

Hierarchical Graph Representations in Digital Pathology

Pushpak Pati^{*1,2} Guillaume Jaume^{*1,3} Antonio Foncubierta-Rodríguez¹ Florinda Feroce⁴ Anna Maria Anniciello⁴
 Giosue Scognamiglio⁴ Nadia Brancati⁵ Maryse Fiche⁶ Estelle Dubruc⁷ Daniel Riccio⁵ Maurizio Di Bonito⁴
 Giuseppe De Pietro⁵ Gerardo Botti⁴ Jean-Philippe Thiran³ Maria Frucci⁵ Orcun Goksel² Maria Gabrani¹

¹*IBM Research Zurich, Switzerland* ²*ETH Zurich, Switzerland* ³*EPFL, Switzerland* ⁴*IRCCS-Fondazione Pascale, Italy* ⁵*ICAR-CNR, Italy* ⁶*Aurigen, Switzerland* ⁷*CHUV, Switzerland*

Abstract—Cancer diagnosis, prognosis, and therapy response predictions from tissue specimens highly depend on the phenotype and topological distribution of constituting histological entities. Thus, adequate tissue representations for encoding histological entities is imperative for computer aided cancer patient care. To this end, several approaches have leveraged cell-graphs that encode cell morphology and cell organization to denote the tissue information. These allow for utilizing graph theory and machine learning to map tissue representations to tissue functionality to help quantify their relationship. Though cellular information is crucial, it is incomplete alone to comprehensively characterize complex tissue structure. We herein treat the tissue as a hierarchical composition of multiple types of histological entities from fine to coarse level, capturing multivariate tissue information at multiple levels. We propose a novel multi-level hierarchical entity-graph representation of tissue specimens to model hierarchical compositions that encode histological entities as well as their intra- and inter-entity level interactions. Subsequently, a hierarchical graph neural network is proposed to operate on the hierarchical entity-graph representation to map the tissue structure to tissue functionality. Specifically, for input histology images we utilize well-defined cells and tissue regions to build Hierarchical Cell-to-Tissue (HACT) graph representations, and devise HACT-Net, a message passing graph neural network, to classify such HACT representations. As part of this work, we introduce the BReAst Carcinoma Subtyping (BRACS) dataset, a large cohort of Haematoxylin & Eosin stained breast tumor regions-of-interest, to evaluate and benchmark our proposed methodology against pathologists and state-of-the-art computer-aided diagnostic approaches. Through comparative assessment and ablation studies, our proposed method is demonstrated to yield superior classification results compared to alternative methods as well as individual pathologists.

Keywords-Digital pathology, Breast cancer classification, Hierarchical tissue representation, Hierarchical graph neural network, Breast cancer dataset

I. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer and registers the highest number of deaths for women with cancer [1]. A study by [2] exhibits that intensive early diagnostic activities have improved 5-year survival to 85% during 2005–09 for breast cancer patients. Early diagnosis

of cancer, primarily through manual inspection of histology slides, enables the acute assessment of risk and facilitates an optimal treatment plan. Though the diagnostic criteria for breast cancer are established, the continuum of histologic features phenotyped across the diagnostic spectrum prevents the distinct demarcation. Thus, manual inspection is tedious and time-consuming with significant intra- and inter-observer variability [3], [4]. Increasing incidence rate of breast cancer cases per year [5] and the complications in manual diagnosis demand for automated computed-aided diagnostic tools.

Whole-slide scanning systems empowered rapid digitization of pathology slides into high-resolution whole-slide images (WSIs), thereby profoundly transforming pathologists' daily practice [6]. Further, they enabled computed aided diagnostics to leverage artificial intelligence [7], [8], especially deep learning, to address various digital pathology tasks in breast diagnosis [9], such as nuclei segmentation [10], [11], nuclei classification [12], [13], gland segmentation [14], [15], tumor detection [16]–[18], and tumor staging [16], [19]. Deep learning techniques primarily employ Convolutional Neural Networks (CNN) [20], [21] to process pathology images in a patch-wise manner. CNNs extract representative patterns from patches and aggregate patch representations to perform image-level tasks. However, patch-wise processing suffers from the trade-off between the resolution of operation and the acquisition of adequate context [22], [23]. Operating at a higher resolution captures local cellular information but limits the field-of-view due to computational burden, thus limiting access to global information that would capture the immediate tissue microenvironment. In contrast, operating at a lower resolution hinders resolvability of cells and access to cellular features. [22]–[24] have proposed CNN methods to address such trade-off by leveraging visual context, however CNNs, which operate on fix-sized input patches, are confined to a fixed field-of-view and are restricted for diagnosis from incorporating information from varying spatial distances. Further, pixel-based processing in CNNs disregards the notion of

pathologically meaningful entities [25], such as cells, glands, and tissue types. Such absence of a notion for relevant entities severely limits the interpretability of CNNs and any utilization of well-established entity-level prior pathological knowledge in CNN-based digital pathology frameworks. Additionally, CNNs disregard the structural composition of tissue, where fine entities hierarchically constitute to form coarser entities, such as, epithelial cells organize to form epithelium, which further constitutes to form glands. Such hierarchical structure is relevant both for diagnostics and for interpretation.

In this paper, we address the aforementioned limitations by shifting the analytical paradigm from pixel to entity-based processing. In an entity paradigm, a histology image is described as an entity-graph, where nodes and edges of a graph denote biological entities and inter-entity interactions, respectively. An entity-graph can be customized in various aspect, *e.g.*, in terms of the type of entity set, entity attributes, and graph topology, by incorporating any task-specific prior pathological knowledge. Thus, the graph representation enables pathology-specific interpretability and human-machine co-learning. In addition, the graph representation is memory efficient compared to pixelated images and can seamlessly describe a large tissue region. [26] first introduced cell-graphs using cells as the entity type. Though a cell-graph efficiently encodes the cell microenvironment, it cannot extensively capture the tissue microenvironment, *i.e.*, the distribution of tissue regions such as necrosis, stroma, epithelium, etc. Similarly, a tissue-graph comprising of the set of tissue regions cannot depict the cell microenvironment. Therefore, an entity-graph representation using a single type of entity set is insufficient to comprehensively describe the tissue structure. To address the limitation, we propose a multi-level entity-graph representation, *i.e.*, HierArchical Cell-to-Tissue (HACT), consisting of multiple types of entity sets, *i.e.*, cells and tissue regions, to encode both cell and tissue microenvironment. The multiset of entities is inherently coupled depicting tissue composition at multiple scales. The HACT graph encodes individual entity attributes and intra- and inter-entity relationships to hierarchically describe a histology image. Upon the graph construction, a graph neural network (GNN), a deep learning technique operating on graph-structured data, processes the entity-graph to perform image analysis. Specifically, we introduce a hierarchical GNN, HierArchical Cell-to-Tissue Network (HACT-Net), to sequentially operate on HACT graph, from fine-level to coarse-level, to provide a fixed dimensional embedding for the image. The embedding encodes morphological and topological distribution of the multiset of entities in the tissue. Interestingly, our proposed methodology resembles the tissue diagnostic procedure in clinical practice, where a pathologist hierarchically analyzes a tissue.

We propose a methodology that consists of HACT graph construction and HACT-Net based histology image analysis,

We characterize breast tumor regions-of-interest (TRoIs) to evaluate our methodology. A preliminary version of this work was presented as [27]. Our substantial extensions herein include 1) an improved HACT representation and HACT-Net architecture, 2) a larger evaluation dataset (twice the earlier size), 3) detailed ablation studies and evaluation on public data, and 4) a benchmark comparison against independent pathologists. Specifically, the major contributions of this paper are:

- A novel hierarchical entity-graph representation (HACT) and hierarchical learning (HACT-Net) methodology for analyzing histology images;
- Introducing a public dataset, BReAst Carcinoma Subtyping (BRACS¹), a large cohort of breast TRoIs annotated with seven breast cancer subtypes. BRACS includes challenging atypical cases and a wide variety of TRoIs representing a more realistic breast cancer analysis;
- An evaluation of our proposed methodology on the BRACS dataset by comparing with three independent pathologists, where an extensive assessment demonstrates our classification performance outperforming several recent CNN and GNN approaches for cancer subtyping, while being comparable to pathologists on per-class and aggregated classification tasks.

II. RELATED WORK

Tumor subtyping in digital pathology: Several deep learning algorithms have been proposed to categorize pathology images or WSIs into cancer subtypes [8], [16], [28]–[31]. For this task, most algorithms in the literature employ CNNs in a patch-wise manner: In [19], [31]–[33] CNNs are utilized to categorize breast cancer pathology images into various subtypes. These methodologies employ single stream patch-wise approaches to capture local patch-level context, unify patch-level information via several aggregation strategies, and classify the image using aggregated information. However, single-stream approaches do not capture adequate context from neighboring tissue microenvironment to aptly classify a patch. In [23] this issue is addressed by including multi-scale information from concentric patches across different magnifications. [24] propose a neural image compression methodology, where WSIs are compressed using a neural network trained in an unsupervised fashion, followed by a CNN trained on the compressed image representations to predict image labels. [34] include an attention module with an auxiliary task to improve neural image compression for classifying histology images. [35] propose a hybrid convolutional and recurrent deep neural network to utilize spatial correlations among patches for pathological image classification. [22] propose a stacked CNN architecture to capture large contexts and perform end-to-end processing

¹The BRACS dataset for breast cancer subtyping will be shortly released

of large-sized histology images. [36] propose a streaming CNN to accommodate multi-megapixel images. [37] utilize a multiple instance learning approach to process whole-slide images in an end-to-end manner, which is extended by [38] to automatically identify sub-regions of high diagnostic value through an attention mechanism. Whereas the aforementioned methodologies employ different strategies to capture useful context information, they all still operate on a square and fix-sized input image. However, actual TRoIs can be of highly varying dimensions and shapes depending on cancer subtype and the site of tissue extraction. Our proposed entity-graph methodology can acquire both local and global context from arbitrary-sized TRoIs to address the aforementioned limitations.

Graphs in digital pathology: Entity-graph based tissue representations can effectively describe the phenotypical and structural properties of tissues by incorporating morphology, topology, and interactions among biologically relevant tissue entities. Using cells as entities, [26] introduced a cell-graph (CG) representation of tissues, where cell morphology can be embedded in the nodes via hand-crafted [26], [27], [39] or deep-learning based features [40]. The graph topology is often heuristically defined, *e.g.*, using k-Nearest Neighbors, probabilistic modeling, or a Waxman model [41]. Subsequently, a CG is typically processed by classical machine learning techniques [42], [43] or GNNs [27], [39], [40], [44] for mapping to tissue function. Recently, improved graph representations using patches [45] and tissue regions [27] have been proposed to enhance the tissue structure-function mapping. Other graph-based applications in computational pathology include cellular community detection [46] and WSI classification [47], [48]. Notably, entity-graph representations consist of biological entities which the pathologists can readily relate to. Thus, graph-based analysis allows for incorporating pathologically-defined, task-specific, entity-level prior knowledge in constructing “meaningful” tissue representations. This implicitly enables and facilitates the study of *interpretability* and *explainability* of such graph-based networks by pathologists. To this end, [39] analyze the clustering of nodes in a CG in order to group cells according to their appearance and tissue types. [49] introduce a post-hoc graph-pruning explainer to identify decisive cells and interactions. [50] employ a robust spatial filtering that utilizes an attention-based GNN and node occlusion to highlight cell contributions. [51] propose quantitative metrics based on pathologically measurable cellular properties in order to characterize several graph explainability algorithms in CG analysis.

III. PRELIMINARIES

A. Notation

We define an attributed, undirected entity-graph $G := (V, E, H)$ as a set of nodes V , edges E , and node features H . Each node $v \in V$ is represented by a feature vector

$h(v) \in \mathbb{R}^d$, thus, $H \in \mathbb{R}^{|V| \times d}$. d denotes the number of features per node, and $|\cdot|$ denotes set cardinality. An edge between two nodes $u, v \in V$ is denoted as e_{uv} . The graph topology is described by a symmetric adjacency matrix $A \in \mathbb{R}^{|V| \times |V|}$, where $A_{u,v} = 1$ if $e_{uv} \in E$. The neighborhood of a node $v \in V$ is denoted as $\mathcal{N}(v) := \{u \in V \mid v \in V, e_{uv} \in E\}$.

B. Graph Neural Networks

GNN [52]–[56] defines a class of neural networks that extend the concept of convolution to operate on graph-structured data. In this work, we employ a GNN based on message-passing. In message-passing GNNs [57], node features $h(v), \forall v \in V$ are iteratively updated in a two-step procedure, i) *AGGREGATE*, and ii) *UPDATE*. In the *AGGREGATE* step for node v , the features of neighboring nodes $\mathcal{N}(v)$ are aggregated into a single feature representation. In the *UPDATE* step, the features of node v is updated by using the current node features and the aggregated representation from the *AGGREGATE* step. A series of T such iterations, in the form of GNN layers, are employed to obtain updated node features $\forall v \in V$, incorporating information within upto T -hops from each node. Finally, the node features $h^T(v)$ are pooled in the *READOUT* step to build a fix-sized graph-level embedding h_G . *AGGREGATE*, *UPDATE*, and *READOUT* operations must be differentiable to allow back-propagation in GNN training. Additionally, *AGGREGATE* and *READOUT* operations must be permutation-invariant such that the aggregated representation is invariant to node ordering. Formally, the three steps are presented as

$$\begin{aligned} a^{t+1}(v) &= \text{AGGREGATE}(\{h^t(u) : u \in \mathcal{N}(v)\}) \\ h^{t+1}(v) &= \text{UPDATE}(h^t(v), a^{t+1}(v)) \\ h_G &= \text{READOUT}(\{h^T(v) : v \in V\}) \end{aligned} \quad (1)$$

An important aspect of designing a GNN is the characterization of its expressive power. A GNN has a strong expressive power if it can map two non-isomorphic graphs to two unique graph embeddings, thus imparting an injective mapping between the graph and embedding spaces. A line-of-research exploring the expressive power of GNNs [54], [58], [59] highlight the connection between iterative message passing procedure of GNN and the popular Weisfeiler-Lehman (WL) [60] test for graph isomorphism. It is established that architectures such as the Graph Isomorphism Network (GIN) [54] can perform as well as the 1-dimensional WL test for *countable* node feature spaces, *i.e.*, when the node features are discrete. An example of graph with discrete node features can be the study of molecule design, where the nodes represent atoms that are discrete in nature. Recent studies show that in case of *continuous* node features, *e.g.*, CNN-based node embeddings, the use of multiple permutation-invariant aggregators, such as sum, max, and mean, can help in building expressive GNNs [61], [62].

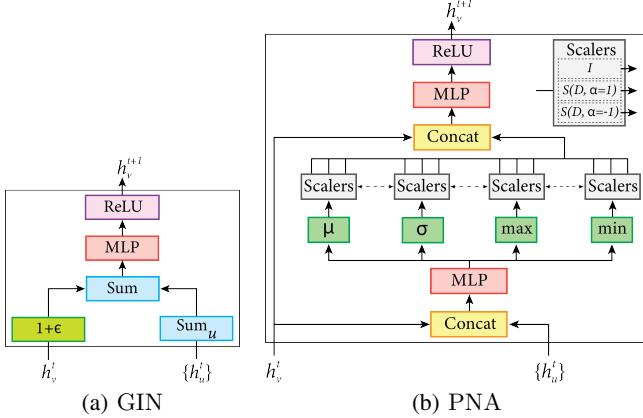


Figure 1: Overview of GIN and PNA layers, where h_v^t and $\{h_u^t\}$ denote the representation of, respectively, node v and its neighbors at layer t . GIN uses sum as the *AGGREGATE* function, followed by a sum and multi-layer perceptron (MLP) for the *UPDATE* function. PNA uses a set of aggregators (element-wise mean, standard deviation, maximum, and minimum) followed by degree-scalers (identity, amplifier, and dampener) as the *AGGREGATE* function. The *UPDATE* function consists of a concatenation followed by an MLP.

To this end, [62] proposed the Principal Neighbourhood Aggregation (PNA) network by using a combination of aggregators followed by *degree-scalers*. The series of aggregators replace the sum operation in GIN while the degree-scalers can amplify or dampen neighboring aggregated-messages based to the degree of a node. Illustrations of GIN and PNA architectures are shown in Figure 1.

IV. METHODOLOGY

In this section, we detail our proposed methodology for hierarchical tissue analysis, as illustrated in Figure 2. For an input Hematoxylin and Eosin (H&E) stained histology TRoI image, first, we apply some pre-processing to standardize the input. Then, we identify pathologically relevant entities and construct a HACT graph representation of the TRoI by encoding the morphological and topological information of such entities. Finally, HACT-Net, a hierarchical GNN, is employed to map the HACT graph to a corresponding tissue class, *e.g.*, a cancer subtype.

A. Pre-processing

H&E stained tissue specimens exhibit appearance variability due to various reasons, such as different specimen preparation techniques, staining protocols (*e.g.*, temperature of the solutions used), fixation characteristics, and imaging device characteristics. Such variability in appearance adversely impacts any computational methods for downstream diagnosis [63], [64]. To alleviate the appearance variability,

we employ the unsupervised, reference-free stain normalization algorithm proposed by [65]. The algorithm is based on the principle that RGB color of each pixel is a linear combination of two unknown stain vectors, Hematoxylin and Eosin, that need to be estimated. First, the algorithm estimates the stain vectors of a TRoI by using a Singular Value Decomposition of the non-background pixels. Second, the algorithm applies a correction to account for the intensity variations due to noise. The algorithm does not involve any intermediate step that may require model parameter tuning or training. Specifically, for stain normalization we employ the scalable and fast pipeline proposed by [66].

B. Graph representation

Stain normalized TRoIs are next processed to identify entities relevant to constructing a hierarchical entity-graph representation. Given a histological task, different choices of entities in histology images can be selected as the relevant biological structures. In this work, we consider nuclei and tissue regions as the relevant entities, such that our HACT graph representations consist of three components: 1) a low-level *cell-graph*, capturing cell morphology and interactions, 2) a high-level *tissue-graph*, capturing morphology and spatial distribution of tissue regions, and 3) cells-to-tissue hierarchies, encoding, *e.g.*, the relative spatial distribution of cells with respect to the tissue distribution. The details of the components are presented in the following subsections.

1) *Cell-graph representation*: A cell-graph (CG) characterizes low-level cell information, where the nodes represent cells encoding cell morphology, and the edges encode cellular interactions depicting cell topology. A CG is constructed in three steps, i) nuclei detection, ii) nuclei feature extraction, and iii) topology configuration. The steps are demonstrated in Figure 3.

Precise nuclei detection facilitates reliable CG representation. To this end, we use HoVer-Net, a nuclei segmentation network proposed by [11], pre-trained on MoNuSeg dataset by [10]. HoVer-Net leverages the instance-rich information encoded within the vertical and horizontal distances of nuclear pixels to their centers of mass. These distances are used to separate clustered nuclei, resulting in an accurate segmentation, particularly in areas with overlapping instances. The centroids of the segmented instances are used as the spatial coordinates of nodes in CG.

Following nuclei detection, morphological features are extracted by processing patches of size $h \times w$ centered around nuclei centroids via ResNet [67] architecture pre-trained on ImageNet dataset [68]. Spatial features of the nuclei are extracted as the spatial coordinates of the nuclei, normalized by the TRoI dimensions. Morphological and spatial features together constitute the nuclei features, which are colocated for all nodes as the node-feature matrix $H_{CG} \in \mathbb{R}^{|V_{CG}| \times d_{CG}}$.

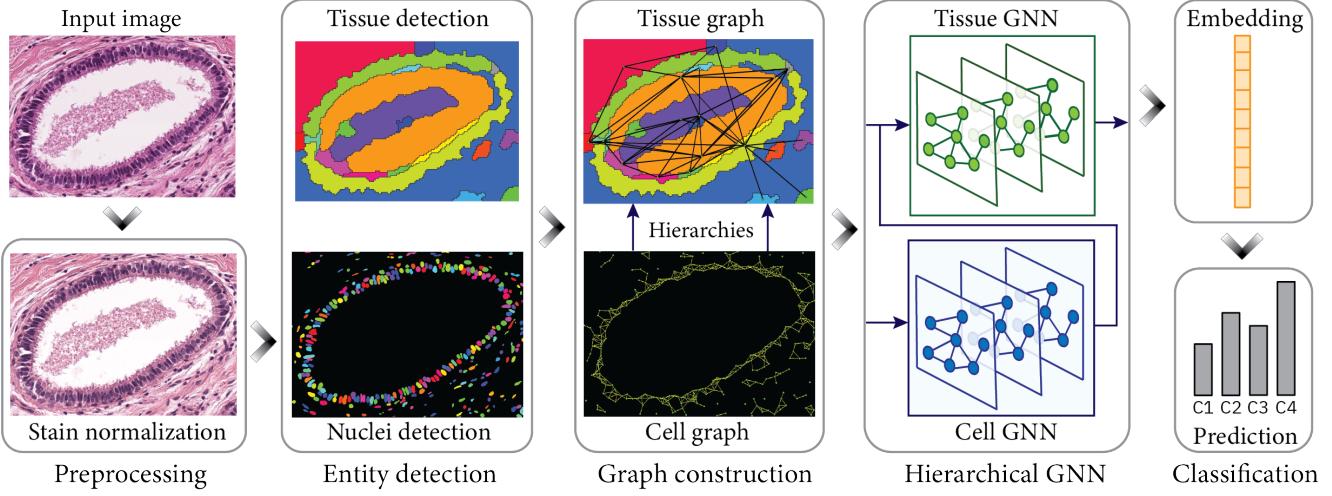


Figure 2: Overview of the proposed hierarchical entity-graph based tissue analysis methodology. Following some preprocessing, a hierarchical entity-graph representation of a tissue is constructed, and it is processed via a hierarchical graph neural network to learn the mapping from tissue compositions to respective tissue categories. (Figure is best viewed in color)

For the CG topology E_{CG} , we utilize the fact that spatially close cells have stronger interactions with distant cells having weaker cellular interactions [69]. Accordingly, we connect nearby cells with edges to model their interactions. To this end, we use the k-Nearest Neighbors (kNN) algorithm to build an initial topology, that we subsequently prune by removing edges longer than a threshold distance d_{\min} . We use Euclidean distances between nuclei centroids in the image space to quantify cellular distances. Formally, for each node v , an edge e_{vu} is built if

$$u \in \{w \mid \text{dist}(v, w) \leq d_k \wedge \text{dist}(v, w) < d_{\min}, \forall w, v \in V_{CG}\}, \\ , d_k = k^{\text{th}} \text{ smallest distance in } \text{dist}(v, w)\} \quad (2)$$

CG topology is represented by a binary adjacency matrix $E_{CG} \in \mathbb{R}^{|V_{CG}| \times |V_{CG}|}$. Figure 3 illustrates the CG representation for a sample TRoI. Formally, a CG representation is formulated as $G_{CG} := \{V_{CG}, E_{CG}, H_{CG}\}$.

2) *Tissue-graph representation:* A tissue graph (TG) depicts a high-level tissue microenvironment, where the nodes and edges of a TG denote tissue regions and their interactions, respectively. Similarly to a CG, a TG is constructed by first identifying tissue regions (*e.g.*, epithelium, stroma, lumen, necrosis, and others in this paper), followed by feature representation of such tissue regions, and finally the TG topology construction. The steps are illustrated in Figure 3.

Tissue regions are identified in a two-step process. First, we oversegment the tissue at a low magnification to detect non-overlapping homogeneous superpixels. Operating at low magnification avoids noisy pixels and provides computational efficiency. To this end, we employ the Simple Linear Iterative Clustering (SLIC) superpixel algorithm [70].

Formally, SLIC follows an unsupervised approach by associating each pixel with a feature vector and merging the pixels using a localized version of k-means clustering. Next, we iteratively merge neighboring superpixels that have similar color attributes, *i.e.*, channel-wise mean and standard deviation, to create superpixels that capture meaningful tissue information. A sample tissue-region instance-map is presented in Figure 3.

To extract feature representations of tissue regions, we follow a two-step procedure: First, we extract CNN-based features for oversegmented superpixels. Patches of size $h \times w$ centered around oversegmented superpixel centroids are processed by ResNet. Second, morphological features of a tissue region are obtained by averaging the deep features of its constituting superpixels. Similar to CG, we include spatial features as the normalized centroids of the tissue region. For a TRoI with a set of V_{TG} tissue regions, we denote the TG node-feature matrix as $H_{TG} \in \mathbb{R}^{|V_{TG}| \times d_{TG}}$.

We assume adjacent tissue regions to biologically interact the most and thus be connected in the TG topology. To this end, we construct a Region Adjacency Graph [71] where an edge is built between each adjacent tissue region. The TG topology is denoted by a binary adjacency matrix $E_{TG} \in \mathbb{R}^{|V_{TG}| \times |V_{TG}|}$. Formally, a TG representation is formulated as $G_{TG} := \{V_{TG}, E_{TG}, H_{TG}\}$.

3) *Hierarchical Cell-to-Tissue graph representation:* Tissues in histopathology can be considered as hierarchical organizations of biological entities ranging from fine-level, *i.e.*, cells, to coarse-level, *i.e.*, tissue regions. There exist intra- and inter-level coupling based on topological distributions and interactions among the entities. Following this motivation, we propose HACT, a HierArchical Cell-to-Tissue (HACT) graph representation to jointly represent

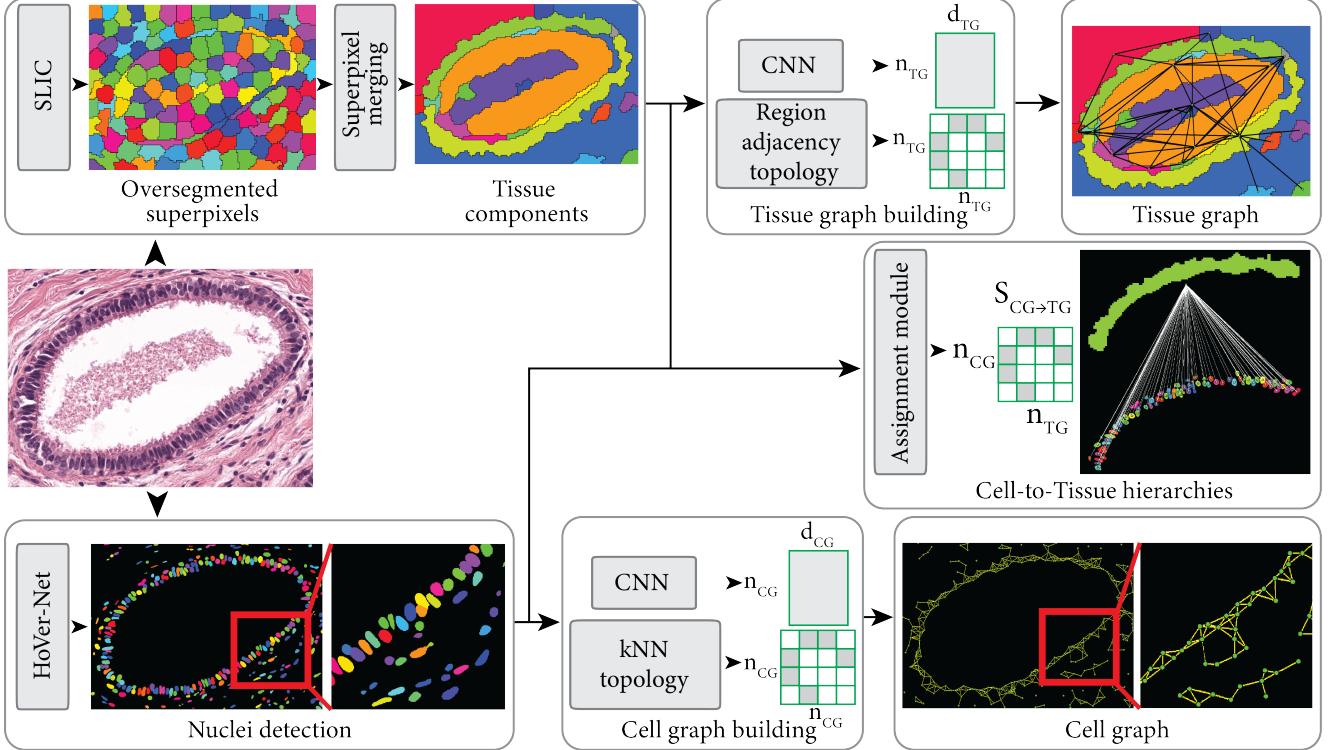


Figure 3: Overview of hierarchical cell-to-tissue (HACT) graph construction for a TRoI. Our HACT graph representation consists of a cell-graph, a tissue-graph, and cell-to-tissue hierarchies, while encoding the phenotypical and topological distributions of tissue entities to describe the cell and tissue microenvironments. (Figure is best viewed in color)

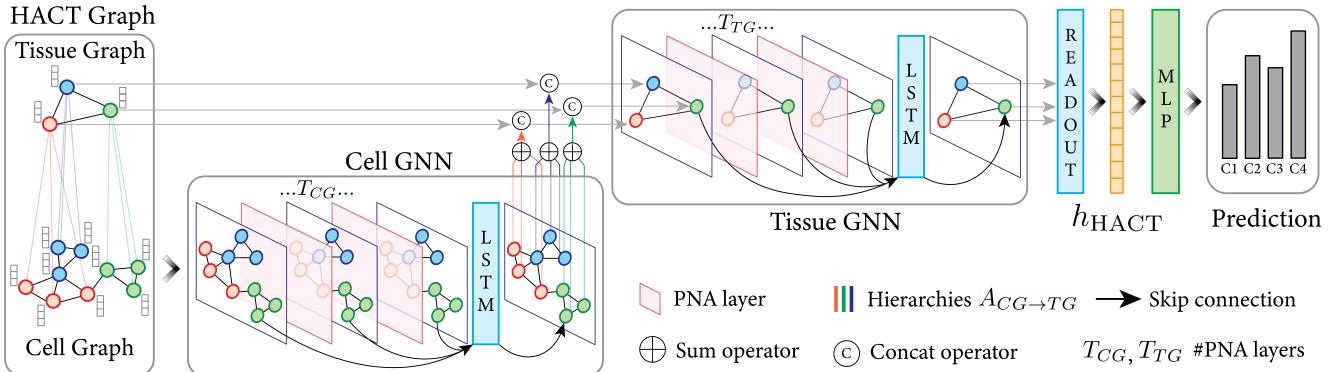


Figure 4: Overview of the proposed HACT-Net architecture. The network processes an input HACT graph representation in a hierarchical manner, from finer cell-level to coarser tissue-region level, to obtain a contextualized graph embedding, which is subsequently used to classify the input graph. (Figure is best viewed in color)

low-level CG and high-level TG. Intra-level topology is already captured by CG and TG standalone. Inter-level topology is presented by a binary assignment (cell-to-tissue hierarchy) matrix $A_{CG \rightarrow TG} \in \mathbb{R}^{|V_{CG}| \times |V_{TG}|}$ that utilizes the relative spatial distributions of nuclei with respect to tissue regions. For the i^{th} nucleus and j^{th} tissue region, the

corresponding assignment is given as

$$A_{CG \rightarrow TG}[i, j] = 1, \text{ if } i^{\text{th}} \text{ nucleus centroid} \in j^{\text{th}} \text{ tissue region} \\ A_{CG \rightarrow TG}[i, j] = 0, \text{ otherwise} \quad (3)$$

Cell-to-tissue hierarchies for a tissue region are presented in Figure 3. Each nucleus is assigned to one and only one tissue region. If a segmented nucleus is at the border of multiple tissue regions, the nucleus is assigned to the tissue

region that it has the maximum overlap with. Formally for a given TRoI, a HACT representation is formulated as $G_{\text{HACT}} := \{G_{\text{CG}}, G_{\text{TG}}, A_{\text{CG} \rightarrow \text{TG}}\}$.

C. Graph learning

The HACT graph representation for a TRoI is processed by a hierarchical GNN to map the TRoI composition to respective TRoI subtype. To this end, we propose a Hierarchical Cell-to-Tissue Network (HACT-Net), a hierarchical GNN architecture illustrated in Figure 4.

1) HACT-Net architecture & learning: HACT-Net takes a HACT representation G_{HACT} as input and outputs a graph-level representation $h_{\text{HACT}} \in \mathbb{R}^{d_{\text{HACT}}}$. Subsequently, a multi-layer perceptron (MLP) classifies h_{HACT} into a respective tissue class, *e.g.*, cancer subtype. Formally, HACT-Net consists of two GNNs, namely Cell-GNN (CG-GNN) and Tissue-GNN (TG-GNN), to hierarchically process the HACT graph from fine to coarse level. In this work, we leverage the recent advances in GNNs and model HACT-Net using Principal Neighbourhood Aggregation (PNA) layers [62].

First, CG-GNN takes $G_{\text{CG}} := \{V_{\text{CG}}, E_{\text{CG}}, H_{\text{CG}}\}$, and applies T_{CG} PNA layers to build contextualized cell-node embeddings. Inline with Equation (1), we iteratively update each node embedding $h_{\text{CG}}^{(t)}(v)$, $\forall v \in V_{\text{CG}}$ as

$$\begin{aligned} a_{\text{CG}}^{(t+1)}(v) &= \bigoplus_{u \in \mathcal{N}_{\text{CG}}(v)} M_{\text{CG}}^{(t)} \left(h_{\text{CG}}^{(t)}(v), h_{\text{CG}}^{(t)}(u) \right) \\ h_{\text{CG}}^{(t+1)}(v) &= U_{\text{CG}}^{(t)} \left(h_{\text{CG}}^{(t)}(v), a_{\text{CG}}^{(t+1)}(v) \right) \end{aligned} \quad (4)$$

where $t = 0, \dots, T_{\text{CG}}$ is the iteration index, $\mathcal{N}_{\text{CG}}(v)$ is the set of cell-nodes neighboring v , and the functions U_{CG}^t and M_{CG}^t are MLPs. \bigoplus denotes the combination of multiple degree-scalers and aggregators, *i.e.*,

$$\begin{aligned} \bigoplus &= \left[I, \mathcal{S}(D, \alpha = 1), \mathcal{S}(D, \alpha = -1) \right] \otimes [\mu, \sigma, \max, \min] \\ \mathcal{S}(D, \alpha) &= \frac{\log(D + 1)^\alpha}{\delta} \\ \delta &= \frac{1}{|\text{train}|} \sum_{i \in \text{train}} \log(d_i + 1) \end{aligned} \quad (5)$$

where I is the identity matrix, \mathcal{S} is the degree-scaler matrix, D is the degree matrix of cell-nodes, $[\mu, \sigma, \max, \min]$ is the list of aggregators that compute statistics on neighboring nodes, \otimes is tensor product, δ is a normalization constant computed as the average log-scale cell-node degree from the training dataset, and α is a variable (that is negative for attenuation, positive for amplification, or zero for no scaling). The schematic diagram of a PNA layer is shown in Figure 1. After T_{CG} PNA layers, an LSTM-based jumping

knowledge technique [72] is employed to adapt to different CG sub-graph structures, *i.e.*,

$$h_{\text{CG}}^{(T_{\text{CG}}+1)}(v) = \text{LSTM} \left(\left\{ h_{\text{CG}}^{(t)}(v) \mid t = 1, \dots, T_{\text{CG}} \right\} \right) \quad (6)$$

The cell node embeddings $h_{\text{CG}}^{T_{\text{CG}}+1}(v) \mid v \in V_{\text{CG}}$ and the assignment matrix $A_{\text{CG} \rightarrow \text{TG}}$ are used as additional hierarchical information to initialize the tissue-node features in the TG, *i.e.*,

$$h_{\text{TG}}^{(0)}(w) = \text{CONCAT} \left(H_{\text{TG}}(w), \sum_{v \in \mathcal{M}(w)} h_{\text{CG}}^{(T_{\text{CG}}+1)}(v) \right) \quad (7)$$

where CONCAT denotes the concatenation operation and $\mathcal{M}(w) := \{v \in V_{\text{CG}} \mid A_{\text{CG} \rightarrow \text{TG}}(v, w) = 1\}$ is the set of nodes in G_{CG} mapping to a node $w \in V_{\text{TG}}$. Analogously to Equation (4), we process G_{TG} using the TG-GNN, based on PNA layers, to compute tissue region node embeddings $h_{\text{TG}}^{(t)}(w)$, $\forall w \in V_{\text{TG}}$. At $t = T_{\text{TG}}$, the embedding of each tissue component node w encodes the cell and tissue information up to T_{TG} -hops from w .

Similar to CG, the tissue-node embeddings in TG are processed with an LSTM-based jumping knowledge technique to aggregate the intermediate tissue-node representations. Finally, the graph-level embedding h_{HACT} is extracted by summing all tissue-node representations. An MLP layer followed by softmax operation provides a mapping from h_{HACT} to the respective TRoI label. The model is trained end-to-end to minimize the cross-entropy loss between the softmax output and the ground-truth TRoI label.

Following [73], after each PNA layer we include graph normalization (GraphNorm) followed by a batch normalization (BatchNorm). Graph normalization scales the node features by the number of nodes in the graph. Intuitively, it prevents the node representations from being at different scales, for graphs of different sizes. This normalization helps the network to learn discriminative topological patterns when the number of nodes vary significantly within a class.

V. DATASETS

BRACS dataset: As part of this work, we introduce a new dataset termed as the BReAst Cancer Subtyping (**BRACS**) dataset. BRACS contains 4391 TRoIs acquired from 325 H&E stained breast carcinoma WSIs. The WSIs were selected from the archives of the Department of Pathology at National Cancer Institute- IRCCS-Fondazione Pascale, Naples, Italy, and are scanned with an Aperio AT2 scanner at 0.25 $\mu\text{m}/\text{pixel}$ resolution. The TRoIs were selected and annotated using QuPath software [74] as being Normal, Benign, Usual ductal hyperplasia (UDH), Atypical Ductal Hyperplasia (ADH), Flat Epithelial Atypia (FEA), Ductal Carcinoma In Situ (DCIS), and Invasive. Figure 5 presents sample TRoIs from all the cancer subtypes in BRACS.

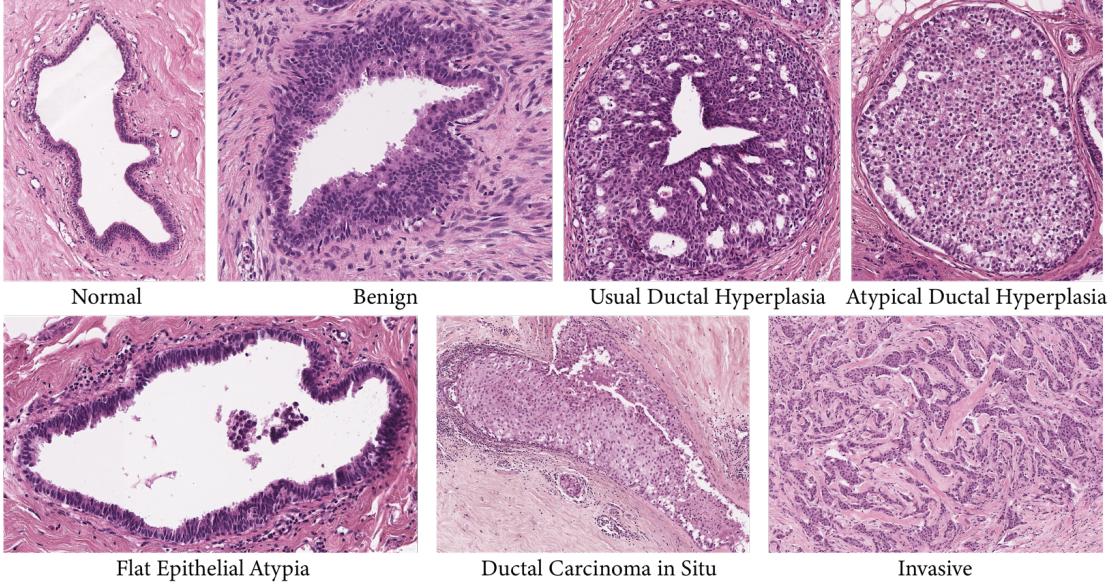


Figure 5: Samples of class-wise tumor regions-of-interest in BRACS dataset. (Figure is best viewed in color)

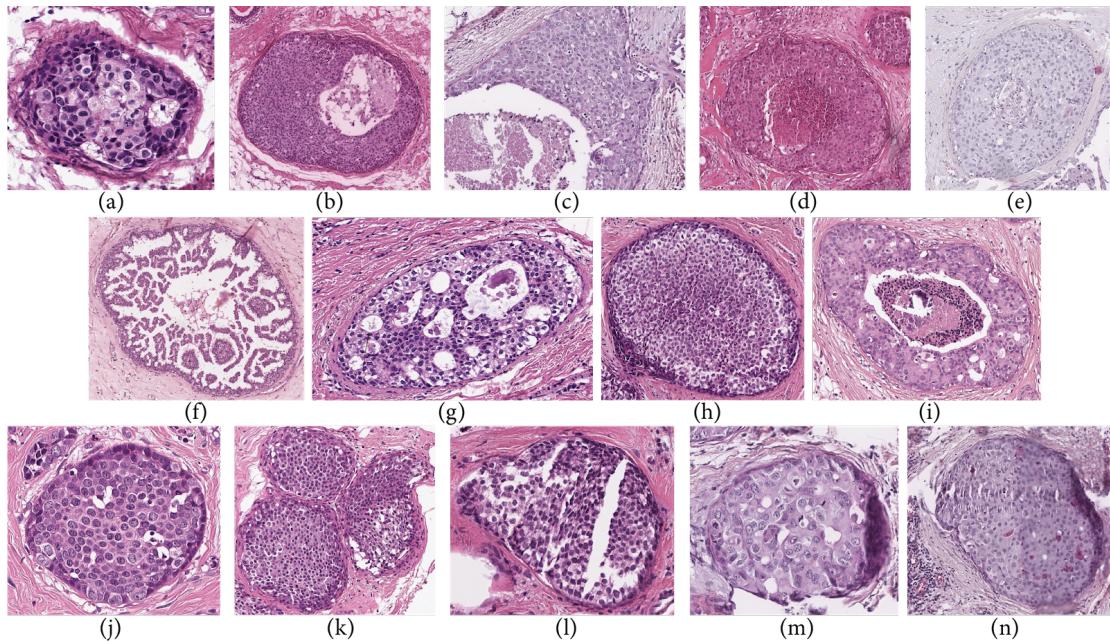


Figure 6: Overview of the variability in Ductal Carcinoma *in situ* tumor regions-of-interest in BRACS dataset. (Figure is best viewed in color)

Each TRoI was first annotated independently by three pathologists, and then any TRoIs with any conflict in annotation were further discussed and annotated based on consensus of the three pathologists. Note that the pathologists utilize the entire WSI-level context during the TRoI annotation. Figure 6 demonstrates a few DCIS samples in BRACS to depict the appearance variability per category. Figure 6(a,b,c) display the variability in the TRoI sizes.

Figure 6(d,e) display the variability in staining appearance. Figure 6(f,g,h,i) display different patterns of low-grade (Papillary), moderate-grade (Cribriform) and high-grade (Solid and Comedo) DCIS in the dataset. Figure 6(j,k) present DCIS TRoIs with single and multiple glandular regions. Figure 6(l,m,n) presents some notable artifacts in tissue and slide preparation, such as tissue-folding, tears, ink stains, and blur. Similar TRoI variability is typical in practice and

Table I: Key statistics of the BRACS dataset.

	Metric	Normal	Benign	UDH	ADH	FEA	DCIS	Invasive	Total
Image	Number of images	512	758	471	568	783	749	550	4391
	Number of pixels (in million)	2.8±2.7	5.7±4.5	2.4±2.9	2.2±2.0	1.2±1.1	5.0±5.0	8.2±5.4	3.9±4.3
	Max/Min pixel ratio	75.3	97.9	180.1	75.3	58.3	128.6	62.4	235.6
CG	Number of nodes	994±732	1826±1547	903±910	863±730	470±352	1723±1598	3609±2393	1468±1642
	Number of edges	3759±2643	6103±5420	3371±3675	3098±2781	1738±1395	5728±5811	12490±10011	5102±6089
	Max/Min node ratio	71.9	126.6	133.3	104.2	45.2	161.3	113.6	256.4
TG	Number of nodes	107±106	217±233	88±93	100±91	45±32	225±217	423±317	172±217
	Number of edges	509±545	1012±1236	393±450	480±474	194±159	1111±1123	2025±1741	815±1125
	Max/Min node ratio	169.5	312.5	125.0	178.6	416.7	312.5	101.0	434.8
Image split	Train	342	586	303	405	599	562	366	3163
	Validation	86	87	88	77	85	97	82	602
	Test	84	85	80	86	99	90	102	626
WSI split	Train	67	86	59	38	37	33	41	198
	Validation	28	24	24	28	17	21	19	68
	Test	15	16	20	17	12	16	16	59

it exists in our dataset for other cancer subtypes as well. Including such natural variability in the dataset (instead of “cleaning” them, *i.e.*, filtering them out) ensures for us a realistic and representative evaluation set, with results readily applicable in the field.

Table I presents category-wise statistics of the TRoIs in BRACS. The statistics demonstrate a high variation in TRoI dimensions. Additionally, we also present the statistics for the CG and TG representations constructed by our framework as described earlier above. These indicate a large variation in the size of the constructed graph representations. For evaluations on BRACS, we partition the TRoIs into train, validation, and test sets at the WSI-level, such that two TRoIs from the same WSI do not fall in different sets. The WSI-level splitting was performed randomly, ensuring a comparable number of TRoIs per cancer subtype. Such partitioning aimed for a fair evaluation of the compared methods.

BACH dataset: We evaluated the proposed methodology also on the publicly available microscopy image dataset called the Grand Challenge on BreAst Cancer Histology images **BACH** [16]. This dataset consists of 400 training and 100 test images from four breast cancer subtypes, *i.e.*, Normal, Benign, DCIS, and Invasive. All images are acquired using a Leica DM 2000 LED microscope and a Leica ICC50 HD camera. These images are in RGB TIFF format and have a fixed size of 2048×1536 pixels and a pixel scale of $0.42 \times 0.42 \mu\text{m}$. Notably, our BRACS dataset has three major advantages over the BACH dataset:

- Number of images: The train and test sets of BRACS are nearly 10 times and 6 times the size of the train and test sets of BACH, respectively. The large test set ensures a robust evaluation of the methods.
- Diverse subtypes: BRACS includes diagnostically com-

plex pre-cancerous atypical (ADH and FEA) categories, which represent a major diagnostic dilemma typical in practice, due to their high risk of progressing to cancer. The seven cancer subtypes in BRACS represent a broad spectrum of breast cancer in histopathology.

- Large variability: The aforementioned high variability in the BRACS dataset in terms of TRoI appearances and dimensions is clinically more representative and corresponds to a more realistic scenario of breast cancer subtyping.

VI. RESULTS

In this section, we comparatively evaluate the proposed methods for breast cancer subtyping. First, we introduce state-of-the-art CNN and GNN baselines and their implementation strategies. Second, we conduct ablation studies on BRACS dataset to examine the impact of various components of our proposed framework. Third, we evaluate the classification performance of our method, comparatively with the given baselines, on BRACS and BACH datasets for different classification settings. Finally, we include a comparison with gold-standard, by analyzing the performance of HACT-Net against three independent expert-pathologists.

A. CNN and GNN baselines for comparative evaluation

- **Single-scale CNN** processes TRoIs at a single magnification. A CNN is trained to predict patch-wise cancer subtypes and we aggregate the patch-wise predictions to produce a TRoI-level prediction. We experiment with images at three magnifications, namely $10\times$, $20\times$, and $40\times$, denoted herein as CNN($10\times$), CNN($20\times$), and CNN($40\times$). Same network architecture and training strategy is employed for all scales. For each scale, we extract patches of size

128×128 pixels with a stride of 64 pixels. The CNN follows the single-scale training procedure by [23] and patch-wise predictions are aggregated using the Agg-Penultimate strategy proposed by [19]. We use transfer learning with a ResNet-50 architecture, pre-trained on ImageNet dataset, as our CNN backbone. Following feature extraction by ResNet-50, we employ a two-layer MLP of 128 channels to classify the patches. To improve classification performance, we used fine-tuning from the original ResNet-50 parameters. Adam optimizer [75] with a learning rate of 10^{-3} , a batch size of 16, and a dropout ratio of 0.2 were used to optimize the categorical cross-entropy objective.

- **Multi-scale CNN** processes the TRoIs at multiple scales. We extract concentric patches of size 128×128 pixels from multiple magnifications and follow the “Late fusion with single-stream + LSTM” training procedure from [23]. This multi-scale approach uses concentric patches to acquire context information from multiple magnifications for improving the patch classification. We operate at two settings, *i.e.*, $(10 \times + 20 \times)$ and $(10 \times + 20 \times + 40 \times)$, and denoted by prepending Multi-scale CNN in front of each. The patch-wise predictions are aggregated using the Agg-Penultimate strategy by [19]. On the concatenated feature representation from the LSTM, we employ a two-layer MLP of 128 channels to classify the patches. The training strategy and hyperparameters are the same as Single-scale CNN.

- **CGC-Net** is the Cell Graph Convolutional Network (CGC-Net) proposed by [39], and it is the state-of-the-art in classifying CG representations for TRoIs. We construct the CG topology for a TRoI using thresholded kNN strategy presented in Section IV-B1. We initialize the CG nodes with hand-crafted features, employ the Adaptive GraphSage-based CGC-Net architecture, and follow the training strategy proposed by [39].

- **Patch-GNN** implements the methodology proposed by [45], which is the state-of-the-art GNN method for classifying patch-graph representations of TRoIs. This methodology incorporates local inter-patch context through a GNN to construct a graph-level feature representation, which is then processed by an MLP to classify the TRoIs. We experiment with Patch-GNN at three scales, *i.e.*, $10 \times$, $20 \times$, and $40 \times$, denoted herein as Patch-GNN($10 \times$), Patch-GNN($20 \times$), and Patch-GNN($40 \times$). At each magnification, we extract patches of size 128×128 to construct a TRoI-specific patch-graph. We employ the network architecture and training strategy proposed by [45].

- **CG-GNN** is provided as a standalone cell-graph based learning baseline, to compare with our proposed hierarchical learning. CG-GNN architecture utilizes PNA layers, an LSTM-based jumping knowledge, sum readout, and a two-layer MLP classifier. We follow the same CG representation strategy as described above in Section IV-B1.

- **TG-GNN** is provided as a standalone tissue graph-based learning baseline, to compare with our proposed hierarchical

learning. TG-GNN employs the same architecture as the CG-GNN, with the node features directly initialized by H_{TG} instead of Equation (7).

- **CONCAT-GNN** is provided to evaluate the impact of hierarchical graph representation and learning. CONCAT-GNN utilizes standalone CG and TG representations, respectively, as input to standalone CG-GNN and TG-GNN to produce h_{CG} and h_{TG} graph-level embeddings. The TRoI level embedding is constructed by concatenating the graph-level embeddings, *i.e.*, $h_{CONCAT} = \text{CONCAT}(h_{CG}, h_{TG})$. Finally, a two-layer MLP classifies h_{CONCAT} into a cancer subtype.

B. Implementation

Graph representations: CG representations (Section IV-B1) use, i) patches of size 72×72 , and ii) a CNN of ResNet-34 or ResNet-50 to initialize the node features. TG representations (Section IV-B2) use, i) patches of size 144×144 , and ii) a CNN of ResNet-34 or ResNet-50 to initialize the node features.

Graph architecture and learning: CG-GNN, TG-GNN, CONCAT-GNN, and HACT-Net all share the same options and hyperparameters below

- # PNA layers in GNN: [3, 4, 5]
- # MLP layers in a PNA layer: 2
- # channels in a PNA-layer MLP: 64
- Graph-level embedding dimension: 128
- # MLP layers in output classifier: 2
- # channels in output MLP classifier: 128
- Training parameters: Adam optimizer [75] with a learning rate of 10^{-3} , batch size of 16, and a categorical cross-entropy objective.

Evaluation metrics: Considering the imbalanced number of TRoIs per class in train, validation, and test sets (see Table I), we evaluate the classification performance using the weighted F1-score, an average weighted by the number of true instances for each class. During the training of each method, the model with the best weighted F1-scores in the validation set is selected as the final trained model of that method. To present any sensitivity to initialization, we report the mean and standard deviation of each model on the test set, by training them three times using random weight initialization. Further, we present precision, recall, and confusion matrices to indicate the distribution of class predictions.

Computational resources: All the experiments were conducted using PyTorch [76] and Deep Graph Library (DGL) [77], on NVIDIA Tesla P100 GPUs and POWER8 processors.

C. Ablation studies

We conduct ablation studies to evaluate the impact of three major components of our proposed methodology on TRoI classification performance, namely i) node feature initialization, ii) GNN layer type, and iii) jumping knowledge

Table II: Ablation: Impact of node features. Mean and standard deviation of 7-class weighted F1-scores. Results expressed in %.

	Weighed F1
CG-GNN: No morphological features	45.24±1.51
CG-GNN: Hand-crafted morphological features	48.34±5.22
CG-GNN: CNN morphological features	55.94±1.01
TG-GNN: No morphological features	36.81±0.71
TG-GNN: Hand-crafted morphological features	51.62±2.11
TG-GNN: CNN morphological features	56.62±1.35
CONCAT-GNN: No morphological features	47.62±1.56
CONCAT-GNN: Hand-crafted morphological features	51.55±1.32
CONCAT-GNN: CNN morphological features	57.01±2.27
HACT-Net: No morphological features	48.70±0.16
HACT-Net: Hand-crafted morphological features	52.46±0.19
HACT-Net: CNN morphological features	61.53±0.87

technique. Each component is analyzed individually, while fixing the others. Ablation studies were performed on the BRACS dataset for classifying the TRoIs into 7-classes.

1) *Impact of node feature initialization:* The performance of GNNs eminently rely on the initial node features [53]. In the context of entity-based tissue analysis, we analyze the impact of initial morphological features of the nodes. To this end, we experiment with the following three morphological feature initialization schemes:

- **No morphological features:** In this setting, the nodes of an entity-graph representation are initialized with only the spatial features. Experiments with this setting demonstrate the impact of standalone graph topology on the GNN performance.

- **Hand-crafted morphological features:** The entity-graph nodes are initialized with hand-crafted morphological features as suggested by [39], *i.e.*, i) *texture features*: difference of average foreground to background; standard deviation, skewness, and mean entropy of intensities; dissimilarity, homogeneity, energy, and angular second moment from Gray-Level Co-occurrence Matrix; and ii) *shape features*: eccentricity, area, maximum and minimum axis lengths, perimeter, solidity, and orientation. Note that, the hand-crafted features for CG and TG are computed, respectively, from the segmented instances of nuclei and tissue regions.

- **CNN morphological features:** The morphological features of the graph nodes are initialized with CNN features (ResNet-34 pre-trained on ImageNet) extracted from patches around the centroids of the nuclei and tissue regions.

Experimental results in Table II indicate that the standalone CG topology is more discriminative for cancer subtyping than TG topology. The combination of CG and TG topologies further improves discriminative ability. The best performance achieved with the HACT topology confirms the strength of hierarchical representations. Further,

Table III: Ablation: Impact of GNN layer. Mean and standard deviation of 7-class weighted F1-scores. Results expressed in %.

	Weighed F1
CG-GNN: GIN	55.70±0.51
CG-GNN: PNA	55.94±1.01
TG-GNN: GIN	55.33±1.36
TG-GNN: PNA	56.62±1.35
CONCAT-GNN: GIN	56.20±2.12
CONCAT-GNN: PNA	57.01±2.27
HACT-Net: GIN	59.73±1.20
HACT-Net: PNA	61.53±0.87

Table IV: Ablation: Impact of GNN jumping knowledge technique. Mean and standard deviation of 7-class weighted F1-scores. Results expressed in %.

	Weighed F1
CG-GNN: No aggregator	55.53±0.75
CG-GNN: Concatenation	55.82±0.97
CG-GNN: LSTM	55.94±1.01
TG-GNN: No aggregator	55.30±0.81
TG-GNN: Concatenation	56.07±0.80
TG-GNN: LSTM	56.62±1.35
CONCAT-GNN: No aggregator	57.67±4.66
CONCAT-GNN: Concatenation	56.28±2.75
CONCAT-GNN: LSTM	57.01±2.27
HACT-Net: No aggregator	49.16±1.15
HACT-Net: Concatenation	59.78±1.59
HACT-Net: LSTM	61.53±0.87

including morphological features significantly improves the discriminative ability. The superiority of graphs with CNN-based morphological features indicate the richness of morphological information acquired by CNNs, compared to hand-crafted measures.

2) *Impact of GNN layer type:* We investigate the impact of two state-of-the-art GNN layers, *i.e.*, GIN and PNA (Figure 1), on the classification performance. The experiments use CNN-based node feature initialization and LSTM-based jumping knowledge. Results in Table III demonstrate that GNNs with PNA layers outperform GNNs with GIN layers, for all the four GNN constructions.

3) *Impact of jumping knowledge technique:* To investigate the impact of the jumping knowledge technique, we experiment with three settings: no jumping knowledge, CONCAT-based, and LSTM-based. LSTM-based technique follows Equation (6). Based on this, CONCAT-based technique replaces the LSTM operation with the concatenation operation. The experiments use CNN-based node feature initialization and PNA layers. Results in Table IV

Table V: Mean and standard deviation of per-class F1-scores and weighted F1-scores for 7-class classification setting. Results are expressed in %. The best result is in **bold** and the second best is underlined.

	Method	Normal	Benign	UDH	ADH	FEA	DCIS	Invasive	Weighted F1
CNN	CNN(10×)	48.67±1.71	44.33±1.89	<u>45.00±4.97</u>	24.00±2.83	47.00±4.32	53.33±2.62	<u>86.67±2.64</u>	50.85±2.64
	CNN(20×)	42.00±2.16	42.33±3.09	39.33±2.05	22.67±2.49	47.67±1.25	50.33±3.09	77.00±1.41	46.85±2.19
	CNN(40×)	32.33±4.64	39.00±0.82	23.67±1.70	18.00±0.82	37.67±2.87	47.33±2.05	70.67±0.47	39.41±1.89
	Multi-scale CNN(10 × +20 ×)	48.33±2.05	45.67±0.47	41.67±4.99	32.33±0.94	46.33±1.41	59.33±2.05	85.67±1.89	52.27±1.93
	Multi-scale CNN(10 × +20 × +40 ×)	50.33±0.94	44.33±1.25	41.33±2.49	31.67±3.30	51.67±3.09	57.33±0.94	86.00±1.41	52.83±1.92
GNN	CGG-Net	30.83±5.33	31.63±4.66	17.33±3.38	24.50±5.24	58.97±3.56	49.36±3.41	75.30±3.20	43.63±0.51
	Patch-GNN(10×)	52.53±3.27	47.57±2.25	23.67±4.65	30.66±1.79	60.73±5.35	58.76±1.15	81.63±2.17	52.10±0.61
	Patch-GNN(20×)	43.86±4.23	43.37±3.21	19.47±2.31	25.73±2.87	55.57±2.08	52.86±1.85	79.20±1.04	47.10±0.70
	Patch-GNN(40×)	41.70±3.06	32.93±1.04	25.07±3.74	25.63±2.01	49.47±3.46	48.60±4.23	71.57±5.15	43.23±0.57
Ours	CG-GNN	58.77±6.82	40.87±3.05	46.82±1.95	<u>39.99±3.56</u>	63.75±10.48	53.81±3.89	81.06±3.33	55.94±1.01
	TG-GNN	63.59±4.88	47.73±2.87	39.41±4.70	28.51±4.29	<u>72.15±1.35</u>	54.57±2.23	82.21±3.99	56.62±1.35
	CONCAT-GNN	60.97±4.54	43.06±2.26	41.96±4.67	26.10±3.73	71.29±2.09	<u>60.83±3.71</u>	85.42±2.70	57.01±2.27
	HACT-Net (Proposed)	61.56±2.15	47.49±2.94	43.60±1.86	40.42±2.55	74.22±1.41	66.44±2.57	88.40±0.19	61.53±0.87

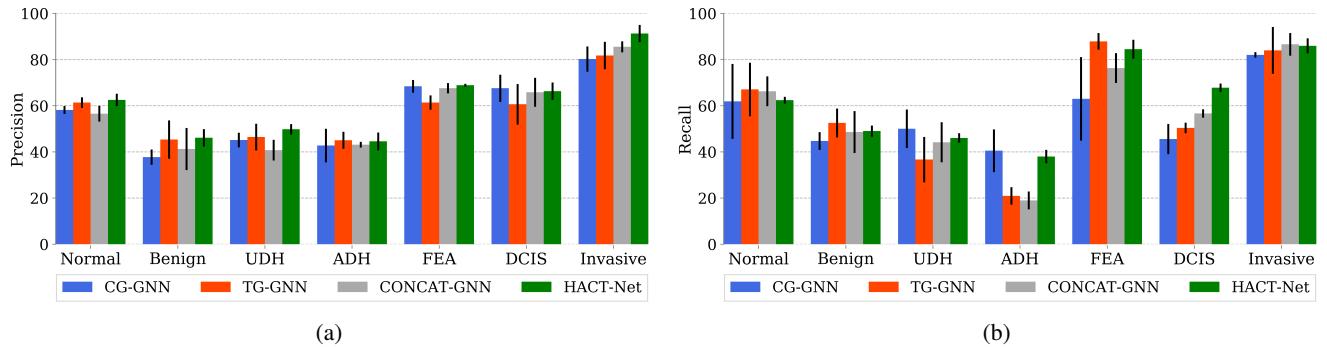


Figure 7: Mean and standard deviation of per-class precision and recall for 7-class classification setting. (Figure is best viewed in color)

demonstrate a generally positive impact of the jumping knowledge technique. Compared to CONCAT, the LSTM-based technique learns better dependencies between GNN layers, thus generates better graph embeddings.

4) *Ablation summary*: The ablation experiments conclude the following choice of components for designing our methodology, i) CNN-based initialization of node-level morphological features, ii) use of PNA layers, and iii) an LSTM-based jumping knowledge technique.

D. Classification results on BRACS dataset

We evaluate our proposed methods, comparatively with CNN and GNN baselines. To analyze the performance for different clinical applications and histopathological needs, we evaluate and report the results separately in the following three settings:

- **Setting 1: 7-class classification**: Here, we classify the TRoIs into 7-classes, *i.e.*, Normal, Benign, UDH, ADH, FEA, DCIS, and Invasive, for the differentiation of a large spectrum of breast cancer subtypes. Table V tabulates the classification performance of the compared methods.

Among single-scale CNNs, CNN(10×) is the best performing one, indicating the importance of global context information for TRoI classification. Multi-scale CNNs using both global and local context information outperform single scale CNNs. Such benefit from context is significant for ADH, FEA, and DCIS categories, which all require both local and global context information for the diagnosis. Multi-scale CNNs also outperform CGC-Net and Patch-GNN baselines. Interestingly, at each magnification, Patch-GNN outperforms single-scale CNN, which affirms the importance of relational and topological information incorporated in the form of patches.

Comparing our proposed GNN solutions, we observe that CG-GNN significantly outperforms CGC-Net, indicating the superiority of CNN-based node feature initialization over handcrafted features, and the significance of GNNs with expressive PNA layers over Adaptive GraphSage in CGC-Net. We notice that CG-GNN and TG-GNN provide comparable performance overall. However, TG-GNN performs better for Normal, Benign, and FEA subtypes, indicating the importance of tissue microenvironment information for these

Table VI: Mean and standard deviation of per-class F1-scores and weighted F1-scores for 4-class classification setting. Results are expressed in %. The best result is in **bold** and the second best is underlined.

	Method	Normal	Non-cancerous	Precancerous	Cancerous	Weighted F1
CNN	CNN(10×)	54.33±3.68	56.00±0.82	56.33±1.25	83.67±0.94	64.36±1.37
	CNN(20×)	45.33±4.64	55.33±0.47	52.33±1.89	81.67±2.05	61.18±1.93
	CNN(40×)	42.00±4.89	51.00±0.82	47.67±4.11	77.67±2.05	56.99±2.72
	Multi-scale CNN(10 × +20×)	51.67±5.79	55.33±1.25	52.67±2.87	80.67±1.89	61.82±2.53
	Multi-scale CNN(10 × +20 × +40×)	51.33±3.27	56.33±2.05	57.00±1.64	81.33±3.68	63.52±2.59
GNN	CGG-Net	34.53±2.93	47.23±3.72	62.90±2.81	82.20±1.04	59.87±2.30
	Patch-GNN(10×)	53.13±4.40	46.23±2.45	63.96±3.82	77.43±3.22	61.93±2.51
	Patch-GNN(20×)	53.46±1.81	47.16±2.81	63.20±3.78	74.90±3.36	61.26±2.90
	Patch-GNN(40×)	40.90±2.75	38.67±2.76	56.77±3.91	72.20±2.61	54.60±1.90
Ours	CG-GNN	52.95±12.11	<u>56.55±3.70</u>	61.53±3.03	<u>84.47±0.87</u>	66.10±2.58
	TG-GNN	52.96±6.81	56.52±2.85	<u>64.36±1.05</u>	82.21±0.78	<u>66.24±1.11</u>
	CONCAT-GNN	<u>54.54±1.64</u>	56.63±1.68	62.58±1.45	81.80±0.77	65.83±0.04
	HACT-Net (Proposed)	66.08±3.69	55.28±1.74	66.21±0.87	84.91±0.79	69.04±0.46

categories. In contrast, CG-GNN surpasses TG-GNN for UDH and ADH that rely on local nuclei-level information. For DCIS and Invasive, both CG-GNN and TG-GNN perform similarly, showing that both low-level and high-level information being useful for these categories. Further, both HACT-Net and CONCAT-GNN provide overall superior performance compared to all CNN and GNN baselines. HACT-Net significantly outperforms CONCAT-GNN indicating the significance of the proposed hierarchical modeling and learning. CONCAT-GNN produces overall comparable or superior performance to CG-GNN and TG-GNN, although for individual classes, CONCAT-GNN is rarely better than the two, suggesting that it may be utilizing complementary information in CG and TG representations. Such complementary information is indeed best utilized by HACT-Net, with high per-class and the overall classification performance. Any of the proposed GNNs often outperform all CNN baselines, establishing the potential of entity-based analysis in digital pathology.

Figure 7 presents per-class precision and recall for CG-GNN, TG-GNN, CONCAT-GNN, and HACT-Net. HACT-Net produces the highest precision values for most of the classes. The recall ranking between CG-GNN and TG-GNN varies across classes, whereas HACT-Net consistently yields good recall values. Further, standard deviation of class-wise precision and recall values are the lowest for HACT-Net, for most classes. Figure 8 presents row-normalized aggregated 7-class confusion matrix across three runs for HACT-Net. It indicates ambiguities between i) Normal and Benign, ii) UDH and ADH, and iii) ADH and DCIS. Notably, these pair-wise classes bear high pathological ambiguity and are diagnostically very challenging.

- **Setting 2: 4-class classification:** This setting categorizes TRoIs into 4-classes based on cancer risk: Normal, Non-cancerous (Benign + UDH), Precancerous (ADH +

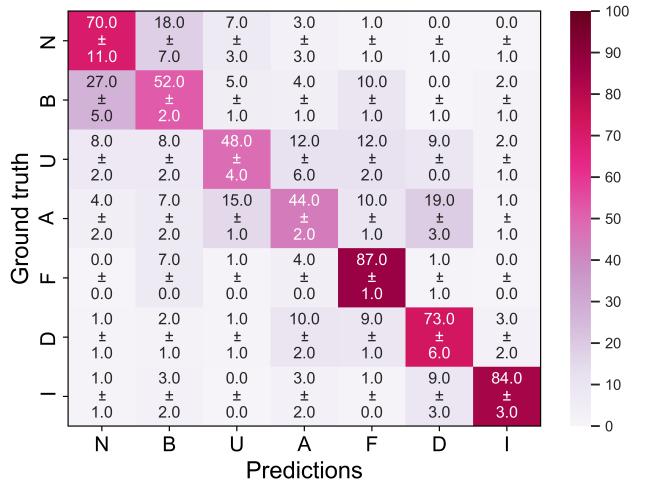


Figure 8: Mean and standard deviation of row-normalized 7-class confusion matrix for HACT-Net.

FEA), and Cancerous (DCIS + Invasive). Classification performance of HACT-Net as well as CNN and GNN baselines are presented in Table VI. Single scale CNNs exhibit the same behavior as in the 7-class setting. However, combining multiple magnifications in multi-scale CNNs does not improve the classification performance over the single scale CNNs. Among the baselines, CGC-Net and Patch-GNNs perform comparable or inferior to the CNNs, with a low-magnification CNN(10×) outperforming the others. Similarly to the 7-class setting, our proposed methods are superior to all the baselines. HACT-Net results in the best overall performance, with the best classification performance for Normal, Precancerous, and Cancerous categories. To highlight, HACT-Net achieves $\approx 66\%$ F1-score for the diagnostically challenging Precancerous category.

Table VII: Mean and standard deviation of weighted F1-scores for binary classification setting. Further, the aggregated mean and standard deviation for the six binary tasks are reported. Results are expressed in %. The best result is in **bold** and the second best is underlined.

	Method	I vs N+B+A+U+F+D	N+B+U vs A+F+D	N vs B+U	B vs U	A+F vs D	A vs F	Aggregated
CNN	CNN(10×)	95.66±0.48	81.24±0.42	69.83±0.38	76.12±1.13	73.44±2.56	77.59±1.73	78.90±1.38
	CNN(20×)	92.39±0.37	80.84±0.36	66.52±2.14	74.75±1.51	67.87±1.82	71.78±2.53	75.69±1.68
	CNN(40×)	90.74±0.59	79.92±1.66	62.36±2.14	68.13±4.30	64.86±2.98	66.91±1.68	72.15±2.51
	Multi-scale CNN(10 × +20 ×)	94.31±1.26	80.89±1.31	67.99±1.86	75.58±2.06	72.07±1.85	76.91±2.22	77.96±1.80
	Multi-scale CNN(10 × +20 × +40 ×)	95.12±1.15	82.21±0.34	70.87±2.07	72.89±2.26	72.08±3.17	75.47±3.69	78.11±2.40
GNN	CGG-Net	91.60±2.09	79.73±1.53	63.67±3.12	62.37±3.00	81.56±1.56	73.80±5.41	75.46±3.09
	Patch-GNN(10×)	95.80±0.43	76.53±0.32	72.57±1.10	72.87±3.07	77.17±0.85	78.26±2.60	78.87±1.75
	Patch-GNN(20×)	93.70±0.36	76.63±1.40	70.10±1.90	69.77±3.13	74.10±0.10	81.03±1.85	77.55±1.78
	Patch-GNN(40×)	92.40±0.95	74.43±0.64	71.10±1.74	67.40±2.46	72.97±0.66	76.40±1.95	75.78±1.56
Ours	CG-GNN (Ours)	94.52±0.43	83.79±0.31	<u>75.71±1.68</u>	73.15±3.32	77.48±1.68	84.33±0.54	81.50±1.70
	TG-GNN	96.00±0.56	80.38±0.80	69.51±3.12	<u>76.12±0.99</u>	<u>80.67±0.22</u>	84.18±3.56	81.14±2.02
	CONCAT-GNN	95.91±0.56	83.21±0.68	71.84±1.46	75.67±1.81	80.14±2.60	<u>88.88±3.86</u>	82.61±2.15
	HACT-Net (Proposed)	96.32±0.64	<u>83.63±0.73</u>	76.84±0.68	77.66±0.37	81.11±0.72	89.35±0.26	84.15±0.60

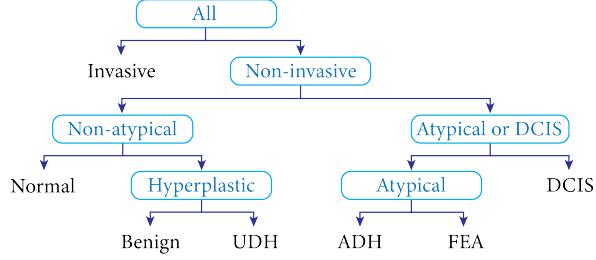


Figure 9: Decision tree used by pathologists for breast cancer diagnosis. The 7-class classification is simplified to a series of binary decision tasks, through which the diagnosis becomes more and more specific until the leaves, *i.e.*, the 7 diagnostic decision classes, are reached.

• **Setting 3: Binary classifications:** In this setting, we replicate the typical decision process of a pathologist following a diagnostic decision tree, presented in Figure 9. This is inspired by the classification scheme presented by [78], which mimics a pathologist’s point-of-view for breast cancer subtyping. Note that such individual binary decisions are less constrained compared to a multi-class classification task, thereby allowing for better discrimination between a selected pair of classes as well as allowing one to incorporate risk in each such decision. Such binary classifiers can assist pathologists in categorizing ambiguous cases, at different parts of a decision tree. Table VII presents the results for six separate binary classification tasks, at the bifurcations in the decision tree. Results are consistent with the 7-class and 4-class classification settings, with HACT-Net consistently outperforming all baselines and providing the best aggregated score.

1) *Domain expert comparison on BRACS dataset:* To further benchmark our proposed methodology as well as to assess the quality of the introduced BRACS dataset, we acquired annotations of the BRACS test set by additional independent pathologists. For such comparison with domain experts, we follow the evaluation protocol in [4]. We recruited three board-certified pathologists (other than the original three pathologists who provided the initial annotations, namely our ground truth labels), from three different medical centers, to further ensure independence: • National Cancer Institute- IRCCS-Fondazione Pascale, Naples, Italy; • Lausanne University Hospital, CHUV, Lausanne, Switzerland; and • Aurigen, Centre de Pathologie, Lausanne, Switzerland. These experts are specialized in breast pathology and have been in practice for over twenty years. The pathologists independently and remotely annotated BRACS test set TRoIs, without having access to respective WSIs. This protocol ensures equal field-of-view for all the pathologists as well as our methodology.

These new pathologist annotations are compared to the ground truth labels, with the results shown in Table VIII. We present per-class F1-scores, overall weighted F1-score, and overall weighted accuracy for each individual pathologist. Further, we include the aggregated statistics of all three pathologists for benchmarking HACT-Net with domain experts. Table VIII indicates that HACT-Net outperforms the domain experts in distinguishing TRoIs of diagnostically challenging classes, *i.e.*, atypia and hyperplasias, while yielding comparable performance for the normal and cancerous categories. Per-class standard deviations of pathologists’ statistics illustrate the expected high inter-observer variability in breast cancer diagnosis. Compared to the pathologists, HACT-Net yields a superior weighted accuracy and weighted F1 given the ground truth diagnoses

Table VIII: Comparison between HACT-Net and domain expert pathologists for 7-class breast cancer subtyping on BRACS dataset. Per-class F1-scores, weighted F1-scores and accuracy for 7-class classification are presented. Results are expressed in %. The best results are in **bold**.

	Normal	Benign	UDH	ADH	FEA	DCIS	Invasive	Weighted F1	Weighted Accuracy
Pathologist 1	67.53	53.92	41.90	36.00	19.13	71.59	94.00	55.30	56.71
Pathologist 2	47.83	52.94	25.00	35.37	65.22	68.00	94.00	57.07	57.99
Pathologist 3	39.66	49.59	49.43	42.29	54.12	65.19	89.47	56.71	56.55
Pathologist statistics	51.57±11.70	52.15±1.85	38.78±10.22	37.89±3.12	46.16±19.64	68.26±2.62	92.49±2.14	56.36±0.76	57.08±0.64
HACT-Net statistics	61.56±2.15	47.49±2.94	43.60±1.86	40.42±2.55	74.22±1.41	66.44±2.57	88.40±0.19	61.53±0.87	63.21±0.27

Table IX: Concordance among three independent pathologists for annotating BRACS test dataset. Results are expressed in %.

	Pathologist 1	Pathologist 2	Pathologist 3	Ground truth
Pathologist 1	-	47.60	50.96	56.71
Pathologist 2	-	-	64.38	57.99
Pathologist 3	-	-	-	56.55

for the 7-class classification task.

To benchmark the BRACS dataset with respect to the dataset by [4], we compare the aggregated pathologist statistics on both datasets for the same set of classes, *i.e.*, Benign without atypia (Normal + Benign + UDH), Atypia (ADH + FEA), DCIS, and Invasive. Note that the dataset by [4] consists of 240 breast biopsy slides, while BRACS consists of 626 TRoI images. For the dataset by [4], class-wise concordance rates (class-weighted average accuracy of 115 pathologists to a three-expert consensus) are 87%, 48%, 84%, and 96%, respectively for the four aforementioned classes. For BRACS, the similar class-wise concordance rates are 87%, 50%, 72%, and 90%, respectively. The class-wise concordance rates exhibit a similar trend in both datasets. Differences can be attributed to differing fields-of-view, *i.e.*, TRoI vs. WSI, accessible to the pathologist during annotation.

Table IX presents the inter-observer concordance rates for the BRACS test set. We notice significant differences in the concordance rates between pathologists 2 vs.3 and pathologist 1 vs. the other two. This can be reasoned to differing histopathology practices across different regions.

2) *Classification results on BACH dataset:* We evaluate the methods on the public BACH dataset. Considering its smaller training set of 400 images, we employ different image augmentation techniques for training HACT-Net. To this end, we employ rotation, mirroring, and color augmentations on the training images before extracting HACT graph representations. We do not use other graph augmentation techniques, such as random node and edge dropping, since these augmentations may hamper the meaningful topological distribution of the biological entities. The implementation strategies and hyperparameters in Section VI-B are em-

Table X: Accuracy of 4-class breast cancer subtyping in BACH dataset. Results are expressed in %.

	Methods	Accuracy
Ensemble networks	Wang et al. (2019) [80] [79] [16]	95.00 94.00 93.00 87.00 Kwok et al. (2018) [82]
Single network	HACT-Net	91.00

ployed for training HACT-Net. Classification performance of HACT-Net and the current state-of-the-art results on the BACH dataset are listed in Table X. Our predictions have been evaluated independently by the organizers of the BACH challenge, ensuring a fair with the other methods. HACT-Net results in comparable classification accuracy with the state-of-the-art methods. The difference in the accuracies are not significant considering only 100 TRoIs in the test set. Notably, our methodology employs a single, unified network, where the other listed competitors all employing an ensemble strategy with multiple networks at inference.

E. Qualitative analysis

Qualitative assessment of some results with our proposed method on the BRACS dataset is presented in Figure 10, where we compare our method with standalone CG and TG based learning. In Figure 11, we employ GRAPH-GRADCAM [51], [83], a post-hoc gradient based feature attribution technique, to highlight the nuclei and tissue-region nodes in CG and TG, respectively, to show what HACT-Net focuses on while classifying the TRoIs. Given the DCIS examples in Figs. 11(a-c&g-i), HACT-Net is seen to focus on the diagnostically relevant tumorous epithelial and the necrotic regions in TG, while ignoring the less important stroma and lumen, cf. Figs. 11(b,h). Further, within these relevant tissue regions, HACT-Net focuses on a subset of tumorous epithelial nuclei in CG, as seen in Figs. 11(c,i). Interestingly, we observe in Figs. 11(h,i) that HACT-Net utilizes complementary information from the necrotic region captured by TG, but not by CG. Similar

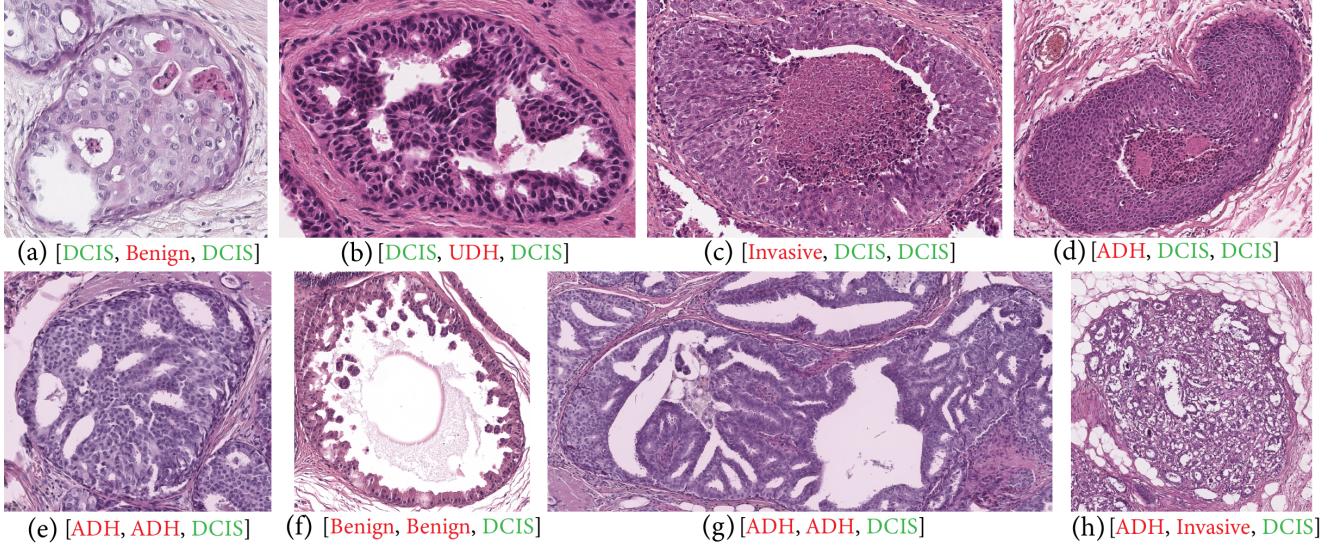


Figure 10: Qualitative comparison of CG-GNN, TG-GNN, and HACT-Net for 7-class classification. Predictions by the classifiers are noted below each example. Red and Green denote incorrect and correct classification, respectively. (a,b) TRoIs which TG-GNN misclassifies, while CG-GNN and HACT-Net classify correctly by using the nuclei characteristics. (c,d) TRoIs misclassified by CG-GNN, while correctly classified by TG-GNN and HACT-Net by using context information from necrotic regions. (e,f,g,h) TRoIs which both CG-GNN and TG-GNN misclassify, where HACT-Net classifies correctly by utilizing both cell and tissue microenvironments together. (Figure is best viewed in color)

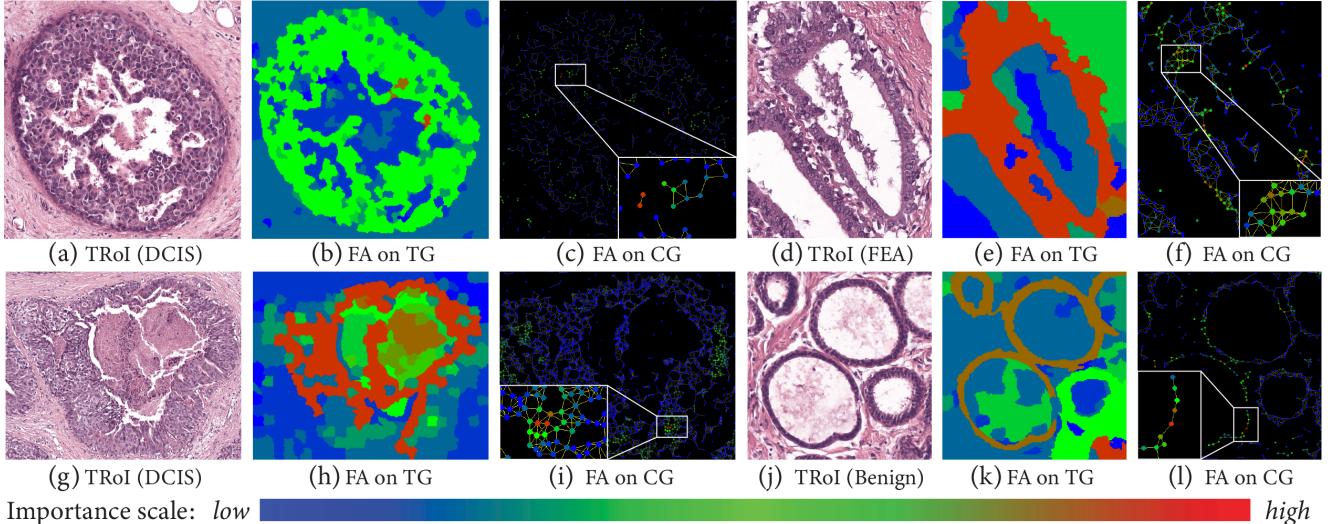


Figure 11: Feature attribution (FA) maps of HACT-Net on TG and CG for four sample TRoIs for 7-class classification: Sample TRoIs of (a,g) DCIS, (d) FEA, and (j) Benign classes, with their corresponding feature attribution maps on (b,h,e,k) TG and (c,i,f,l) CG. (Figure is best viewed in color)

observations of HACT-Net considering the diagnostically relevant regions can also be made for FEA and Benign examples seen in Figs 11(d-f&j-l). Noticeably, such feature attribution analysis GNNs localizes and highlights the focus of deep networks in the given entity-paradigm, which is both more interpretable and more explainable compared to feature attribution strategies in a pixel-paradigm [49], [51].

VII. CONCLUSION

Pixel-based processing of pathology images suffers from the context-resolution trade-off, and misses the notion of biological entity and tissue composition. In this work, we propose an entity-based tissue representation and learning to address the issues with pixel-based processing. To that end, our two specific contributions

are: (i) a hierarchical entity-graph representation of a tissue image by incorporating multisets of pathologically intuitive biological entities, and (ii) a hierarchical graph neural network for sequentially processing the entity-graph representation for mapping tissue compositions to respective tissue classes. Further, we introduce BReAst Cancer Subtyping (BRACS), a large cohort of breast tumor regions-of-interest, annotated with breast cancer subtypes. BRACS encompasses seven breast cancer subtypes to represent a realistic breast cancer diagnosis scenario. Using BRACS as well as a public breast cancer subtyping dataset BACH, we demonstrate herein the superior performance of our proposed methodology for classifying breast tumor regions-of-interest into cancer subtypes. Under various experimental settings, our methodology is shown to outperform state-of-the-art pixel-based and entity-graph based classification approaches. Furthermore, we benchmark our methodology on the BRACS dataset by comparing it to three independent pathologists. Notably, our method achieves better performance for per-cancer subtype and overall aggregated classification. Although we have evaluated our method for breast cancer classification, the technology is easily extendable to other tissue types and diseases. Notably, the proposed hierarchical graph methodology can also be adapted to other image modalities, such as natural images, multiplexed images, hyperspectral images, satellite images, and other medical imaging domains, by utilizing domain and task-specific entities.

REFERENCES

- [1] H. Sung *et al.*, “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” in *CA Cancer J. clinicians*, 2021, pp. 1–41.
- [2] C. Allemani *et al.*, “Global surveillance of cancer survival 1995–2009: Analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (concord-2),” in *The Lancet*, vol. 385, no. 9972, mar 2015, pp. 977–1010.
- [3] D. Gomes *et al.*, “Inter-observer variability between general pathologists and a specialist in breast pathology in the diagnosis of lobular neoplasia, columnar cell lesions, atypical ductal hyperplasia and ductal carcinoma in situ of the breast,” in *Diagnostic Pathology*, vol. 9, no. 121, 2014.
- [4] J. Elmore *et al.*, “Diagnostic concordance among pathologists interpreting breast biopsy specimens,” in *JAMA*, vol. 313, no. 11, 2015, pp. 1122–1132.
- [5] R. Siegel, K. Miller, and A. Jemal, “Cancer statistics, 2020,” in *CA: A Cancer Journal for Clinicians*, vol. 70, 2020, pp. 7–30.
- [6] S. Mukhopadhyay *et al.*, “Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivot study),” in *Am J Surg Pathol*, vol. 42, no. 1, 2017, pp. 39–52.
- [7] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” in *Medical Image Analysis*, vol. 42, 2017, pp. 60–88.
- [8] S. Deng *et al.*, “Deep learning in digital pathology image analysis: A survey,” in *Frontiers of Medicine*, 2020.
- [9] A. Ibrahim *et al.*, “Artificial intelligence in digital breast pathology: Techniques and applications,” in *The Breast*, vol. 49, 2020, pp. 267–273.
- [10] N. Kumar *et al.*, “A dataset and a technique for generalized nuclear segmentation for computational pathology,” in *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, 2017, pp. 1550–1560.
- [11] S. Graham *et al.*, “Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images,” in *Medical Image Analysis*, vol. 58, 2019.
- [12] P. Pati *et al.*, “Reducing annotation effort in digital pathology: A co-representation learning framework for classification tasks,” in *Medical Image Analysis*, vol. 67, 2021.
- [13] R. Verma *et al.*, “Multi-organ nuclei segmentation and classification challenge,” 2020.
- [14] S. Graham *et al.*, “Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images,” in *Medical Image Analysis*, vol. 52, 2019, pp. 199–211.
- [15] T. Binder *et al.*, “Multi-organ gland segmentation using deep learning,” in *Frontiers in Medicine*, 2019.
- [16] G. Aresta *et al.*, “Bach: Grand challenge on breast cancer histology images,” in *Medical Image Analysis*, vol. 56, 2019, pp. 122–139.
- [17] B. Bejnordi *et al.*, “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” in *JAMA*, vol. 318, no. 22, 2019, p. 2199–2210.
- [18] P. Pati *et al.*, “Deep positive-unlabeled learning for region of interest localization in breast tissue images,” in *SPIE Medical Imaging 2018: Digital Pathology*, vol. 10581, 2018, p. 1058107.
- [19] C. Mercan *et al.*, “From patch-level to roi-level deep feature representations for breast histopathology classification,” in *SPIE Medical Imaging 2019: Digital Pathology*, vol. 10956, 2019, p. 109560H.
- [20] A. Madabhushi and G. Lee, “Image analysis and machine learning in digital pathology: Challenges and opportunities,” in *Medical Image Analysis*, vol. 33, 2016, pp. 170–175.
- [21] A. Parwani, “Next generation diagnostic pathology: use of digital pathology and artificial intelligence tools to augment a pathological diagnosis,” in *Diagnostic Pathology*, vol. 14, no. 138, 2019.
- [22] B. Bejnordi *et al.*, “Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images,” in *Journal of Medical Imaging*, vol. 4, no. 4, 2017.

- [23] K. o. Sirinukunwattana, “Improving whole slide segmentation through visual context - a systematic study,” in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 11071, 2018.
- [24] D. Tellez *et al.*, “Neural image compression for gigapixel histopathology image analysis,” vol. 58, 2019.
- [25] M. Hagele *et al.*, “Resolving challenges in deep learning-based analyses of histopathological images using explanation methods,” vol. 10, 2020.
- [26] C. Demir *et al.*, “The cell graphs of cancer,” in *Bioinformatics*, vol. 20, 2004, pp. 145–151.
- [27] P. Pati *et al.*, “Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification,” in *Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on GRaphs in biomedical Image anaLysis*, 2020.
- [28] D. Komura and S. Ishikawa, “Machine learning methods for histopathological image analysis,” in *Computational and Structural Biotechnology Journal*, 2018, pp. 34–42.
- [29] C. Srinidhi *et al.*, “Deep neural network models for computational histopathology: A survey,” in *Medical Image Analysis*, vol. 67, 2021.
- [30] F. Spanhol *et al.*, “A dataset for breast cancer histopathological image classification,” in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, 2016, pp. 1455–1462.
- [31] T. Araujo *et al.*, “Classification of breast cancer histology images using convolutional neural networks,” in *PloS one*, vol. 12, no. 6, 2005.
- [32] D. Bardou *et al.*, “Classification of breast cancer based on histology images using convolutional neural networks,” in *IEEE Access*, vol. 6, 2018, pp. 24 680–24 693.
- [33] K. Roy *et al.*, “Patch-based system for classification of breast histology images using deep learning,” in *Computerized Medical Imaging and Graphics*, vol. 71, 2019, pp. 90–103.
- [34] M. Shaban *et al.*, “Context-aware convolutional neural network for grading of colorectal cancer histology images,” in *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, 2020, pp. 2395 – 2405.
- [35] R. Yan *et al.*, “Breast cancer histopathological image classification using a hybrid deep neural network,” in *Methods*, vol. 173, 2020, pp. 52–60.
- [36] H. Pinckaers, B. van Ginneken, and G. Litjens, “Streaming convolutional neural networks for end-to-end learning with multi-megapixel images,” in *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, 2020, pp. 1306–1315.
- [37] G. Campanella *et al.*, “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images,” in *Nature Medicine*, vol. 25, no. 8, 2019, p. 1301–1309.
- [38] M. Lu *et al.*, “Data efficient and weakly supervised computational pathology on whole slide images,” in *Nat Biomed Eng*, 2021.
- [39] Y. Zhou *et al.*, “CGC-net: Cell graph convolutional network for grading of colorectal cancer histology images,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, 2019.
- [40] R. Chen *et al.*, “Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis,” in *IEEE Transactions on Medical Imaging*, 2020.
- [41] H. Sharma *et al.*, “A review of graph-based methods for image analysis in digital histopathology,” in *Diagnostic Pathology*, 2015.
- [42] ———, “Cell nuclei attributed relational graphs for efficient representation and classification of gastric cancer in digital histopathology,” in *SPIE Medical Imaging: Digital Pathology*, vol. 9791, 2016.
- [43] ———, “A comparative study of cell nuclei attributed relational graphs for knowledge description and categorization in histopathological gastric cancer whole slide images,” in *IEEE Symposium on Computer-Based Medical Systems*, 2017, pp. 61–66.
- [44] D. Anand *et al.*, “Histograms: graphs in histopathology,” in *SPIE Medical Imaging 2020: Digital Pathology*, vol. 11320, 2020, p. 113200O.
- [45] B. Aygunes *et al.*, “Graph convolutional networks for region of interest classification in breast histopathology,” in *SPIE Medical Imaging 2020: Digital Pathology*, vol. 11320, 2020, p. 113200K.
- [46] S. Javed *et al.*, “Cellular community detection for tissue phenotyping in colorectal cancer histology images,” in *Medical Image Analysis*, vol. 63, 2020.
- [47] Y. Zhao *et al.*, “Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4837–4846.
- [48] M. Adnan *et al.*, “Representation learning of histopathology images using graph neural networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020, pp. 4254–4261.
- [49] G. Jaume *et al.*, “Towards explainable graph representations in digital pathology,” in *International Conference on Machine Learning (ICML), Workshop on Computational Biology*, 2020.
- [50] M. Sureka *et al.*, “Visualization for histopathology images using graph convolutional neural networks,” in *arXiv:2006.09464*, 2020.
- [51] G. Jaume *et al.*, “Quantifying explainers of graph neural networks in computational pathology,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [52] M. Defferrard *et al.*, “Convolutional neural networks on graphs with fast localized spectral filtering,” in *Neural Information Processing Systems (NeurIPS)*, 2016, pp. 3844–3852.

- [53] T. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *International Conference on Learning Representations (ICLR)*, 2017.
- [54] K. Xu *et al.*, “How powerful are graph neural networks?” in *International Conference on Learning Representations (ICLR)*, 2019.
- [55] W. Hamilton *et al.*, “Inductive representation learning on large graphs,” in *Neural Information Processing Systems (NeurIPS)*, 2017, pp. 1024–1034.
- [56] P. Velickovic *et al.*, “Graph attention networks,” in *International Conference on Learning Representations (ICLR)*, 2018.
- [57] J. Gilmer *et al.*, “Neural message passing for quantum chemistry,” in *International Conference on Machine Learning (ICML)*, vol. 70, 2017, pp. 1263–1272.
- [58] C. Morris *et al.*, “Weisfeiler and leman go neural: Higher-order graph neural networks,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2018.
- [59] G. Jaume *et al.*, “edggnn: a simple and powerful gnn for directed labeled graphs,” in *International Conference on Learning Representations (ICLR) Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- [60] B. Weisfeiler and A. A. Lehman, “A reduction of a graph to a canonical form and an algebra arising during this reduction,” in *Nauchno-Technicheskaya Informatsia*, vol. 2, no. 9, 1968, pp. 12–16.
- [61] N. Dehmamy *et al.*, “Understanding the representation power of graph neural networks in learning graph topology,” in *Neural Information Processing Systems (NeurIPS)*, 2019, pp. 15 413–15 423.
- [62] G. Corso *et al.*, “Principal neighbourhood aggregation for graph nets,” in *Neural Information Processing Systems (NeurIPS)*, 2020.
- [63] M. Veta *et al.*, “Breast cancer histopathology image analysis: A review,” in *IEEE Transactions on Biomedical Engineering*, no. 5, 2014, pp. 1400–1411.
- [64] D. Tellez *et al.*, “Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology,” in *Medical Image Analysis*, vol. 58, 2019.
- [65] M. Macenko *et al.*, “A method for normalizing histology slides for quantitative analysis,” in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2009, pp. 1107–1110.
- [66] M. Stanisljevic *et al.*, “A fast and scalable pipeline for stain normalization of whole-slide images in histopathology,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
- [67] K. He *et al.*, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [68] J. Deng *et al.*, “Imagenet: A large-scale hierarchical image database,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.
- [69] K. Francis and B. Palsson, “Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion,” vol. 94, no. 23, 1997, pp. 12 258–12 262.
- [70] R. Achanta *et al.*, “Slic superpixels compared to state-of-the-art superpixel methods,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, 2012, pp. 2274–2282.
- [71] F. Potjer, “Region adjacency graphs and connected morphological operators,” in *Mathematical Morphology and its Applications to Image and Signal Processing. Computational Imaging and Vision*, vol. 5, 1996, p. 111–118.
- [72] K. Xu *et al.*, “Representation learning on graphs with jumping knowledge networks,” in *International Conference on Machine Learning (ICML)*, 2018.
- [73] V. Dwivedi *et al.*, “Benchmarking graph neural networks,” in *arXiv:2003.00982*, 2020.
- [74] P. Bankhead *et al.*, “Qupath: Open source software for digital pathology image analysis,” in *Scientific reports*, vol. 7, no. 1, 2017, pp. 1–7.
- [75] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *International Conference on Learning Representations (ICLR)*, 2015.
- [76] A. Paszke *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” in *Neural Information Processing Systems (NeurIPS)*, 2019, pp. 8024–8035.
- [77] M. Wang *et al.*, “Deep graph library: Towards efficient and scalable deep learning on graphs,” in *CoRR*, vol. abs/1909.01315, 2019.
- [78] E. Mercan *et al.*, “Automated diagnosis of breast cancer and pre-invasive lesions on digital whole slide images,” in *ICPRAM*, 2018.
- [79] G. Aresta *et al.*, “Iciar 2018 grand challenge on breast cancer histology images,” in *ICCIAR*, 2018. [Online]. Available: <https://iciar2018-challenge.grand-challenge.org/Home/>
- [80] B. Marami *et al.*, “Ensemble network for region identification in breast histopathology slides,” in *International Conference Image Analysis and Recognition*, 2018, pp. 861–868.
- [81] S. Chennamsetty, M. Safwan, and V. Alex, “Classification of breast cancer histology image using ensemble of pre-trained neural networks,” in *International Conference Image Analysis and Recognition*, 2018, pp. 804–811.
- [82] N. Brancati, M. Frucci, and D. Riccio, “Multi-classification of breast cancer histology images by using a fine-tuning strategy,” in *International Conference Image Analysis and Recognition*, 2018, pp. 771–778.
- [83] P. Pope *et al.*, “Explainability methods for graph convolutional neural networks,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10 764–10 773.