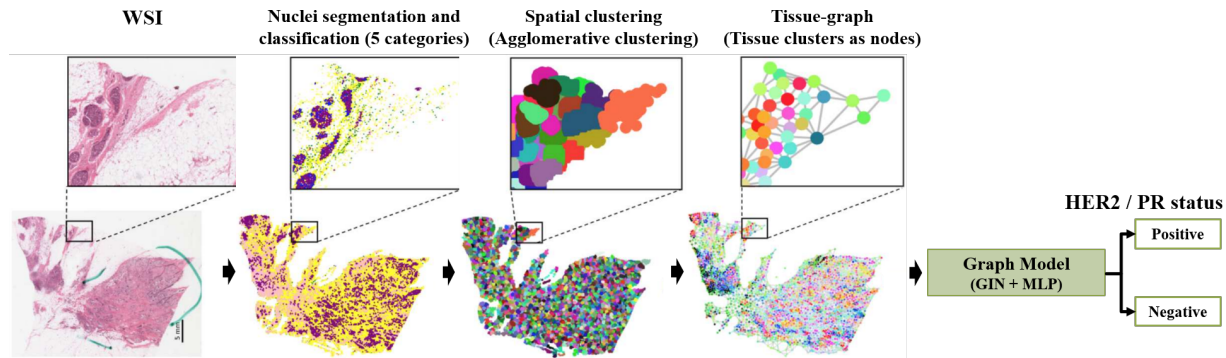


Fig. 10: Slide Graph Model which constructs a graph from the nuclei level to the entire WSI-level. The main steps are as follows: segmenting and classification of nuclei; clustering; constructing the graph and graph classification. Recreated from [36].

graph from the entire WSI using the nuclei level information, including geometry and cellular organization in tissue slides (termed the histology landscape). After building the graph, the authors used a GIN model to predict the positive or negative human epidermal growth factor receptor 2 (HER2), and the progesterone receptor (PR), which are two valuable biomarkers for breast cancer prognosis.

The proposed method in [36] consists of four steps as illustrated in Fig. 10. This work first used Hover-Net [50] to simultaneously segment and classify the individual nuclei and extract their features. Then, agglomerative clustering [56] is used to group spatially neighboring nuclei into clusters which results in reduced computational cost for downstream analysis. Using these clusters, a graph is generated by assigning the tissue clusters to nodes and the edges of the graph encode the cellular topology of the WSI. Lastly, the graph generated from the entire WSI is used as an input to a GCN to predict HER2 or PR status at the WSI-level. The performance of this method is evaluated on the hematoxylin and eosin (H&E) stained WSI images from the TCGA-BRCA [94] dataset, which consist of 608 HER2 negative and 101 HER2 positive, and 452 PR positive and 256 PR negative samples.

Content-based histopathological image retrieval has also been investigated for decision support in digital pathology. This system scans a pre-existing WSI database for regions that the pathologist is interested in and returns related regions to the pathologists for comparison. These methods can provide valuable information including diagnosis reports from experts for similar regions. Retrieval methods can also be used for classification by considering the most likely diagnosis [110]. However the amount of manually labelled training data limits their power. Ozen et al. [35] suggested a generic method that combines GNNs with a self-supervised training method that employs a contrastive loss function without requiring labeled data. In this framework, fixed-size patches and their spatial proximity relations are represented by undirected graphs. The simple framework for contrastive learning of visual representation (SimCLR) [111] is adopted for learning representations of ROIs. Using the contrastive loss, the GNN encoder and MLP projection head are trained to maximise the agreement between the representations. A GCN followed by a DiffPool operation is selected as the model configuration.

For content-based retrieval tasks, this GNN is trained in a self-supervised setting and is used to extract ROI representations where the Euclidean distance between the extracted representations is used to determine how similar two ROIs are. Quantitative results demonstrated that contrastive learning can improve the quality of learned representations, and despite not utilizing class labels could outperforming supervised classification methods.

2) *Colorectal cancer*: Although CNN-based approaches have practical merits when identifying important patches for predicting CRC, they do not take into account the spatial relationships between patches, which is important for determining the stage of the tumor. The size and the relative location of the tumor in relation to other tissue partitions are used for tumor node metastasis staging estimation. Furthermore, traditional approaches require the presence of expert pathologists to annotate each WSI. Weakly-supervised learning is an important and potentially viable solution to dealing with sparse annotations in medical imagery. Multiple instance learning (MIL) is well-suited to histology slide classification, as it is designed to operate on weakly-labeled data [4].

Raju et al. [27] considered the spatial relationship between tumor and other tissue partitions with a graph attention multi-instance learning framework to predict colorectal tumor node metastasis staging. Each graph with nodes representing different tissues serves as an instance, and the multiple instances for a WSI form a bag that aids in tumour stage prediction.

In [27], given a WSI, a texture autoencoder [64] is used to encode the texture from random sample patches. Then a cluster embedding network based on a Siamese architecture [112] is trained on a binary classification task to group similar texture features into multiple graphs. Each WSI is divided into multiple graphs and each graph has features from all cluster labels. The authors used a tissue wise annotated CRC dataset [113] to assign cluster labels for similar image patches. The authors consider the multiple graphs as multiple instances in a bag which are used to predict the tumor staging using an attention MIL method [114]. The authors adopted an Adaptive GraphSage [39] approach with learnable attention weights to assign more importance to instances which contain more information towards predicting the tumor stage. The authors demonstrated that graph attention multi-instance learning can

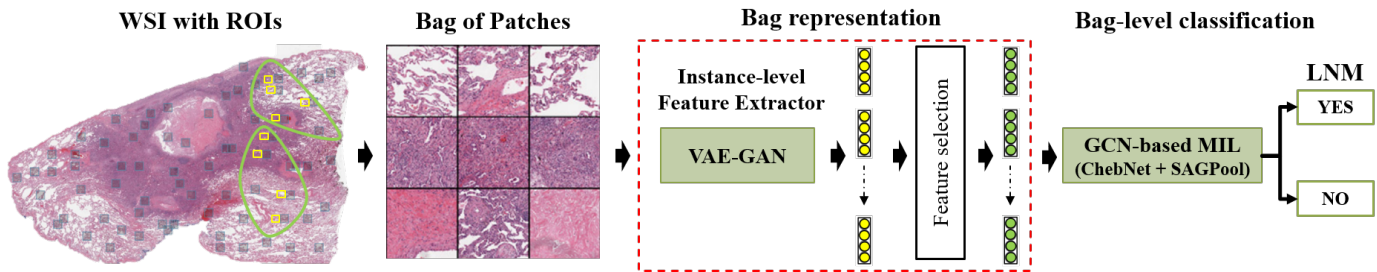


Fig. 11: Representation of a GCN-based MIL method. Once the bag of patches are extracted, instance-level feature extraction and selection is conducted followed by a bag-level classification. Recreated from [40].

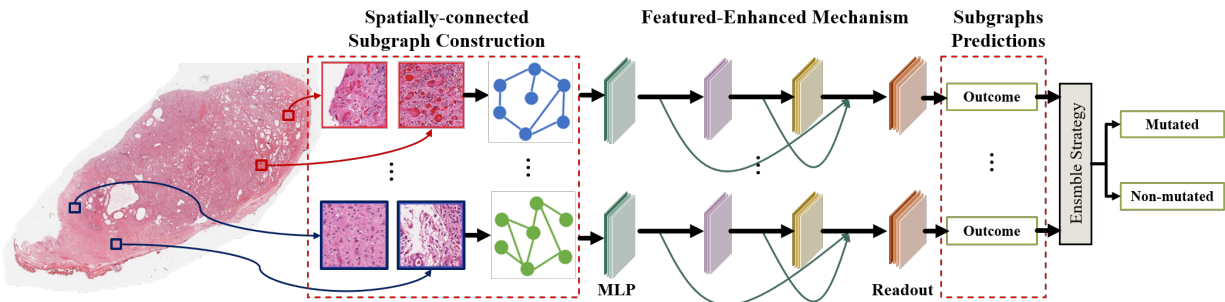


Fig. 12: The proposed FENet architecture. For each WSI, patches are randomly selected. Each patch corresponds to a node in each non-isomorphic subgraph where a CNN is used to extract node attributes. A feature-enhanced mechanism is adopted to consider all topological structural information. An ensemble approach used majority voting to aggregate all subgraphs' prediction outcomes. Recreated from [41].

perform better than a GCN on the Molecular and Cellular Oncology (MCO) [96] dataset.

Colorectal cancer lymph node metastasis (LNM) is a crucial factor in patient management and prognosis, and its identification suggests the need for dissection to avoid further spread. Zhao et al. [40] introduced a GCN-based multiple instance learning method combined with a feature selection strategy to predict LNM in the colon adenocarcinoma (COAD) cohort of the Cancer Genome Atlas (TCGA) project [95]. Following the MIL approach, the training dataset is composed of bags where each bag contains a set of instances. The goal of this work is to teach a model to predict the bag label, where only the bag-level label is available.

The overall framework has three major components: instance-level feature extraction, instance-level feature selection, and bag-level classification, as illustrated in Fig. 11. First, non-overlapping patches are extracted from a WSI which is represented as a bag of patches. Since instance labels are unavailable, the authors introduced a combination of a variational autoencoder (VAE) [65] and a generative adversarial network (GAN) for fine-tuning the encoder component as an instance-level feature extractor in a self-supervised manner. In this VAE-GAN model, the architecture of the network for the decoder of the VAE and generator of the GAN is the same network. Then, a feature selection component is incorporated to remove redundant and unhelpful features to alleviate the workload when generating the bag representation. The maximum mean discrepancy is used to evaluate the feature importance. Finally, the authors employed ChebNet [66] followed by SAGPool [80] to generate the bag representation and perform

the bag-level classification. The authors demonstrated that the proposed model outperformed CNN-based and attention-based MIL models.

Colon adenoma and carcinoma may occur as a result of a series of histopathological changes due to key genetic alterations. Thus, the ability to predict genetic mutations is important for the diagnosis of colon cancer. Ding et al. [41] proposed a feature-enhanced graph network (FENet) using a spatial-GCNs, based on GIN, to predict gene mutations across all three key mutational prediction tasks (APC, KRAS, and TP53) that are associated with colon cancer evolution. In this approach, multiple spatial graphs are created using randomly selected image patches from each patient's WSI.

The feature-enhanced mechanism aggregates features from neighboring patches and combines them as the central node representation to increase feature learning performance. The authors introduced GlobalAddPooling as a READOUT function to convert the node representation into a graph representation. The prediction outcome for each sub-graph is classified by fully-connected layers. Finally, an ensemble strategy combines the prediction results of all sub-graphs to predict mutated and non-mutated classes. Fig. 12 illustrates the proposed FENet networks. The authors demonstrated that the integration of multiple sub-graph outcomes in the proposed model leads to a significant improvement in prediction performance on the Cancer Genome Atlas Colon Adenocarcinoma dataset [97], outperforming graph-based baseline models such as ChebNet, GraphSAGE and GAT.

3) *Lung cancer*: Lung adenocarcinoma and lung squamous cell carcinoma are the most common subtypes of lung cancer,

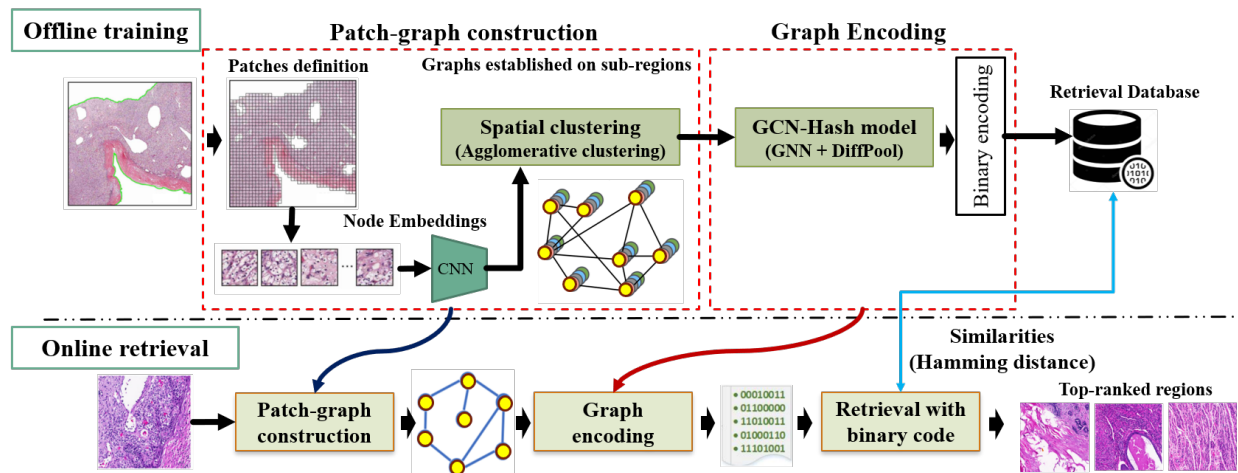


Fig. 13: Representation of the retrieval framework. Patch-graphs are constructed based on spatial relationships and feature distances between patches, and are fed into the developed GNN-Hash model for graph encoding. When retrieving, the query region is converted into a patch-graph and a binary code for similarity comparison with samples in the database. Recreated from [44].

and distinguishing between them requires a visual examination by an experienced pathologist. Efficient mining of survival-related structural features on a WSI is a promising way to improve survival analysis. Li et al. [45] introduced a GCN-based survival prediction model that integrated local patch features with global topological structures (patch-graph) through spectral graph convolution operators (ChebNet) using the TCGA-LUSC [98] and NLST [100] datasets. The model utilized a survival-specific graph trained under supervision using survival labels. A parallel graph attention mechanism is used to learn attention node features to improve model robustness by reducing the randomness of patch sampling (*i.e.* an adaptive patch selection by learning the importance of individual patches). This attention network is trained jointly with the prediction network. The authors demonstrated that topological features fine-tuned with survival-specific labels outperformed CNN-based models.

Adnan et al. [43] explored the application of GNNs for MIL. The authors sampled important patches from a WSI and model them as a fully-connected graph where the graph is converted to a vector representation for classification. Each instance is treated as a node of the graph in order to learn end-to-end relationships between nodes. In this approach, a DenseNet is used to extract features from all important patches sampled from a segmented tissue using color thresholds [57]. Then, an adjacency learning layer which uses global information about the patches is adopted to define the connections within nodes in an end-to-end manner. The adjacency matrix is calculated by an adjacency learning block using a series of dense layers and cross-correlation. The constructed graph is passed through two types of graph models (ChebNet and GraphSAGE), followed by a graph pooling layer to get a single feature vector to compare the discrimination of sub-types of lung cancer on the TCGA [98] and MUSK1 [99] datasets. With the adopted global attention pooling [78] which uses a soft attention mechanism, it is possible to visualise the importance that the network places on each patch when making the prediction.

The pooled representation is fed to two fully connected dense layers to achieve the final classification between lung adenocarcinoma and lung squamous cell carcinoma. The proposed model outperformed CNN-based models that use attention-MIL.

As discussed previously, content-based image retrieval seeks to find images that have morphological characteristics that are most similar to a query image. Binary encoding and hashing techniques have been successfully adopted to speed up the retrieval process in order to satisfy efficiency requirements [115]. However, WSI are commonly divided into small patches to index WSIs for region-level retrieval. This process does not consider the contextual information from a broad region surrounding the nuclei and the adjacency relationships that exist for different types of biopsy.

Zheng et al. [44] proposed a retrieval framework for a large-scale WSI database based on GNNs and hashing, which is illustrated in Fig. 13. Patch-graphs are first built in an offline stage based on patch spatial adjacency, and feature similarity extracted with a pre-trained CNN. Then, the patch-graphs are processed by a GNN-Hash model designed to use a graph encoding, and stored in the retrieval database. The GNN-Hash structure was created by stacking GNN modules and a DiffPool module [79]. The output of the hierarchical GNN-Hash is modified with a binary encoding layer in the final graph embedding layer. Finally, the relevant regions are retrieved and returned to pathologists after the region the pathologist queries is converted to a binary code. The similarities between the query code and those in the database are measured using Hamming distance. Experiments to estimate the adjacency relationships between local regions in WSIs and the similarities with query regions were conducted using the lung cancer ACDC-LungHP [98] dataset. The results demonstrated that the proposed retrieval model is scalable to different query region sizes and shapes, and returns tissue samples with similar content and structure.

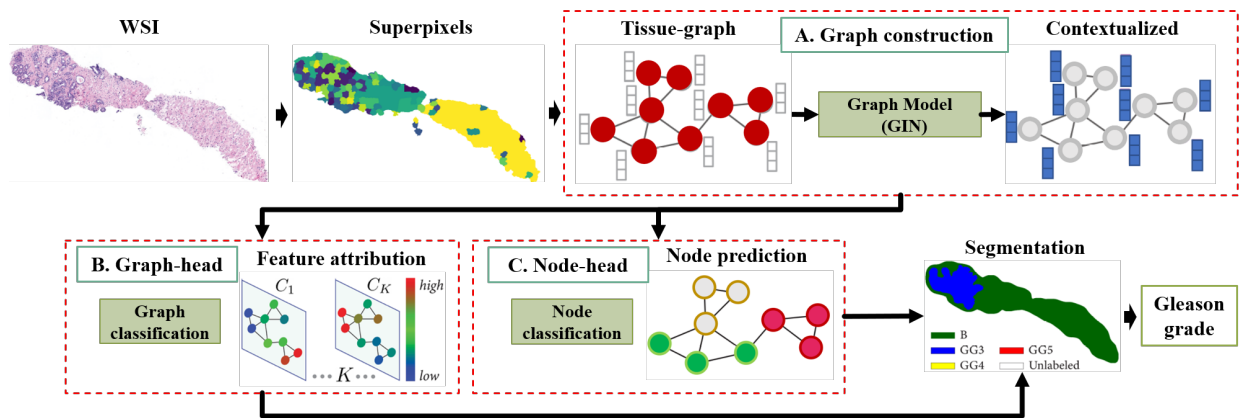


Fig. 14: Representation of the proposed SegGini methodology. a) Tissue graph construction with tissue superpixels as nodes, and edges computed using a region adjacency graph from the spatial connectivity of superpixels. A GNN is used to learn discriminative node embeddings to perform semantic segmentation. b) Graph-head: graph classification and feature attribution based on GraphGrad-CAM. c) Node-head: node classification. Recreated from [42].

4) *Skin cancer*: One of the most common types of skin cancer is basal cell carcinoma (BCC) which can look similar to open sores, red patches and shiny bumps. Several studies have demonstrated the ability to identify BCC from pathological images. Wu et al. [47] introduced a model that predicted BCC on WSI using a weakly- and semi-supervised formulation by combining prior knowledge from expert observations, and structural information between patches into a graph-based model. A sample of this prior knowledge is the fact that a dense patch with predictive cancer cells is more likely to have a cluster of cancer cells, and more patches with high cancer likelihoods increase the overall likelihood of an image being positive.

The framework consists of two modules, a GCN that propagates supervisory information over patches to learn patch-aware interpretability in the form of a probability score; and an aggregation function that connects patch-level and image-level predictions using prior knowledge. The proposed model makes full use of different levels of supervision, using a mix of weak supervision from image-level labels and available pixel-wise segmentation labels as a semi-supervised signal. By incorporating prior knowledge and structure information, both image-level classification and patch-level interpretation are significantly improved.

5) *Prostate cancer*: Pathologists must go above and beyond normal clinical demands and norms when precisely annotating image data. As a result, a semantic segmentation method should be able to learn from inexact, coarse, and image-level annotations without complex task-specific post-processing steps. To this end, Anklin et al. [42] proposed a weakly-supervised semantic segmentation method based on graphs (SegGini) that incorporates both local and global inter-tissue-region relations to perform contextualized segmentation using inexact and incomplete labels. The model is evaluated on the UZH (TMAs) [93] and SICAPv2 (WSI) [101] prostate cancer datasets for Gleason pattern segmentation and Gleason grade classification. Fig. 14 depicts the proposed SegGini methodology. A tissue-graph representation for an input histology image is constructed as proposed in [17], where the graph

nodes depict tissue superpixels. As the rectangular patches can span multiple distinct structures, superpixels are used [59]. To characterize the nodes, morphological and spatial features are extracted, and the graph topology is computed with a region adjacency graph (RAG) [62], using the spatial connectivity of superpixels.

Given a tissue graph, a GIN model learns contextualized features from the tissue microenvironment and inter-tissue interactions to perform semantic segmentation, where the proposed SegGini model assigns a class label to each node. The resulting node features are processed by a graph-head (image label), a node-head (node label), or both, based on the type of weak supervision. The graph-head consists of a graph classification and a feature attribution technique. The authors employed GraphGrad-CAM to measure importance scores towards the classification of each class, where the node attribution maps determine the node labels. Further, the authors in [42] found that the node-head simplifies image segmentation into classifying nodes where the node labels are extracted by assigning the most prevalent class within each node. For inexact image label and incomplete scribbles, both heads are jointly trained to improve the individual classification tasks. The outcomes of the heads are used to segment Gleason patterns. Finally, to identify image-level Gleason grades from the segmentation map, a classification approach [90] is used. SegGini outperforms prior models such as HistoSegNet [116] in terms of per-class and average segmentation, as well as classification metrics. This model also provides comparable segmentation performance for both inexact and complete supervision; and can be applied to a variety of tissues, organs, and histology tasks.

C. Hierarchical graph representation (macro and micro architectures)

In previous approaches, pathological images have been represented by cell-graphs, patch-graphs or tissue-graphs. However, cellular or tissue interactions alone are insufficient to fully represent pathological structures. A cell-graph incorporates only the cellular morphology and topology, and discards

tissue distribution information that is vital for appropriate representation of histopathological structures. A tissue-graph made up of a collection of tissue areas, on the other hand, is unable to portray the cell microenvironment. Thus, to learn the intrinsic characteristics of cancerous tissue it is necessary to aggregate multilevel structural information, which seeks to replicate the tissue diagnostic process followed by a pathologist when analyzing images at different magnification levels.

1) *Breast cancer*: Early detection of cancer can significantly reduce the mortality rate of breast cancer, where it is crucial to capture multi-scale contextual features in cancerous tissue. Combinations of CNNs have been used to encode multi-scale information in pathology images via multi-scale feature fusion, where scale is often associated with spatial location.

Zhang and Li [29] introduced a multi-scale graph wavelet neural network (MS-GWNN) that uses graph wavelets with different scaling parameters in parallel to obtain multilevel tissue structural information in a graph topology. The graph wavelet neural network (GWNN) [77] replaces the graph convolution in a spectral GCN with the wavelet transform which has an excellent localization capability. For breast cancer classification, the authors first transformed pathological images into graph structures where nodes are non-overlapping patches. Then, node classification is performed via a GWNN at different scales in parallel (node-level prediction). After that, multi-level node representations are incorporated to perform graph-level classification. The results and the visualization of the learned node embeddings demonstrated the strong capacity of the model to encode different structural information on two public datasets: BACH [9] and BreakHis [102]. However, this approach is limited by the manual selection of the appropriate scaling parameter.

A hierarchy defined from the cells with learned pooling layers [39] does not include high-level tissue features and approaches that concatenate cell-level and tissue-level information [32] cannot leverage the hierarchy between the levels of the tissue representation. To address these issues, Pati et al. [17] proposed a hierarchical-cell-to-tissue (HACT) representation that utilizes both nuclei and tissue distribution properties for breast cancer subtype classification. The HACT representation consists of a low-level cell-graph (CG) that captures the cellular morphology and topology; a tissue-graph (TG) at a high-level that captures the properties of the tissue sections as well as their spatial distribution; and the hierarchy between the cell-graph and the tissue-graph that captures the cells' relative distribution within the tissue.

Fig. 15 illustrates samples of the CG, the TG and the hierarchical cell-to-tissue representation. To construct a CG, each node represents a cell and edges encode cellular interactions, where for each nucleus hand-crafted features such as shape, texture and spatial location are extracted. Then, a KNN algorithm is adopted to build the initial topology based on the assumption that a close cell should be connected and a distant cell should remain disconnected. The Euclidean distances between nuclei centroids in the image space are used to quantify cellular distances. The TG is constructed by first identifying tissue regions (*e.g.*, epithelium, stroma, lumen, necrosis) by

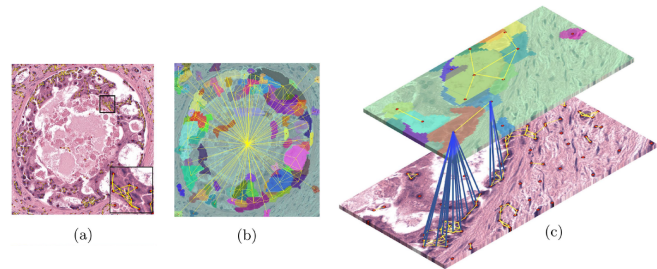


Fig. 15: Representation of a) CG, b) TG, and c) Hierarchical cell-to-tissue. Image adapted from [17].

detecting non-overlapping homogeneous superpixels of the tissue and iteratively merging neighboring superpixels that have similar colour attributes. The TG topology is generated assuming that adjacent tissue parts should be connected by constructing a region adjacency graph [62] with the spatial centroids of the superpixels. The HACT representation, that jointly represents the low-level (CG) and high-level (TG) relationships, is processed with a hierarchical model (HACT-Net) that employs two GIN models [73]. The learned cell-node embeddings are combined with the corresponding tissue-node embeddings to predict the classes.

To demonstrate the hierarchical-learning, the authors introduce the BRACS dataset to classify five breast cancer subtypes: normal, benign, atypical, ductal carcinoma in situ, and invasive. The authors also evaluate the generalizability to unseen data by splitting the data at the WSI-level (two images from the same slide do not belong to different splits) different from previous approaches that split at the image-level [15], [39]. The enriched multi-level HACT representation for classification outperformed CNN-based models and standalone cell-graph and tissue-graph models, confirming that for better structure-function mapping, the link between low-level and high-level information must be modelled at the local node level rather than at the graph level.

Later, Pati et al. [31] exploited hierarchical modeling for interpretability in digital pathology, aiming to map the tissue structure to tissue functionality. The authors adopt the hierarchical entity-graph representation of a tissue which is processed via a hierarchical GNN to learn the mapping from tissue compositions to respective tissue categories. In this work, Pati et al. [31] improved the HACT representation and the HACT-Net model. HACT-Net is modeled using principal neighborhood aggregation (PNA) [117] layers, which use a combination of aggregators to replace the sum operation in GIN and adopt degree-scalers to amplify or dampen neighboring aggregated messages based on the degree of a node. Graph normalization followed by batch normalization is incorporated after each PNA layer [118], which aids the network in learning discriminative topological patterns when the number of nodes within a class varies dramatically. To further assess the quality of the methodology, a comparison with independent pathologists is conducted. Three board-certified pathologists were recruited to annotate the BRACS test set without having access to the respective WSIs. The results indicate that the model outperforms the domain experts in the 7-class classification

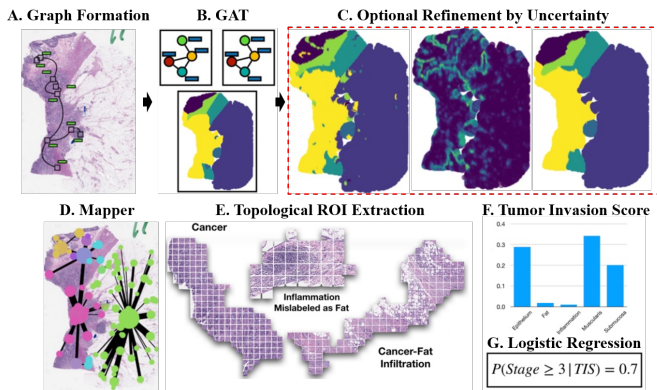


Fig. 16: A-C. Patch-level embeddings, graph representation and classification via a GCN. A refinement phase is incorporated through estimation of uncertainty. D-E. The Graph Mapper summarizes high-order relationships over a WSI as a graph, where meaningful histology regions are captured. F-G. Tumor invasion scores are used in the prediction model to form an interpretable staging score. Image adapted from [16].

task. The authors employed the GraphGrad-CAM to highlight the nuclei and tissue region nodes to show what the HACT-Net focuses on while classifying the tumor regions-of-interest.

2) *Colorectal cancer*: Tumor staging includes both tissue and nodal stages, with higher numbers indicating a greater depth of invasion and a greater number of lymph nodes implicated in the tumor, respectively. Levy et al. [16] introduced a framework that used varied levels of structure to learn both local and global patterns from histological images for determining the degree of tumor invasion. Fig. 16 illustrates the proposed framework where the authors combined GCNs to explain the mechanisms by which tissue regions interact, and topological feature extraction methods [119] to extract essential contextual information. Patch-level classification of colon sub-compartments was conducted via a GCN as well as a refinement of patch-level predictions, in which nodes with high uncertainty were deleted, and the remaining class labels were propagated to unlabeled patches. A topological data analysis (TDA) tool for graphs known as Graph Mapper [89] was adopted as a post-hoc model explanation technique to elucidate the high-level topology of the WSI. The mapper generates a graph in which each node represents a cluster of WSI patches and each edge represents the degree of shared patches between the clusters. This tool can offer higher level information flow descriptors in a GNN model, substantially simplifying analysis. With the regions of interest (collection of patches) extracted with the mapper, the authors compute tumor invasion scores that measure the degree of overlap between the tumor and adjacent tissue region. Finally, cancer staging is predicted via derived invasion scores using a private colon and lymph node dataset collected from the Dartmouth Hitchcock Medical Center, where the results demonstrated the potential of topological methods in the analysis of GNN models.

D. Unimodal and multi-modal feature level fusion

In this subsection we introduce works that have used fusion techniques to extract and combined multiple rich visual

representations of the same input data (unimodal fusion), or integrate information from various input modalities (multi-modal fusion) to enable more accurate and robust decisions. The former involves integrating several feature sets acquired from different networks into a single vector, which is then used for classification. This fusion occurs in two stages: normalization of a feature, and selection of a feature. The latter seeks to correlate and combine disparate heterogeneous modalities, such that the model can learn pairwise feature interactions and control the expressiveness of each modality. The main challenges in multi-modal data fusion are the dissimilarity of the data types being fused, and the interpretation of the results.

1) *Unimodal fusion (Cervical cancer)*: Cervical cancer is one of the most common causes of cancer death in women, and screening for abnormal cells from a cervical cytology slide is a common procedure for early detection of cervical cancer. In contrast with conventional CNNs which learn multi-level features through hierarchical deep architectures, Shi et al. [25] combined a GCN output with deep CNN features to classify images of isolated cervical cells into five and seven classes using the SIPakMeD [103] and Motic (liquid-based cytology image) [46] datasets, respectively.

First a CNN model pretrained for a cervical cell classification task is used to extract features of each individual cervical cell image. Then, K-means clustering is computed on the extracted features from all images to construct a graph where the centre of each cluster represents a node. The constructed graph of intrinsic similarities can be used to further investigate the potential relationships between images. Consequently, a stacked two-layer GCN generates a relation-aware representation which is encoded into CNN features for classification, as illustrated in Fig. 17. The authors demonstrated that the relation-aware representation generated by the GCN greatly enhances the classification performance. Extensive experiments to validate the performance of cervical cytology classification with a GCN were also published by the same authors in [46].

2) *Multi-modal fusion (Renal cancer)*: To predict clinical outcomes, oncologists often use both quantitative and qualitative information from genomics and histology [120]. However, current automated histology methods do not take genomic details into account. The following work exploits the complementary knowledge within morphological information and molecular information from genomics to better quantify tumors using graph-based methods.

Renal cell carcinoma is the most common malignant tumor of the kidney, and it is a diverse category of tumor with varying histology, clinical outcomes, and therapeutic responses. Renal cell carcinoma subtypes can be automatically classified through Deep learning frameworks. These algorithms can also identify features that predict survival outcomes from digital histopathological images. Several authors have used GCNs for cancer histology classification, however, its application to survival outcome prediction is less explored. Chen et al. [32] proposed a framework for multi-modal fusion of histology and genomic features for renal cancer survival outcome prediction on the TCGA datasets (glioma and clear cell renal cell carcinoma) [98], which contains paired whole slide images,

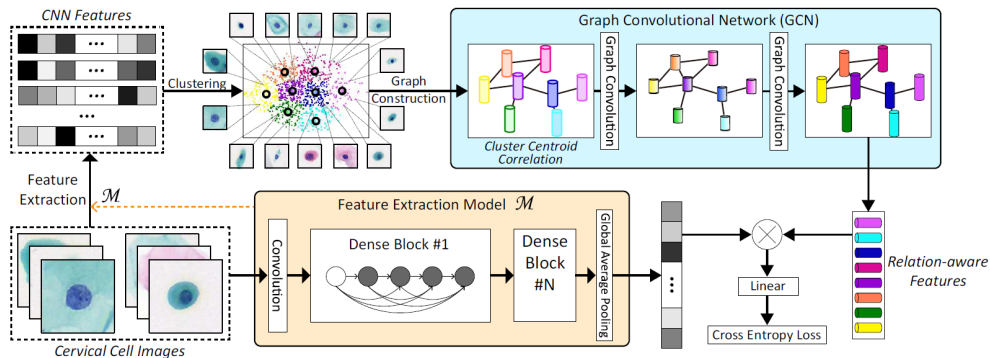


Fig. 17: Classification framework for cervical cell images. Features are extracted with a CNN pre-trained on a cervical cell classification task. K-means clustering is performed on these CNN features. A graph of cluster centroid correlations is built based on intrinsic similarities, and is the input to a GCN model. The encoded representations are incorporated into the CNN features for classification. Image reproduced from [25], [46].

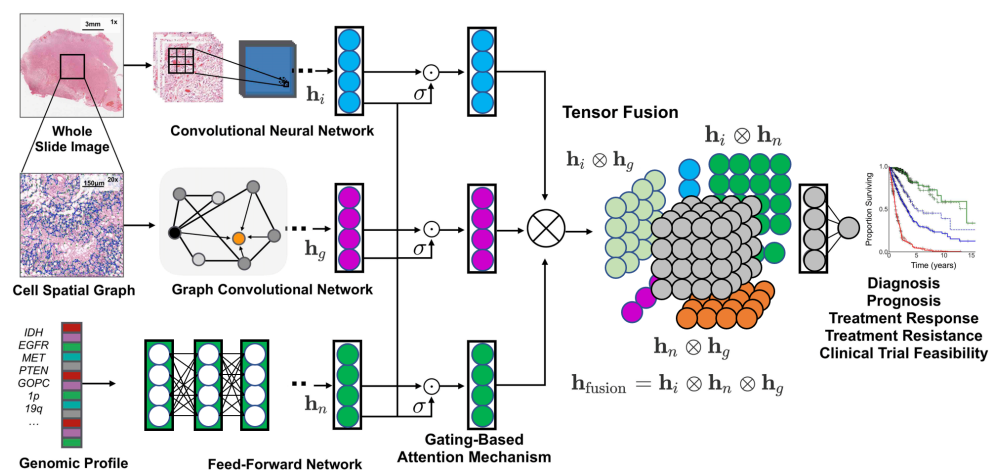


Fig. 18: An integrated framework for multi-modal fusion of histology and genomics features for survival outcome prediction. Image-based features using CNNs, and graph-based features using GCNs. Recreated from [32].

genotype, and transcriptome data. Their model fuses the histology image (patch features), cell-graph and genomic features into a multi-modal tensor that models interactions between the different modalities and outperforms deep learning-based feature fusion for survival outcome prediction. This framework is illustrated in Fig. 18.

The authors first extract morphological features from image-based features using CNNs, and graph-based features using GCNs, to learn cell-to-cell interactions in WSI. Cells are represented as nodes in a graph, with cells segregated using a nuclei segmentation method and connections established using KNN. CPC is also adopted as a self-supervised method for cell feature extraction. The authors adopted the aggregating functions of the GraphSAGE architecture. The hierarchical self-attention pooling strategy, SAGPool [80], is adopted to encode the hierarchical structure of cell graphs. Then, to monitor the expressiveness of each modality, a gating-based attention system is used to perform uni-modal function fusion. Multi-modal interpretability was considered by adopting an integrated gradient method for visualizing image saliency feature importance.

IV. DISCUSSION AND OPEN CHALLENGES

Beyond generating predictions relating to biology and medicine at molecular, genomic and therapeutic levels [24], graph representation learning has also been used to support medical diagnosis through the representation of patient records as graphs by using information including brain electrical activity, functional connectivity and anatomical structures [21]. As demonstrated throughout this review, graph-based deep learning has been successfully used to capture phenotypical and topological distributions in histopathology to better enable precision medicine. Numerous entity-graph based tissue representations and GNN models have been proposed for computer-aided detection and diagnosis of breast, colorectal, prostate, lung, lymphoma, skin, colon, cervical and renal cancers.

Given the utility of graphs across biomedical domains, especially to model the histology of cancer tissue, there has been a major push to exploit recent developments in deep learning for graphs in this domain. However, these applications are still in their nascent stages compared to existing research concerning conventional deep learning methods. There are challenges associated with the adoption of GNNs, and there are graph approaches yet to be explored in this domain that

potentially allow a more robust and comprehensive investigation of complex biological processes that merit further investigation. In this section, we discuss several future research directions that need to be addressed to unlock the full power of graph deep learning in digital pathology: 1) Entity graph construction; 2) Embedding expert knowledge and clinical adoption of graph analytics; 3) Complexity of graph models; 4) Training paradigms; and 5) Explainability of graph models.

A. Entity-graph construction

Defining an appropriate graph representation where vertices correspond to entities, and edges represent the connectivity of these entities, is highly relevant. Given a pathology task, different choices of entities in histology images can be selected as the relevant biological structures. Several graph representations have been customized according to the relevant entity such as nuclei, tissue regions, glands or just traditional patches. However, in the majority of methods discussed in this survey, graph structures are designed manually.

1) *Pros and cons of current preprocessing steps for entity-graph construction:*

a) *Entity definition:* Cell-graphs have been one of the most popular graph representations, where cells are the entities used to encode cell microenvironments, including morphology of cells and cellular interactions. Such cell-graph representations were proposed in [15], [28], [30], [32]–[34], [38], [39]. However, modeling a WSI as a cell-graph is non-trivial due to the large number of cells and the many possibly isolated cells and weak nuclear boundaries. This representation relies heavily on cell detection or segmentation methods. Although some works have used representative node sampling [39] or agglomerative clustering [36] to remove redundancy in the graph and reduce computation cost, the majority of cell-graph based proposals assume that cell-cell interactions are the most salient sources of information. Cell-graphs do not exploit tissue macro-architectural structures, or the hierarchical nature of the tissue.

Another traditional technique for analysing WSI that include context information of ROIs is patch-graphs. Although patch-graph representations have been adopted in a number of studies [26], [35], [37], [40], [41], [43]–[45], [47], not all entities are biologically-defined and methods are limited by the patch definition. The resolution and optimal size of each image patch and the level of context offered are trade-off against one another, and are determined by the data. For example, variations in glandular morphology and size make determining an acceptable image patch size problematic. Operating at lower magnification levels may not capture cell-level features, and higher resolutions limits the ability to capture the tissue micro-environment. Thus, an automated technique that defines these patch regions and an appropriate scaling parameter from the input data is vital.

To improve the tissue structure-function mapping, graph representations based on tissue regions have been proposed, which can also deal with one of the limitations of cell-graph as important regions may not need to only contain cells [27], [36], [42]. Tissue-graphs represent well-defined tissue regions and are used to propagate information across neighboring nodes

in a progressive manner at a gland or region level. Although superpixel-based approaches are proposed to address patch-graph limitations, a tissue-graph alone cannot capture local cellular information. A combination of cell-level and patch-level features was proposed to capture local and global patterns from histological images [32]. However, this fusion approach cannot take advantage of the hierarchy between levels.

Hierarchical graph representations were proposed as an adequate tissue representation as histological structures cannot be fully represented by cellular or tissue interactions alone. It has been shown that cell-graphs and tissue-graphs provide valuable complementary information (cellular and tissue interactions) to learn the intrinsic characteristics of cancerous tissues. Such hierarchical analysis that captures multivariate tissue information at multiple levels has been addressed only by [16], [17], [29], [31]. Nevertheless, this approach is still dependent on the construction of a cell-centered graph, which itself is limited by cell detection accuracy and is subjected to the complexity constraints of the model driven by the number of nodes. Other works have dealt with cell detection limitations by exploiting graph wavelets with different scaling parameters [29] to obtain multilevel tissue structural information in a tissue-graph. Further, in [16] micro- and macro architectures of histology images were captured with the combination of a topological data analysis tool (cell-level) and GCN (tissue-level).

b) *Feature extraction:* Handcrafted and CNN-based features have been the typical methods to characterize entities. Such deep feature extraction allows use of features from a pre-trained deep architecture. However, the performance of these methods is compromised because the authors usually utilize a pre-trained model (*e.g.*, trained on ImageNet) due to a lack of patch labels to fine-tune the network, and thus suffer from the domain gap between natural scene images and histopathological images. To address this limitation, a small number of works trained a feature extractor using self-supervised approaches such as CPC, VAE-GAN and auto encoder in [15], [27], [32], [40].

c) *Graph topology:* On current entity-graphs, each node is only connected to its spatially nearest neighbors, resulting in relatively limited information exchange during the message passing phase. Only one approach to date has computed the connections between nodes by using an adjacency learning layer in an end-to-end manner that considered the global context of all patches [43]. Edge embeddings in cell-graph and tissue-graph topologies are a poorly studied field with few approaches. Learning takes place primarily at the vertices, with edge attributes serving as auxiliary information. The EGNN has only been applied in [38] for colorectal cancer classification, and shows similar performance to the best model based on a 1-dimensional GNN [108]. Edge attributes can also directly inform the message passing phase operating over the vertices. In the MEGNet [121] model, vertices are updated by an aggregation of features from adjacent edges.

2) *Automated graph generation:* Automated graph structure estimation aims to find a suitable graph to represent the data as input to the GNN model. By modeling graph generation as a sequential process, the graph representation (nodes, edges

and embeddings) can be inferred directly from data which would be especially useful when representing tissues with a variety of complex micro- and macro environments. However, the majority of methods surveyed follow a standard sequential workflow which is highly dependent on the individual performance of each preprocessing step, including tissue mask detection, nuclei detection, super-pixel detection, deep feature extraction, and graph building. The use of neural networks to build generative graph models is gaining popularity to capture both their topology and their attributes, which can in turn lead to more robust algorithms and help to provide more accurate results. However, the effectiveness of such algorithms have not been investigated for histopathology images. Therefore, several requirements are still needed to enable the generation process.

Several works that have adopted GCNs for brain electrical activity analysis tasks [21] have demonstrated that learning the graph structure from data improves classification performance in comparison to approaches where a predefined graph topology is used. In digital pathology these predefined parameters per histology task are represented by fixed threshold to differentiate non-tissue pixels; patch size and number of patches for nuclei detection, and nuclei and tissue feature extraction; sample ratio of representative nuclei; thresholded KNN and distance that define topology and edges; the number of superpixels and downsampling factor per image; and selection of handcrafted features and CNN layer from which deep features are extracted. Such definitions limit the generalization of entity-graphs to different tissues, organs, and histology tasks. Some graph generation approaches that are worthy of exploration within histopathology diagnosis are GraphGAN [122], DGMG [123], and GCPN [124]. For instance, DGMG [123] can be used to generate one node at a time from each histopathology patch and then create edges one by one, to connect each node to the existing partial graph using probabilistic dependencies among nodes.

In summary, the preceding discussion exemplified the difficulties in estimating a graph structure with the desired properties from data. While there is emerging work in this field, it is ripe for further investigation. In digital pathology, automated graph generation, in which a graph model infers structural content from data, and the integration of domain knowledge, are also underutilised.

B. Embedding expert knowledge and clinical adoption of graph analytics

Incorporating domain knowledge into the model has emerged as a promising method for improving medical image analysis [125]. The use of graph-based mappings with label representations (word embeddings) have been investigated to guide information propagation among nodes [126]. For example, in basal cell carcinoma classification [47], the embedding knowledge is represented by encoding patches based on prior expert knowledge, which bridges the gap between patch-level and image-level predictions and results in better performance. Further, pathologists' feedback can help to improve the graph representation in terms of how to best mirror the biological relationship between cells and tissues. Thus, graph-based

analysis motivates exploring the inclusion of task-specific pathological prior knowledge in the construction of the graph representations [17].

Another open research question is how to incorporate interdisciplinary knowledge in a principled way, rather than on a case-by-case basis. Integrating electronic health records for personalized medicine can also boost the diagnostic power of digital pathology. The hierarchical information inherent in medical ontologies naturally lends itself to creating a rich network of medical knowledge, and other data types such as symptoms and genomics [127]. Thus, by integrating patient records into the graph representation learning environment, tailored predictions can be generated for individual patients.

Among the AI-techniques, graph-based tissue image analysis demonstrated performance superior or comparable to domain experts in breast cancer analysis [31]. These results combined with studies examining the effect of explanations on clinical end-user decisions [33] show generally positive results in the translation of this technology into diagnostic pathology. Such translation will require to considered integration of standardised technologies into digital pathology workflows, resulting in an integrated approach to diagnosis and offering pathologists new tools that accelerate their workflow, increase diagnostic consistency, and reduce errors.

While, there is considerable promise for graph analytics in digital pathology, there are some challenges ahead. These include, for example, the ability to generalize a diagnosis technique to a large population of patients which contain outliers; and to develop problem-solving skills that demand complex interactions with other medical disciplines. Thus, more work should be conducted to investigate how a pathologist could refine a graph model decision via a human-in-the-loop system [128], [129] Such approaches provide an important safety mechanism for detecting and correcting algorithmic errors that may occur. A remaining challenge here is to provide frameworks with the above functionalities with reduced complexity to lower the barriers between the systems and clinicians, to help facilitate system uptake.

Entity-graph analysis has the ability to transform pathology by providing applications that speed up workflow, improve diagnosis, and improve patient clinical outcomes. However, there is still a gap between research studies and the effort required to deliver reliable graph analytics that incorporate expert knowledge into the system, and can be integrated into existing clinical workflows.

C. Complexity of graph models

Graph-based approaches for histology analysis have a high representational power, and can describe topological and geometric properties of multiple types of cancers. When compared to pixel-based approaches, the graph representation can more seamlessly describe a large tissue region. However, classical graph-based models have a high computational complexity. As a result, in the suggested learning approach, the choice of GNN architecture should be handled as a hyper-parameter.

The most common GNNs used by methods in this survey include ChebNet [66], GCN [67], GraphSAGE [68], GAT [72], GIN [73], and variants such as Adaptive GraphSAGE [39],

RSF [75], MS-GWNN [29] and FENet [41]. Spatial-GCNs such as GraphSAGE and GIN demonstrated their learning ability using max-, mean-, or sum-pooling aggregators. GIN has been particularly effective in computational pathology with a provably strong expressive power to learn fixed-size discriminative graph embeddings from cellular and tissue architectures in WSIs, which demonstrate translation and rotation invariance. However, it is noted that these GNN models inherit considerable complexity from their deep learning lineage, which can be burdensome when scaling and deploying GNNs. This is likely one of the reasons that has seen patch-based approaches remain a popular approach for many problems.

The training of GNNs remains one of the most difficult tasks due to their high memory consumption and inference latency compared to patch-based deep learning approaches. GNNs usually require the whole graph and the intermediate states of all nodes to be saved in memory. However, the adoption of an efficient training approach is uncommon in the applications surveyed. Various graph sampling approaches have been proposed as a way to alleviate the cost of training GNNs. Rather than training over the full graph, each iteration is run over a sampled sub-graph, whether they are sampled node-wise (GraphSage [68]), layer-wise (FastGCN [130], L^2 -GCN [131]), or by clustering (Cluster-GCN [132]).

Some works have proposed more efficient and simple architectures that deserve attention for their potential to be adopted in computational histopathology. The simple graph convolution (SGC) [133] reduces the complexity of GCNs by repeatedly removing the non-linearities between GCN layers and collapsing multiple weight matrices into a single linear transformation. This model was adopted for emotion recognition and increased the performance speed with a comparable classification accuracy in comparison to other networks [134]. The simple scalable inception GNN (SIGN) [135] is explicitly designed as a shallow architecture that combines graph convolutional filters of different sizes that allow efficient pre-computation. The efficient graph convolution (EGC) [136] method does not require trading accuracy for runtime memory or latency reductions based on an adaptive filtering approach. GNNs can also deliver high performance for feature matching across images [137], which can be incorporated for content-based histopathological image retrieval.

It is also important to highlight that some works exploit the cell-graph representation without the complexity of GCN processing. The tissue classification problem was proposed in [138] as a cellular community detection based on cell detection and classification into distinct cellular components (cell-graphs), and clustering of image patches (patch-level graphs) into biologically meaningful communities (specific tissue phenotype). The concept of constructing a graph and then using geodesic distance for community detection has outperformed deep neural networks and graph-based deep learning methods such as ChebNet, GCNs and deep graph infomax learning (DGI) [139].

In the coming years, a key research topic will be how to effectively learn and compute GNNs in order to realise their full potential. Deep learning on graphs is inherently difficult due to the graphs' complex topological structure, which can be

made up of many different types of entities and interactions. As such, the appropriate selection of key parameters of a model prior to representation learning is essential to capture the structural information of the histopathology slides.

D. Training paradigms

As stated in previous sections, training paradigms can be divided into two main categories: training a network to learn the node embeddings used in the graph representation; and the training of the GNN model.

1) *Node embeddings*: The node embeddings are the features that are learned to represent the defined node (*e.g.* cells, nucleus, patches, super-pixels). Some of the embedding features extracted through attribute networks require labeled datasets and need to be trained in a supervised manner as explained in [16], [17], [25], [26], [28]–[31], [33]–[39], [41], [43], [44], [46], [47]. However, one of the main challenges in deep learning is the lack of large corpora of manually labeled data for training, which often imposes a limitation on problems in the medical domain. Thus, self-supervised methods are gaining interest to improve the quality of learned node embeddings [15], [27], [32], [40], [45] by learning embedding features directly from histopathology images, rather than relying on extracting features using transfer learning, which is discussed in Subsection IV-A.

2) *Node/graph classification*: Training a GCN for node or graph level classification can be performed in supervised, semi-supervised or even in a self-supervised manner. If sufficient labels are available for nodes or graph data, the common practice is a supervised training approach, such as the methods of [16], [17], [25], [28]–[31], [33], [34], [36]–[39], [41], [43], [45].

Though supervised methods can achieve high performance, they can place limitations on model complexity and can suffer when annotations are inconsistent or imprecise. In the absence of sufficient labeled data, weakly-supervised or semi-supervised frameworks are proposed to better capture the structure of histopathology data and reduce the human annotation workload. Although the issue of missing labels is not specific to the graph domain, only a few works have adopted such frameworks (pixel or patch level labels).

In semi- or weakly-supervised approach, the node embeddings are learnt from few labeled samples per class [15], [26], [46], [47]. For example, in a weakly supervised learning approach, the contributions of the individual patches to the ROI-level diagnosis are not known during training [26].

In addition to the above, extensive research over past years in deep learning [140]–[142] showed that a decision classifier based on Multiple Instance Learning (MIL) can boost the performance in classifying cancer by aggregating instance-level predictions. MIL only requires labels for the bag of instances rather than individual instances, which makes it well-suited for histology slide classification. One example is CLAM (Clustering-constrained attention multiple instance learning [143]). Even though these approaches have practical merits and can consider the important patches for predicting the staging, they do not consider the spatial relationships between patches. Current multiple instance learning approaches

using deep graphs [27], [40], [43] follow this line of research. They can seamlessly scale to arbitrary tissue dimensions by incorporating an arbitrary number of entities and interactions, thus offering an alternative to traditional MIL [143]. MIL methods can be incorporated with a GCN to take advantage of the structural information among instances [4]. For example, the SegGini model [42] outperforms several traditional state-of-the-art methods such as CLAM [143] and Context-Aware CNN (CACNN) [144] for weakly-supervised classification of prostate cancer.

Self supervised methods have also been successfully deployed as a training paradigm for GCNs. For example, Ozen et al. [35] adopted a SimCLR framework [111] along with contrastive loss to learn a representation of ROIs and perform classification. Although the aforementioned training paradigms demonstrate remarkable performance, few works [43] have considered end-to-end training and the challenge that brings such as dealing with complexity of constructing a graph or labeled data, and thus this requires investigation in future works.

Training paradigms are dependent on the availability of manually labeled data. In medical imaging and specifically histopathology obtaining a large set of labeled data is a tedious process and so weakly- and self-supervised algorithms are receiving increasing interest for learning node embeddings and performing graph classification. It is expected that in future, further research carry out on a large-scale to analyse histopathology data using GCNs in a weakly- or self-supervised manner.

E. Explainability of graph models

To effectively translate graph models into clinical practice, clinicians' trust must be established. Explainability, or a model's ability to justify its outcomes and therefore assist clinicians in understanding a model's prediction, has long been seen as crucial to building trust. Understanding model behaviour beyond traditional performance indicators has thus become an important part of machine learning research, particularly in healthcare [145].

Explainability in deep models has focused on providing input-dependent explanations and understanding model behavior from different perspectives, including visual explanations and highlighting salient regions. We can examine the sensitivity between the input features and the predictions, for example, by looking at the gradients or weights. We can also highlight important features or regions of an image by incorporating attention mechanisms [146]. Nevertheless, compared with traditional image domains, explainability and visualization of deep learning for graphs is less explored [84], yet explainability is critical to highlight informative structural compositions of tissue and inter-nuclear relationships, as is desired for computational histopathology.

While interpretability approaches are generally lacking within most graph network methods, it is worth noting that a few methods exist and incorporate such explanations in digital pathology as illustrated in Table II: i) In [47] a GCN propagated supervisory information over patches to learn patch-aware interpretability in the form of a probability score. ii) A

robust spatial filtering with an attention-based architecture and node occlusion was used to capture the contribution of each nucleus and its neighborhood to the prediction [30]. iii) The Graph Mapper, a topological data analysis tool, was adopted to compress histological information to its essential structures, where meaningful histology regions are captured [16]. iv) In [32], an integrated gradient method was used to visualise image saliency feature importance. v) A graph clustering visualization was used in [39] to group cells with similar tissue structures. vi) A post-hoc graph-pruning explainer, GCExplainer, was designed to identify decisive cells and interactions from the input graph [34]. vii) The gradient-based saliency method, GraphGrad-CAM, was adopted in [31] and [42] to measure importance scores and regions that contributed towards the classification of each class.

The majority of approaches that have incorporated explainers are limited to cell-graph analysis. Considering the pathologically aligned multi-level hierarchical tissue attributes [31], the interpretability can reveal crucial entities such as nuclei, tissue parts and interactions which can mimic the pathologist's assessment and therefore, increase the level of trust between experts and AI frameworks.

Existing works however lack the definition of objectives to validate a model in terms of effective explainability, and only a single work has looked at the quality and utility of the proposed explanation methods for the intended audience (*i.e.* clinicians). In [33], the authors evaluated several graph explainers (GNNExplainer, GraphGrad-CAM, GraphGrad-CAM++, GraphLRP) to provide domain-understandable quantitative metrics based on pathologically measurable cellular properties, to make graph decisions understandable to pathologists. The authors found that at the concept-level, GraphGrad-CAM++ has the highest overall agreement with the pathologists, followed by GraphGrad-CAM and GNNExplainer.

Other methods not investigated in this survey that focus on instance-level interpretation of deep graph models that deserve attention in digital pathology for explainability at the node, edge, or node feature levels are: excitation BP [83], PGM-explainer [147], GraphMask [148], Graphlime [149], and Relax [150]. Other methods such as SubgraphX [151] provide subgraph-level explanations which may be more intuitive and human-intelligible for digital pathology.

Knowing the subset of features from which the model outcome is derived is critical. This allows clinicians to compare model decisions with clinical judgement, which is especially useful when there is a discrepancy. It is also worth noting that clinicians expect variation in the importance of inputs to exist both across patients and populations [145]. However, the explanations provided by methods discussed in this survey using gradient-based (GraphGrad-CAM) and perturbation-based methods (GNNExplainer) are limited to single instances. To verify and understand a deep model, pathologists need to check explanations for all input graphs, which is time-consuming and impractical. Models that interpret each instance independently, as previously stated, are insufficient to provide a global understanding of the trained model [152]. Thus, methods to provide GNN predictions on a group of instances collectively (*i.e.* a population) and provide a global understanding of GNN

predictions is less explored in the literature.

Instance-level methods explain GNNs with respect to each input graph, whereas model-level methods explain GNNs without regard for any specific input example. The latter specifically investigates what input graph patterns can lead to a specific GNN behaviour, such as maximising a target prediction. However, no research on interpreting GNNs at the model-level exists in digital pathology. XGNN [153] provides model-level explanations by training a graph generator to build graph patterns that optimize a specific model prediction. The authors formulated graph creation as a reinforcement learning problem, with the graph generator predicting how to add an edge to a given graph and build a new graph at each step. The generator is then trained using a policy gradient based on feedback from the trained graph models. Several graph rules are also used to ensure that the explanations are both valid and human-readable. PGExplainer [154] can also provide an explanation for each instance with a global view of the GNN model by incorporating a generative probabilistic model. Nonetheless, it is unknown whether XGNN and PG-Explainer can be used to perform node classification tasks for histopathology analysis, which is an important area for future research.

Given the trend of graph-based processing for a variety of applications in computational pathology, graph explainability and quantitative evaluation with a focus on clinician usability are critical. Interpretability is essential because it can aid, for example, in informed decision-making during cancer diagnosis and treatment planning. However, interpretability of GNNs within digital pathology has received insufficient attention to date.

V. CONCLUSION

Through the use of whole-slide images (WSIs) and tissue microarrays (TMAs), digital pathology has transformed pathology diagnosis. The growing use of this data has also given rise to a new field of study known as computational pathology, which aims to develop machine learning techniques to provide more objective and reproducible results. Deep learning, in particular Convolutional Neural Networks (CNNs), have demonstrated efficacy in visual representation learning in digital pathology. To obtain image-level representations, mainstream CNN architectures typically aggregate feature representations over fixed-sized patches of the WSI. However, the patch-wise and pixel-based processing used by CNNs lacks the ability to capture global contextual information relating to meaningful entities such as cells, glands, and tissue types. As demonstrated throughout this review, histopathology knowledge graphs enable the capture of more comprehensive and interpretable information relating to the underlying mechanisms of a disease. Several works have attempted to adopt graph-based deep learning models to learn both local and global patterns. Entity-based analysis has the potential to improve the interpretability of deep learning techniques by identifying decisive nuclei, tissue regions and interactions. This can also potentially replicate holistic and context aware parts of a pathologist's assessment.

Our survey has provided a detailed overview of a new rapidly growing field of representation learning for computational histopathology. The enriched graph representation and learning in digital pathology has resulted in superior performance for diverse types of cancer analysis. Nevertheless, we highlight open research directions concerning the adoption of graph-based deep learning, including the explainability of graph representation learning, methods of graph construction, and the complexity of graph models and their limited training efficiency.

VI. CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest.

REFERENCES

- [1] S. Deng, X. Zhang, W. Yan, I. Eric, C. Chang, Y. Fan, M. Lai, and Y. Xu, "Deep learning in digital pathology image analysis: a survey," *Front. Med.*, pp. 1–18, 2020.
- [2] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.
- [3] J. van der Laak, G. Litjens, and F. Ciompi, "Deep learning in histopathology: the path to the clinic," *Nat. Med.*, vol. 27, no. 5, pp. 775–784, 2021.
- [4] C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Med. Image Anal.*, p. 101813, 2020.
- [5] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [6] F. Xing, Y. Xie, H. Su, F. Liu, and L. Yang, "Deep learning in microscopy image analysis: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4550–4568, 2017.
- [7] Y. He, H. Zhao, and S. T. Wong, "Deep learning powers cancer diagnosis in digital pathology," *Comput. Med. Imaging Graph.*, p. 101820, 2020.
- [8] M. Amgad, L. A. Atteya, H. Hussein, K. H. Mohammed, E. Hafiz, M. A. Elsebaie, A. M. Alhusseiny, M. A. AlMoslemany, A. M. Elmatboly, P. A. Pappalardo *et al.*, "Nucls: A scalable crowdsourcing, deep learning approach and dataset for nucleus classification, localization and segmentation," *arXiv preprint arXiv:2102.09099*, 2021.
- [9] G. Aresta, T. Araújo, S. Kwok, S. S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan *et al.*, "Bach: Grand challenge on breast cancer histology images," *Med. Image Anal.*, vol. 56, pp. 122–139, 2019.
- [10] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P.-A. Heng, J. Li, Z. Hu *et al.*, "A multi-organ nucleus segmentation challenge," *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1380–1391, 2019.
- [11] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.
- [12] S. Qi, W. Wang, B. Jia, J. Shen, and S.-C. Zhu, "Learning human-object interactions by graph parsing neural networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 401–417.
- [13] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, and Y. Kalantidis, "Graph-based global reasoning networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 433–442.
- [14] J. Li, X. Xie, Z. Zhao, Y. Cao, Q. Pan, and G. Shi, "Temporal graph modeling for skeleton-based action recognition," *arXiv preprint arXiv:2012.08804*, 2020.
- [15] J. Wang, R. J. Chen, M. Y. Lu, A. Baras, and F. Mahmood, "Weakly supervised prostate tma classification via graph convolutional networks," in *Proc. IEEE Int. Symp. Biomed. Imaging (ISBI)*, 2020, pp. 239–243.
- [16] J. Levy, C. Haudenschild, C. Bar, B. Christensen, and L. Vaickus, "Topological feature extraction and visualization of whole slide images using graph neural networks," in *Proc. Pac. Symp. Biocomput. (PSB)*, vol. 26, 2021, pp. 285–296.
- [17] P. Pati, G. Jaume, L. A. Fernandes, A. Foncubierta-Rodríguez, F. Ferrocce, A. M. Anniciello, G. Scognamiglio, N. Brancati, D. Riccio, M. Di Bonito *et al.*, "Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification," in *UNSURE and GRAIL in conjunction with MICCAI*, 2020, pp. 208–219.

- [18] I. Makarov, D. Kiselev, N. Nikitinsky, and L. Subelj, "Survey on graph embeddings and their applications to machine learning problems on graphs," *PeerJ Comput. Sci.*, vol. 7, 2021.
- [19] S. Georgousis, M. Kenning, and X. Xie, "Graph deep learning: State of the art and challenges," *IEEE Access*, 2021.
- [20] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, 2020.
- [21] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, "Graph-based deep learning for medical diagnosis and analysis: Past, present and future," *Sensors*, vol. 21, no. 14, 2021.
- [22] H. Sharma, N. Zerbe, S. Lohmann, K. Kayser, O. Hellwich, and P. Hufnagl, "A review of graph-based methods for image analysis in digital histopathology," *Diagnostic Pathol.*, vol. 1, no. 1, 2015.
- [23] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2013.
- [24] M. M. Li, K. Huang, and M. Zitnik, "Representation learning for networks in biology and medicine: Advancements, challenges, and opportunities," *arXiv preprint arXiv:2104.04883*, 2021.
- [25] J. Shi, R. Wang, Y. Zheng, Z. Jiang, and L. Yu, "Graph convolutional networks for cervical cell classification," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2019.
- [26] B. Aytüneş, S. Aksoy, R. G. Cinbiş, K. Kösemehmetoğlu, S. Önder, and A. Üner, "Graph convolutional networks for region of interest classification in breast histopathology," in *Med. Imaging 2020: Digit. Pathol.*, vol. 11320, 2020, p. 113200K.
- [27] A. Raju, J. Yao, M. M. Haq, J. Jonnagaddala, and J. Huang, "Graph attention multi-instance learning for accurate colorectal cancer staging," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2020, pp. 529–539.
- [28] D. Anand, S. Gadiya, and A. Sethi, "Histograms: graphs in histopathology," in *Med. Imaging 2020: Digit. Pathol.*, vol. 11320, 2020, p. 1132000.
- [29] M. Zhang and Q. Li, "Ms-gwnn: multi-scale graph wavelet neural network for breast cancer diagnosis," *arXiv preprint arXiv:2012.14619*, 2020.
- [30] M. Sureka, A. Patil, D. Anand, and A. Sethi, "Visualization for histopathology images using graph convolutional neural networks," in *Proc. IEEE Int. Conf. Bioinform. BioEng. (BIBE)*, 2020, pp. 331–335.
- [31] P. Pati, G. Jaume, A. Foncubierta, F. Feroce, A. M. Anniciello, G. Scognamiglio, N. Brancati, M. Fiche, E. Dubruc, D. Riccio *et al.*, "Hierarchical cell-to-tissue graph representations for breast cancer subtyping in digital pathology," *arXiv preprint arXiv:2102.11057*, 2021.
- [32] R. J. Chen, M. Y. Lu, J. Wang, D. F. Williamson, S. J. Rodig, N. I. Lindeman, and F. Mahmood, "Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," *IEEE Trans. Med. Imaging*, 2020.
- [33] G. Jaume, P. Pati, B. Bozorgtabar, A. Foncubierta, A. M. Anniciello, F. Feroce, T. Rau, J.-P. Thiran, M. Gabrani, and O. Goksel, "Quantifying explainers of graph neural networks in computational pathology," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2021, pp. 8106–8116.
- [34] G. Jaume, P. Pati, A. Foncubierta-Rodriguez, F. Feroce, G. Scognamiglio, A. M. Anniciello, J.-P. Thiran, O. Goksel, and M. Gabrani, "Towards explainable graph representations in digital pathology," in *Proc. Int. Conf. Mach. Learn. (ICML) Workshop CompBio*, 2020, pp. 5453–5462.
- [35] Y. Ozen, S. Aksoy, K. Kösemehmetoğlu, S. Önder, and A. Üner, "Self-supervised learning with graph neural networks for region of interest retrieval in histopathology," in *Proc. Int. Conf. Pattern. Recog. (ICPR)*, 2021, pp. 6329–6334.
- [36] W. Lu, S. Graham, M. Bilal, N. Rajpoot, and F. Minhas, "Capturing cellular topology in multi-gigapixel pathology images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2020, pp. 260–261.
- [37] H. Ye, D.-H. Wang, J. Li, S. Zhu, and C. Zhu, "Improving histopathological image segmentation and classification using graph convolution network," in *Proc. Int. Conf. Comput. Pattern Recog. (ICCP)*, 2019, pp. 192–198.
- [38] L. Studer, J. Wallau, H. Dawson, I. Zlobec, and A. Fischer, "Classification of intestinal gland cell-graphs using graph neural networks," in *Proc. Int. Conf. Pattern. Recog. (ICPR)*, 2021, pp. 3636–3643.
- [39] Y. Zhou, S. Graham, N. Alemi Koohbanani, M. Shaban, P.-A. Heng, and N. Rajpoot, "Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 0–0.
- [40] Y. Zhao, F. Yang, Y. Fang, H. Liu, N. Zhou, J. Zhang, J. Sun, S. Yang, B. Menze, X. Fan *et al.*, "Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2020, pp. 4837–4846.
- [41] K. Ding, Q. Liu, E. Lee, M. Zhou, A. Lu, and S. Zhang, "Feature-enhanced graph networks for genetic mutational prediction using histopathological images in colon cancer," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2020, pp. 294–304.
- [42] V. Anklin, P. Pati, G. Jaume, B. Bozorgtabar, A. Foncubierta-Rodríguez, J.-P. Thiran, M. Sibony, M. Gabrani, and O. Goksel, "Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs," *arXiv preprint arXiv:2103.03129*, 2021.
- [43] M. Adnan, S. Kalra, and H. R. Tizhoosh, "Representation learning of histopathology images using graph neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2020, pp. 988–989.
- [44] Y. Zheng, B. Jiang, J. Shi, H. Zhang, and F. Xie, "Encoding histopathological wsis using gnn for scalable diagnostically relevant regions retrieval," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2019, pp. 550–558.
- [45] R. Li, J. Yao, X. Zhu, Y. Li, and J. Huang, "Graph cnn for survival analysis on whole slide pathological images," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2018, pp. 174–182.
- [46] J. Shi, R. Wang, Y. Zheng, Z. Jiang, H. Zhang, and L. Yu, "Cervical cell classification with graph convolutional network," *Comput. Methods Biomed.*, vol. 198, p. 105807, 2020.
- [47] J. Wu, J.-X. Zhong, E. Z. Chen, J. Zhang, J. Y. Jay, and L. Yu, "Weakly- and semi-supervised graph cnn for identifying basal cell carcinoma on pathological images," in *Proc. Int. Workshop Graph Learn. Med. Imaging (GLMI)*, 2019, pp. 112–119.
- [48] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, 2013.
- [49] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.
- [50] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, p. 101563, 2019.
- [51] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P.-A. Heng, "Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation," in *Proc. Int. Conf. Inf. Process. Med. Imaging (IPMI)*, 2019, pp. 682–693.
- [52] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
- [53] F. Mahmood, D. Borders, R. J. Chen, G. N. McKay, K. J. Salimian, A. Baras, and N. J. Durr, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE Trans. Med. Imaging*, vol. 39, no. 11, pp. 3257–3267, 2019.
- [54] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.
- [55] J. Gamper, N. A. Koohbanani, K. Benet, A. Khuram, and N. Rajpoot, "Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification," in *Eur. Congr. Digit. Pathol.*, 2019, pp. 11–19.
- [56] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," *arXiv preprint arXiv:1109.2378*, 2011.
- [57] S. Kalra, H. R. Tizhoosh, C. Choi, S. Shah, P. Diamandis, C. J. Campbell, and L. Pantanowitz, "Yottixel—an image search engine for large archives of histopathology whole slide images," *Med. Image Anal.*, vol. 65, p. 101757, 2020.
- [58] J. N. Kather, N. Halama, and A. Marx, "100,000 histological images of human colorectal cancer and healthy tissue," Apr. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.1214456>
- [59] B. E. Bejnordi, G. Litjens, M. Hermsen, N. Karssemeijer, and J. A. van der Laak, "A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images," in *Med. Imaging 2015: Digit. Pathol.*, vol. 9420, 2015, p. 94200H.
- [60] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

- [61] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2016, pp. 770–778.
- [62] F. K. Potjer, "Region adjacency graphs and connected morphological operators," in *Math. Morphol. Appl. Image Signal Process.*, 1996, pp. 111–118.
- [63] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.
- [64] H. Zhang, J. Xue, and K. Dana, "Deep ten: Texture encoding network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2017, pp. 708–717.
- [65] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [66] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2016, pp. 3844–3852.
- [67] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2017.
- [68] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2017, pp. 1024–1034.
- [69] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2014.
- [70] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 1263–1272.
- [71] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2017, pp. 5998–6008.
- [72] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2018.
- [73] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2019.
- [74] L. Gong and Q. Cheng, "Exploiting edge features for graph neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 9211–9219.
- [75] F. P. Such, S. Sah, M. A. Dominguez, S. Pillai, C. Zhang, A. Michael, N. D. Cahill, and R. Ptucha, "Robust spatial filtering with graph convolutional neural networks," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 6, pp. 884–896, 2017.
- [76] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 5453–5462.
- [77] B. Xu, H. Shen, Q. Cao, Y. Qiu, and X. Cheng, "Spherical cnns on unstructured grids," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2019.
- [78] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2016.
- [79] R. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2018.
- [80] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *Proc. Int. Conf. Mach. Learn. (ICML)*. PMLR, 2019, pp. 3734–3743.
- [81] J. Arevalo, T. Solorio, M. Montes-y Gómez, and F. A. González, "Gated multimodal units for information fusion," *arXiv preprint arXiv:1702.01992*, 2017.
- [82] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *NAACL HLT*, 2016, pp. 1480–1489.
- [83] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martín, and H. Hoffmann, "Explainability methods for graph convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 10772–10781.
- [84] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnnexplainer: Generating explanations for graph neural networks," *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, vol. 32, p. 9240, 2019.
- [85] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
- [86] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2018, pp. 839–847.
- [87] R. Schwarzenberg, M. Hübner, D. Harbecke, C. Alt, and L. Hennig, "Layerwise relevance visualization in convolutional text graph classifiers," in *Proc. Graph-based methods Nat. Lang. Process. (TextGraphs)*, 2019, pp. 58–62.
- [88] F. Baldassarre and H. Azizpour, "Explainability techniques for graph convolutional networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019.
- [89] C. Bodnar, C. Cangea, and P. Liò, "Deep graph mapper: Seeing graphs through the neural lens," in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2020.
- [90] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rueschoff, and M. Claassen, "Automated gleason grading of prostate cancer tissue microarrays via deep learning," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.
- [91] L. Studer, S. Toneyan, I. Zlobec, H. Dawson, and A. Fischer, "Graph-based classification of intestinal glands in colorectal cancer tissue images," in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2019.
- [92] R. Awan, K. Sirinukunwattana, D. Epstein, S. Jefferyes, U. Qidwai, Z. Aftab, I. Mujeeb, D. Snead, and N. Rajpoot, "Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, 2017.
- [93] Q. Zhong, T. Guo, M. Recheisner, J. H. Rüschoff, N. Rupp, C. Fankhauser, K. Saba, A. Mortezaei, C. Poyet, T. Hermanns *et al.*, "A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients," *Sci. Data*, vol. 4, no. 1, pp. 1–9, 2017.
- [94] C. G. A. Network *et al.*, "Comprehensive molecular portraits of human breast tumours," *Nature*, vol. 490, no. 7418, p. 61, 2012.
- [95] C. Kandath, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski *et al.*, "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, no. 7471, pp. 333–339, 2013.
- [96] R. L. Ward and N. J. Hawkins, "Molecular and cellular oncology (mco) study tumour collection," *UNSW Australia*, 2015.
- [97] S. Kirk, Y. Lee, C. Sadov, S. Levine, C. Roche, E. Bonaccio, and J. Filiippini, "Radiology data from the cancer genome atlas colon adenocarcinoma [tcga-coad] collection," *The Cancer Imaging Archive*, 2016.
- [98] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemp. Oncol.*, vol. 19, no. 1A, p. A68, 2015.
- [99] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [100] B. S. Kramer, C. D. Berg, D. R. Aberler, and P. C. Prorok, "Lung cancer screening with low-dose helical ct: results from the national lung screening trial (NLST)," 2011.
- [101] J. Silva-Rodríguez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "Going deeper through the gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection," *Comput. Methods Programs Biomed.*, vol. 195, p. 105637, 2020.
- [102] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2015.
- [103] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images," in *Proc. IEE Int. Conf. Image Process. (ICIP)*, 2018, pp. 3144–3148.
- [104] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology," *Nat. Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, 2019.
- [105] H. Sharma, N. Zerbe, D. Heim, S. Wienert, S. Lohmann, O. Hellwich, and P. Hufnagel, "Cell nuclei attributed relational graphs for efficient representation and classification of gastric cancer in digital histopathology," in *Med. Imaging 2016: Digit. Pathol.*, vol. 9791, pp. 97910X.
- [106] S. Gadiya, D. Anand, and A. Sethi, "Some new layer architectures for graph cnn," *arXiv preprint arXiv:1811.00052*, 2018.
- [107] A. Fischer, K. Riesen, and H. Bunke, "Improved quadratic time approximation of graph edit distance by combining hausdorff matching and greedy assignment," *Pattern Recog. Lett.*, vol. 87, pp. 55–62, 2017.
- [108] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, "Weisfeiler and leman go neural: Higher-order graph neural networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, 2019, pp. 4602–4609.

- [109] S. Mehta, E. Mercau, J. Bartlett, D. Weaver, J. G. Elmore, and L. Shapiro, “Y-net: joint segmentation and classification for diagnosis of breast biopsy images,” in *Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2018, pp. 893–901.
- [110] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, and Y. Zhao, “Histopathological whole slide image analysis using context-based cbr,” *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1641–1652, 2018.
- [111] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 1597–1607.
- [112] M. Ye, X. Zhang, P. C. Yuen, and S.-F. Chang, “Unsupervised embedding learning via invariant and spreading instance feature,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 6210–6219.
- [113] P. Gupta, S.-F. Chiang, P. K. Sahoo, S. K. Mohapatra, J.-F. You, D. D. Onthoni, H.-Y. Hung, J.-M. Chiang, Y. Huang, and W.-S. Tsai, “Prediction of colon cancer stages and survival period with machine learning approach,” *Cancers*, vol. 11, no. 12, p. 2007, 2019.
- [114] M. Ilse, J. Tomczak, and M. Welling, “Attention-based deep multiple instance learning,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 2127–2136.
- [115] Z. Li, X. Zhang, H. Müller, and S. Zhang, “Large-scale retrieval for medical image analytics: A comprehensive review,” *Med. Image Anal.*, vol. 43, pp. 66–84, 2018.
- [116] L. Chan, M. S. Hosseini, C. Rowsell, K. N. Plataniotis, and S. Damaskinos, “Histosegnet: Semantic segmentation of histological tissue type in whole slide images,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 10662–10671.
- [117] G. Corso, L. Cavalleri, D. Beaini, P. Liò, and P. Veličković, “Principal neighbourhood aggregation for graph nets,” in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2020.
- [118] V. P. Dwivedi, C. K. Joshi, T. Laurent, Y. Bengio, and X. Bresson, “Benchmarking graph neural networks,” *arXiv preprint arXiv:2003.00982*, 2020.
- [119] F. Chazal and B. Michel, “An introduction to topological data analysis: fundamental and practical aspects for data scientists,” *arXiv preprint arXiv:1710.04019*, 2017.
- [120] O. Gallego, “Nonsurgical treatment of recurrent glioblastoma,” *Curr. Oncol.*, vol. 22, no. 4, p. e273, 2015.
- [121] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, “Graph networks as a universal machine learning framework for molecules and crystals,” *Chemistry of Materials*, vol. 31, no. 9, pp. 3564–3572, 2019.
- [122] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, “Graphgan: Graph representation learning with generative adversarial nets,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018.
- [123] Y. Li, O. Vinyals, C. Dyer, R. Pascanu, and P. Battaglia, “Learning deep generative models of graphs,” *arXiv preprint arXiv:1803.03324*, 2018.
- [124] J. You, B. Liu, R. Ying, V. Pande, and J. Leskovec, “Graph convolutional policy network for goal-directed molecular graph generation,” in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2018.
- [125] X. Xie, J. Niu, X. Liu, Z. Chen, and S. Tang, “A survey on domain knowledge powered deep learning for medical image analysis,” *arXiv preprint arXiv:2004.12150*, 2020.
- [126] Z.-M. Chen, X.-S. Wei, P. Wang, and Y. Guo, “Multi-label image recognition with graph convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019, pp. 5177–5186.
- [127] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun, “Gram: graph-based attention model for healthcare representation learning,” in *Proc. Int. Conf. Knowledge Discov. Data Mining (KDD)*, 2017, pp. 787–795.
- [128] A. Singh, S. Sengupta, and V. Lakshminarayanan, “Explainable deep learning models in medical image analysis,” *J. Imaging*, vol. 6, no. 6, p. 52, 2020.
- [129] W. Bulten, M. Balkenhol, J.-J. A. Bellinga, A. Brillhante, A. Çakır, L. Egevad, M. Eklund, X. Farré, K. Geronatsiou, V. Molinić *et al.*, “Artificial intelligence assistance significantly improves gleason grading of prostate biopsies by pathologists,” *Modern Pathology*, vol. 34, no. 3, pp. 660–671, 2021.
- [130] J. Chen, T. Ma, and C. Xiao, “Fastgcn: Fast learning with graph convolutional networks via importance sampling,” in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2018.
- [131] Y. You, T. Chen, Z. Wang, and Y. Shen, “L2-gcn: Layer-wise and learned efficient training of graph convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2020, pp. 2127–2135.
- [132] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh, “Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks,” in *Proc. Int. Conf. Knowledge Discov. Data Mining (KDD)*, 2019, pp. 257–266.
- [133] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, “Simplifying graph convolutional networks,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6861–6871.
- [134] P. Zhong, D. Wang, and C. Miao, “Eeg-based emotion recognition using regularized graph neural networks,” *IEEE Trans. Affect. Comput.*, 2020.
- [135] E. Rossi, F. Frasca, B. Chamberlain, D. Eynard, M. Bronstein, and F. Monti, “Sign: Scalable inception graph neural networks,” *arXiv preprint arXiv:2004.11198*, 2020.
- [136] S. A. Tailor, F. L. Opolka, P. Liò, and N. D. Lane, “Adaptive filters and aggregator fusion for efficient graph convolutions,” *arXiv preprint arXiv:2104.01481*, 2021.
- [137] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, “Super-gluon: Learning feature matching with graph neural networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2020, pp. 4938–4947.
- [138] S. Javed, A. Mahmood, M. M. Fraz, N. A. Koohbanani, K. Benes, Y.-W. Tsang, K. Hewitt, D. Epstein, D. Snead, and N. Rajpoot, “Cellular community detection for tissue phenotyping in colorectal cancer histology images,” *Med. Image Anal.*, vol. 63, p. 101696, 2020.
- [139] P. Velickovic, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm, “Deep graph infomax,” in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2019.
- [140] J. N. Kather, A. T. Pearson, N. Halama, D. Jäger, J. Krause, S. H. Loosen, A. Marx, P. Boor, F. Tacke, U. P. Neumann *et al.*, “Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer,” *Nat. Med.*, vol. 25, no. 7, pp. 1054–1056, 2019.
- [141] G. Campanella, M. G. Hanna, L. Geneslaw, A. Mirafior, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images,” *Nat. Med.*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [142] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, “Patch-based convolutional neural network for whole slide tissue image classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2016, pp. 2424–2433.
- [143] M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, “Data-efficient and weakly supervised computational pathology on whole-slide images,” *Nat. Biomed. Eng.*, vol. 5, no. 6, pp. 555–570, 2021.
- [144] M. Shaban, R. Awan, M. M. Fraz, A. Azam, Y.-W. Tsang, D. Snead, and N. M. Rajpoot, “Context-aware convolutional neural network for grading of colorectal cancer histology images,” *IEEE Trans. Med. Imaging*, vol. 39, no. 7, pp. 2395–2405, 2020.
- [145] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, “What clinicians want: contextualizing explainable machine learning for clinical end use,” in *Proc. Mach. Learn. Healthc. Conf.*, 2019, pp. 359–380.
- [146] M. Du, N. Liu, and X. Hu, “Techniques for interpretable machine learning,” *Commun. ACM*, vol. 63, no. 1, pp. 68–77, 2019.
- [147] M. N. Vu and M. T. Thai, “Pgm-explainer: Probabilistic graphical model explanations for graph neural networks,” in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2020.
- [148] M. S. Schlichtkrull, N. De Cao, and I. Titov, “Interpreting graph neural networks for nlp with differentiable edge masking,” in *Proc. Int. Conf. Learn. Repr. (ICLR)*, 2021.
- [149] Q. Huang, M. Yamada, Y. Tian, D. Singh, D. Yin, and Y. Chang, “Graphlime: Local interpretable model explanations for graph neural networks,” *arXiv preprint arXiv:2001.06216*, 2020.
- [150] Y. Zhang, D. Defazio, and A. Ramesh, “Relax: A model-agnostic relational model explainer,” in *Proc. AAAI Conf. AI Ethics Soc.*, 2021, pp. 1042–1049.
- [151] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, “On explainability of graph neural networks via subgraph explorations,” *arXiv preprint arXiv:2102.05152*, 2021.
- [152] W. Guo, S. Huang, Y. Tao, X. Xing, and L. Lin, “Explaining deep learning models—a bayesian non-parametric approach,” in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2018, pp. 4514–4524.
- [153] H. Yuan, J. Tang, X. Hu, and S. Ji, “Xggn: Towards model-level explanations of graph neural networks,” in *Proc. Int. Conf. Knowledge Discov. Data Mining (KDD)*, 2020, pp. 430–438.
- [154] D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, “Parameterized explainer for graph neural network,” in *Proc. Adv Neural Inf. Process. Syst (NeurIPS)*, 2020.