Medical Image Analysis 75 (2022) 102264

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Hierarchical graph representations in digital pathology

Pushpak Pati[a,b,*], Guillaume Jaume[a,c], Antonio Foncubierta-Rodríguez[a], Florinda Feroce[d], Anna Maria Anniciello[d], Giosue Scognamiglio[d], Nadia Brancati[e], Maryse Fiche[f], Estelle Dubruc[g], Daniel Riccio[e], Maurizio Di Bonito[d], Giuseppe De Pietro[e], Gerardo Botti[d], Jean-Philippe Thiran[c], Maria Frucci[e], Orcun Goksel[b,h], Maria Gabrani[a]

[a] IBM Zurich Research Lab, Zurich, Switzerland
[b] Computer-Assisted Applications in Medicine, ETH Zurich, Zurich, Switzerland
[c] Signal Processing Laboratory 5, EPFL, Lausanne, Switzerland
[d] National Cancer Institute - IRCCS-Fondazione Pascale, Naples, Italy
[e] Institute for High Performance Computing and Networking - CNR, Naples, Italy
[f] Aurigen- Centre de Pathologie, Lausanne, Switzerland
[g] Lausanne University Hospital, Lausanne, Switzerland
[h] Department of Information Technology, Uppsala University, Sweden

ARTICLE INFO

Article history:
Received 11 March 2021
Revised 22 September 2021
Accepted 6 October 2021
Available online 27 October 2021

Keywords:
Digital pathology
Breast cancer classification
Cell graph representation
Tissue graph representation
Hierarchical tissue representation
Hierarchical graph neural network
Breast cancer dataset

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, capturing the cell-microenvironment, to depict the tissue. These allow for utilizing graph theory and machine learning to map the tissue representation to tissue functionality, and 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 the 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 and 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 the 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. The code, data, and models can be accessed at https://github.com/histocartography/hact-net.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Breast cancer is the most commonly diagnosed cancer and registers the highest number of deaths for women with cancer (Sung et al., 2021). A study by Allemani et al. (2015) exhibits that intensive early diagnostic activities have improved 5-year survival to 85% during 200509 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.

* Corresponding author.
E-mail address: pus@zurich.ibm.com (P. Pati).

https://doi.org/10.1016/j.media.2021.102264
1361-8415/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Thus, manual inspection is tedious and time-consuming with significant intra- and inter-observer variability (Gomes et al., 2014; Elmore et al., 2015). The increasing incidence rate of breast cancer cases per year (Siegel et al., 2020) 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) and profoundly transformed pathologists' daily practice (Mukhopadhyay et al., 2017). Further, they enabled computer aided diagnostics to leverage artificial intelligence (Litjens et al., 2017; Deng et al., 2020), especially deep learning, to address various pathology tasks, such as nuclei segmentation (Kumar et al., 2017; Graham et al., 2019a), nuclei classification (Pati et al., 2021; Verma et al., 2020), gland segmentation (Graham et al., 2019b; Binder et al., 2019), tissue segmentation (Mehta et al., 2018; Mercan et al., 2019b), tumor detection (Aresta et al., 2019; Bejnordi et al., 2019; Pati et al., 2018), tumor staging (Aresta et al., 2019; Mercan et al., 2019a), and survival analysis (Zhu et al., 2017; Yao et al., 2021). Deep learning techniques primarily use Convolutional Neural Networks (CNN) (Madabhushi and Lee, 2016; Parwani 2019) to process histology images in a patch-wise manner. CNNs extract representative patterns from patches and aggregate them to perform image-level tasks. However, patch-wise processing suffers from the trade-off between the resolution of operation and the utilization of adequate context (Bejnordi et al., 2017; Sirinukunwattana et al., 2018). Operating at a higher resolution captures local cellular information but limits the field-of-view due to computational burden and limits the access to global tissue microenvironment information. In contrast, operating at a lower resolution hinders resolvability of cells and access to cellular properties. Bejnordi et al. (2017); Sirinukunwattana et al. (2018); Tellez et al. (2019a) 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 to incorporate information from varying spatial distances. Further, pixel-based processing in CNNs disregards the notion of histologically meaningful entities (Hagele et al., 2020), such as cells, glands, and tissue types. The inattention to histological entities severely limits the interpretability of CNNs by pathologists, and any utilization of established entity-level prior pathological knowledge in the CNN-based diagnostic 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 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 aspects, *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. Demir et al. (2004) 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 composition. To address this, 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 Pati et al. (2020). 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 (BRACS[1]), 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 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.

2. Related work

Tumor subtyping in digital pathology: Several deep learning algorithms have been proposed to categorize histopathology images into cancer subtypes (Komura and Ishikawa, 2018; Srinidhi et al., 2021; Deng et al., 2020; Spanhol et al., 2016; Araujo et al. (2005); Aresta et al., 2019). For this task, most algorithms employ CNNs in a patch-wise manner: In Araujo et al. (2005); Bardou et al. (2018); Roy et al. (2019); Mercan et al. (2019a), CNNs are used to classify breast histology images. These methods use single stream patch-wise approaches to capture local patch-level context, aggregate the patch-level information, and classify the image using aggregated information. However, single-stream approaches do not capture adequate context from the tissue microenvironment to aptly encode a patch. Sirinukunwattana et al. (2018) address this issue by including multi-scale information from concentric patches across different magnifications. Tellez et al. (2019a) propose neural image compression, where WSIs are compressed using a neural network trained in an unsupervised fashion, followed by a CNN trained on the compressed representations to classify the images.

[1] BRACS dataset for breast cancer subtyping: https://www.bracs.icar.cnr.it.

Shaban et al. (2020) include an attention module with an auxiliary task to improve neural image compression for histology image classification. Yan et al. (2020) propose a hybrid convolutional and recurrent neural network to utilize spatial correlations among patches for analyzing histology images. Bejnordi et al. (2017) propose a stacked CNN architecture to capture large contexts and perform end-to-end processing of large histology images. Pinckaers et al. (2020) propose a streaming CNN to accommodate multi-megapixel images. Campanella et al. (2019) utilize a multiple instance learning approach to process whole-slide images in an end-to-end manner, which is extended by Lu et al. (2021) to automatically identify sub-regions of high diagnostic value through an attention mechanism. Though the aforementioned methods use different strategies to encode a tissue, they all operate on a square and fix-sized patches. However, actual TRoIs can be of highly varying dimensions and shapes depending on the 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 tissue composition by incorporating morphology, topology, and interactions among biologically comprehensible entities, unlike CNNs. Using cells as entities, Demir et al. (2004) introduced a cell-graph (CG) representation of a tissue, where cell morphology can be embedded in the nodes via hand-crafted (Demir et al., 2004; Zhou et al., 2019; Pati et al., 2020) or deep-learning based features (Chen et al., 2020). The graph topology is often heuristically defined, *e.g.*, using k-Nearest Neighbors, probabilistic modeling, or a Waxman model (Sharma et al., 2015). Subsequently, a CG is processed by classical machine learning techniques (Sharma et al., 2016, 2017) or GNNs (Zhou et al., 2019; Pati et al., 2020; Chen et al., 2020; Anand et al., 2020) for mapping to tissue function. Recently, graph representations using patches (Aygunes et al., 2020) and tissue regions (Pati et al. (2020); Anklin et al., 2021) as entities have been proposed for better tissue representation. Other graph-based applications in computational pathology include cellular community detection (Javed et al., 2020), WSI classification (Zhao et al., 2020; Adnan et al., 2020), WSI segmentation (Anklin et al., 2021). Notably, entity-graphs consist of biological entities to which the pathologists can readily relate. So, the entity-graph paradigm enables to incorporate pathologically-defined, task-specific entity-level prior knowledge in constructing "meaningful" tissue representations. This implicitly enables *interpretability* and *explainability* of graph-based networks for pathologists. To this end, Zhou et al. (2019) analyzes the clustering of nodes in a CG to group cells according to their appearance and tissue types. Jaume et al. (2020) introduces a post-hoc graph-pruning explainer to identify decisive cells and interactions. Sureka et al. (2020) employs robust spatial filtering that utilizes an attention-based GNN and node occlusion to highlight cell contributions. Jaume et al. (2021b) propose quantitative metrics leveraging pathologically relevant cellular properties to characterize graph explainability for CG analysis.

3. Preliminaries

3.1. 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 $|.|$ 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 $\mathcal{N}(v) := \{u \in V \mid v \in V,\ e_{uv} \in E\}$.

3.2. Graph neural networks

GNN (Kipf and Welling 2017; Xu et al., 2019; Hamilton et al., 2017; Velickovic et al., 2018) defines a class of neural networks that operate on graph-structured data. In this work, we use message-passing GNNs (Gilmer et al., 2017), where node features $h(v), \forall v \in V$ are iteratively updated in two-steps, 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 features 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 up to 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 for 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,

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

An important aspect of designing a GNN is the characterization of its expressive power, which is measured by the GNN's ability to map non-isomorphic graphs to unique graph embeddings, denoting an injective mapping between the graph and the embedding space. A line-of-research exploring the expressive power of GNNs (Xu et al., 2019; Morris et al., 2018; Jaume et al., 2019) highlight the connection between iterative message passing steps of GNN and the popular Weisfeiler-Lehman (WL) (Weisfeiler and Lehman 1968) test for graph isomorphism. It is established that architectures such as the Graph Isomorphism Network (GIN) Xu et al. (2019) can perform as well as the 1-dimensional WL test for *discrete* node feature spaces, *e.g.*, the study of molecule design where the nodes represent atoms that are discrete in nature. Recent studies show that for *continuous* node features, *e.g.*, CNN-based node features, the use of multiple permutation-invariant aggregators, such as sum, max and mean, can build expressive GNNs (Dehmamy et al. (2019); Corso et al., 2020). To this end, Corso et al. (2020) proposed the Principal Neighbourhood Aggregation (PNA) network by using a combination of aggregators with *degree-scalers*. The series of aggregators replace the sum operation in GIN and the degree-scalers scale neighboring aggregated-messages according to the node degree. Illustrations of GIN and PNA architectures are presented in Fig. 1.

4. Methodology

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

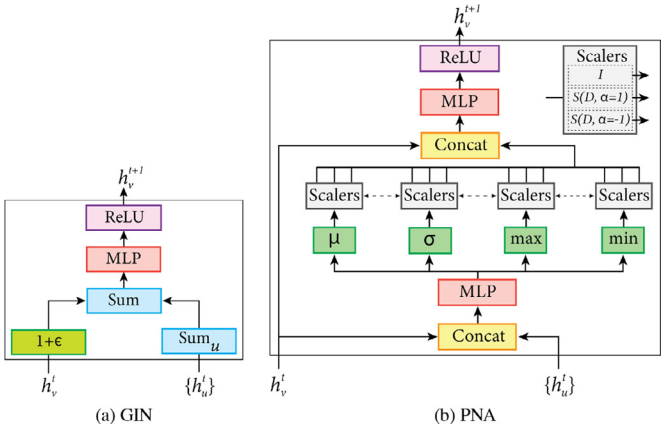

Fig. 1. Overview of GIN and PNA layers, where h_v^t and $\{h_u^t\}$ denote the representation of, respectively, node v and its neighbors at layer t. GIN uses sum as the *AGGREGATE* function, followed by a sum and multi-layer perceptron (MLP) for the *UPDATE* function. PNA uses a set of aggregators (element-wise 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.

4.1. Pre-processing

H&E stained tissue images exhibit appearance variability due to various reasons, such as different specimen preparation techniques, staining protocols, fixation characteristics, and imaging device characteristics. Such variability adversely impacts computational methods for downstream diagnosis Veta et al. (2014); Tellez et al. (2019b). To alleviate the variability, we use the unsupervised, reference-free stain normalization algorithm proposed by Macenko et al. (2009). 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 requiring no model training is computationally inexpensive. Specifically, for stain normalization we employ the scalable and fast pipeline proposed by Stanisavljevic et al. (2018).

4.2. Graph representation

A stain normalized TRoI is processed to identify relevant entities and construct a hierarchical entity-graph representation. In this work, we consider nuclei and tissue regions as the entities. Therefore, the HACT graph 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 the relative spatial distribution of cells with respect to the tissue distribution. The details of the components are presented in the following subsections.

4.2.1. Cell-graph representation

A cell-graph (CG) characterizes cell microenvironment, where nodes denote cells and encode cell morphology, and edges denote cellular interactions and encode cell topology. It is constructed in three steps, i) nuclei detection, ii) nuclei feature extraction, and iii) topology configuration, as shown in Fig. 3.

Precise nuclei detection enables reliable CG representation. To this end, we use HoVer-Net, a nuclei segmentation network proposed by Graham et al. (2019a), pre-trained on MoNuSeg dataset by Kumar et al. (2017). HoVer-Net leverages the instance-rich information encoded in the vertical and horizontal distances of nuclear pixels to their centers of mass. These distances are used to accurately segment clustered nuclei, particularly in areas with overlapping nuclei. The centroids of the segmented nuclei form 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 He et al. (2016) pre-trained on ImageNet dataset Deng et al. (2009). 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 collocated for all nodes as the node-feature matrix $H_{CG} \in \mathbb{R}^{|V_{CG}| \times d_{CG}}$.

For the CG topology E_{CG}, we utilize the fact that spatially close cells have stronger interactions (Francis and Palsson 1997) with distant cells having weaker cellular interactions. 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

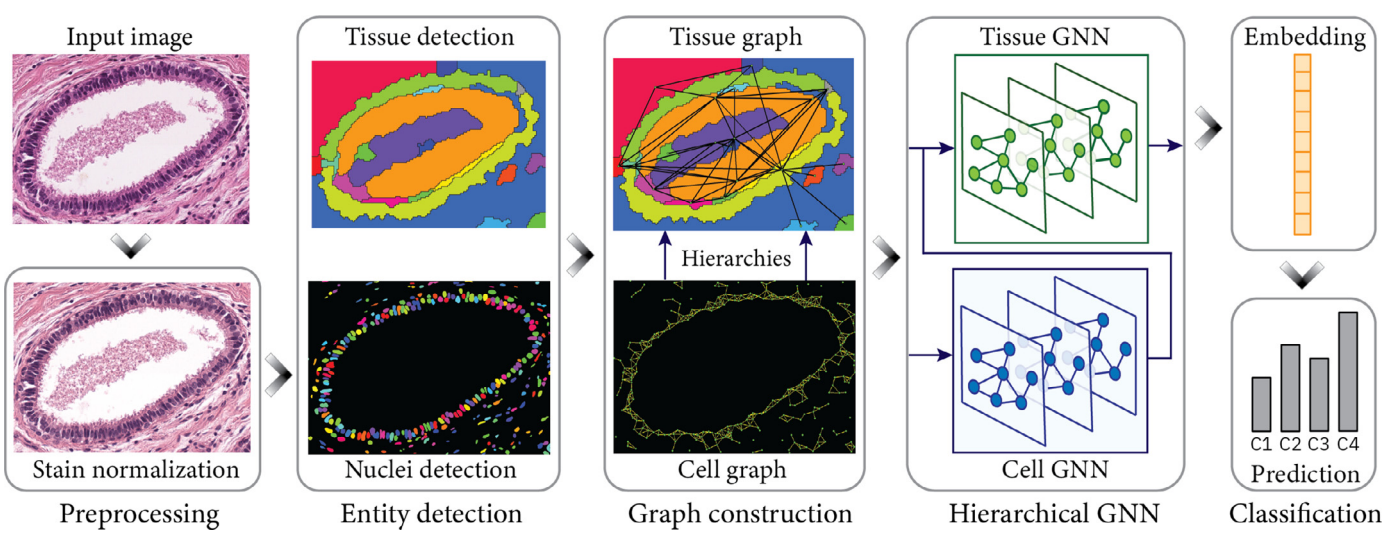

Fig. 2. Overview of the proposed hierarchical entity-graph based tissue analysis methodology. Following some pre-processing, a hierarchical entity-graph representation of a tissue is constructed, and it is processed via a hierarchical graph neural network to learn the mapping from tissue compositions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

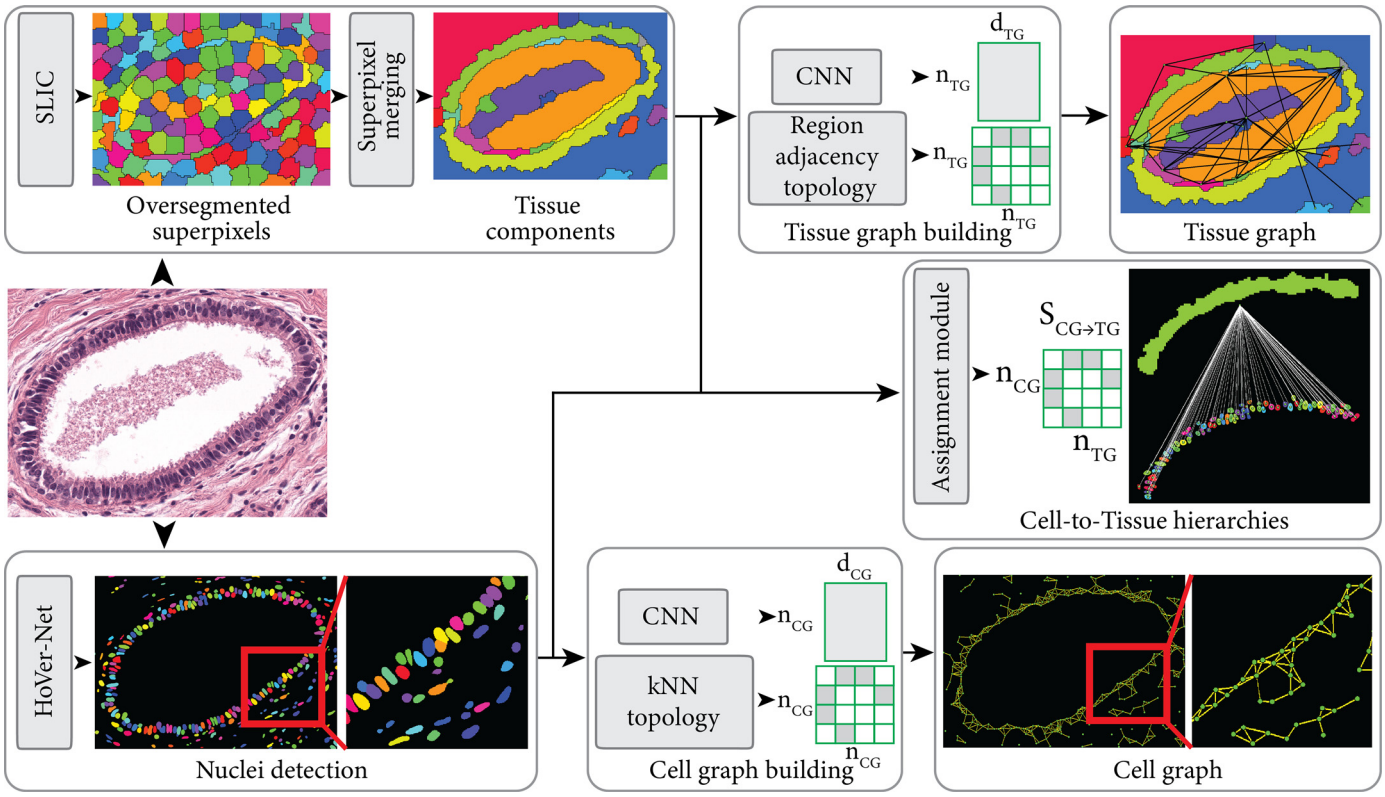

Fig. 3. Overview of hierarchical cell-to-tissue (HACT) graph construction for a TRoI. Our HACT graph representation consists of a cell-graph, a tissue-graph, and cell-to-tissue hierarchies, while encoding the phenotypical and topological distributions of tissue entities to describe the cell and tissue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

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

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

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

4.2.2. Tissue-graph representation

A tissue graph (TG) depicts a high-level tissue microenvironment, where the nodes and edges denote tissue regions and their interactions, respectively. A TG is constructed by first identifying tissue regions (*e.g.*, epithelium, stroma, lumen, necrosis etc.), followed by encoding the tissue regions, and finally the topology building. The steps are illustrated in Fig. 3. A parallel approach involving superpixel detection and neighborhood information aggregation is adopted by Mercan et al. (2018) to semantically segment tissue regions in histology images.

Tissue regions are identified in two-steps. First, we oversegment the tissue to detect non-overlapping homogeneous superpixels. We operate at a low magnification to avoid noisy pixels and compute efficiently. Specifically, we use the Simple Linear Iterative Clustering (SLIC) algorithm (Achanta et al., 2012). 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, to create superpixels that capture meaningful tissue information. A sample tissue-region instance-map is shown in Fig. 3.

To extract feature representations of tissue regions, we follow a two-step procedure: first, we extract CNN-based features for oversegmented superpixels, *i.e.*, patches of size $h \times w$ centered around the 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 connect in the TG topology. To this end, we construct a Region Adjacency Graph (Potjer 1996) where an edge is built between adjacent tissue region. The topology is presented by a binary adjacency matrix $E_{TG} \in \mathbb{R}^{|V_{TG}| \times |V_{TG}|}$. Formally, a TG representation is formulated as $G_{TG} := \{V_{TG}, E_{TG}, H_{TG}\}$.

4.2.3. 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 Hier-Archical Cell-to-Tissue (HACT) graph representation to jointly represent 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 $A_{CG \to TG} \in \mathbb{R}^{|V_{CG}| \times |V_{TG}|}$ that utilizes the relative spatial distributions of nuclei with respect to tissue regions. For the ith nucleus and j^{th} tissue region, the corresponding assignment is given as,

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

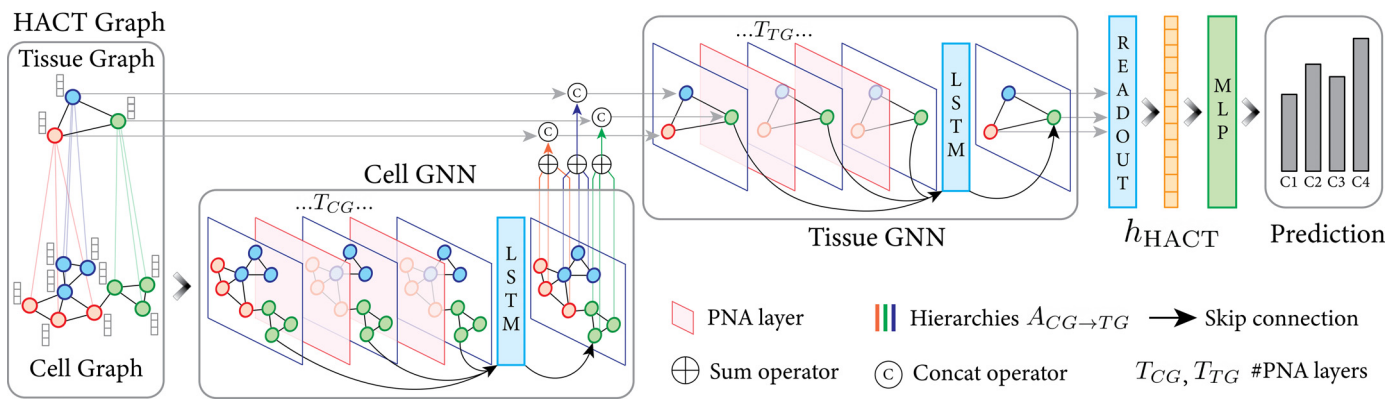

Fig. 4. Overview of the proposed HACT-Net architecture. The network processes an input HACT graph representation in a hierarchical manner, from fine cell-level to coarse tissue-region level, to obtain a contextualized graph embedding, and consequently classify the input graph. (Figure is best viewed in color).

Cell-to-tissue hierarchies for a tissue region are presented in Fig. 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 $G_{\text{HACT}} := \{G_{\text{CG}}, G_{\text{TG}}, A_{\text{CG}\to\text{TG}}\}$.

4.3. Graph learning

The HACT graph for a TRoI is processed by a hierarchical GNN to map TRoI composition to TRoI subtype. To this end, we propose HierArchical Cell-to-Tissue Network (HACT-Net), a hierarchical GNN architecture shown in Fig. 4.

4.3.1. HACT-Net Architecture & learning

HACT-Net intakes 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) categorizes h_{HACT}, *e.g.*, to a cancer subtype. Formally, HACT-Net consists of two GNNs, *i.e.*, 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 PNA layers (Corso et al., 2020).

First, CG − GNN intakes $G_{\text{CG}} := \{V_{\text{CG}}, E_{\text{CG}}, H_{\text{CG}}\}$, and applies T_{CG} PNA layers to build contextualized cell-node embeddings, inline with Eq. (1). The node embeddings $h^{(t)}(v), \forall v \in V_{\text{CG}}$ are iteratively updated as,

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

where $t = 0, \ldots, T_{\text{CG}}$ is the iteration index. As shown in Fig. 1, for a node v, first, the set of neighboring node embeddings $\{h_{CG}^{(t)}(u)\}, \forall u \in \mathcal{N}_{\text{CG}}(v)$ is concatenated with $h_{CG}^{(t)}(v)$, and processed by M_{CG}^{t}, a MLP, to produce a set of neighborhood-aware embeddings. Then, multiple aggregators with degree-scalers denoted by $\oplus$ operate on the set of MLP embeddings to extract a set of multivariate information that express the neighborhood distribution of node v. Finally, the set of information is concatenated to produce the aggregated message $a_{CG}^{(t+1)}(v)$ for node v. Afterwards, $a_{CG}^{(t+1)}(v)$ and $h_{CG}^{(t)}(v)$ are concatenated and processed by U_{CG}^{t}, a MLP, to update the node embedding, *i.e.*, $h_{CG}^{(t+1)}(v)$. Details of $\oplus$ is presented as,

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

where I is identity matrix, S is degree-scaler matrix, D is node degree matrix, δ is normalization constant, α is scaling variable, and V_{train} is nodes in the training dataset. $[I, \mathcal{S}(D,\alpha=1), \mathcal{S}(D,\alpha=-1)]$ and $\left[\mu, \sigma, \max, \min\right]$ denote the list of scalers and the list of aggregators, respectively. The aggregators compute statistics on neighboring multiset of nodes, and the injective scalers discriminate between the multisets of various sizes. $\alpha = \{-1, 0, 1\}$ controls the attenuation, no scaling, or amplification of the scaling, respectively. $\otimes$ denotes tensor product between scalers and aggregators, and produces twelve operations that extract the set of multivariate information. The schematic diagram of a PNA layer is shown in Fig. 1.

After T_{CG} PNA layers, an LSTM-based jumping knowledge technique (Xu et al., 2018) is employed to adapt to different CG subgraph structures, *i.e.*,

$$h_{\text{CG}}^{(T_{\text{CG}}+1)}(v) = \text{LSTM}\left(\left\{h_{\text{CG}}^{(t)}(v) \;\middle|\; t = 1, \ldots, T_{\text{CG}}\right\}\right) \tag{6}$$

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

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

where CONCAT denotes concatenation and $\mathcal{M}(w) := \{v \in V_{\text{CG}} \mid A_{\text{CG}\to\text{TG}}(v,w) = 1\}$ is the set of nodes in G_{CG} mapping to a node $w \in V_{\text{TG}}$. Analogous to Eq. (4), G_{TG} is processed by TG − GNN to compute tissue-node embeddings $h_{\text{TG}}^{(t)}(w), \forall w \in V_{\text{TG}}$. At $t = T_{\text{TG}}$, the embedding of each tissue-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 via an LSTM-based jumping knowledge technique to combine the intermediate tissue-node embeddings. Finally, the graph-level embedding h_{HACT} is produced by summing all the tissue-node embeddings. An MLP and a softmax operation follows to map h_{HACT} to respective TRoI label. HACT-Net is trained end-to-end by minimizing the cross-entropy loss between the softmax output and the ground-truth TRoI label.

Following Dwivedi et al. (2020), 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.

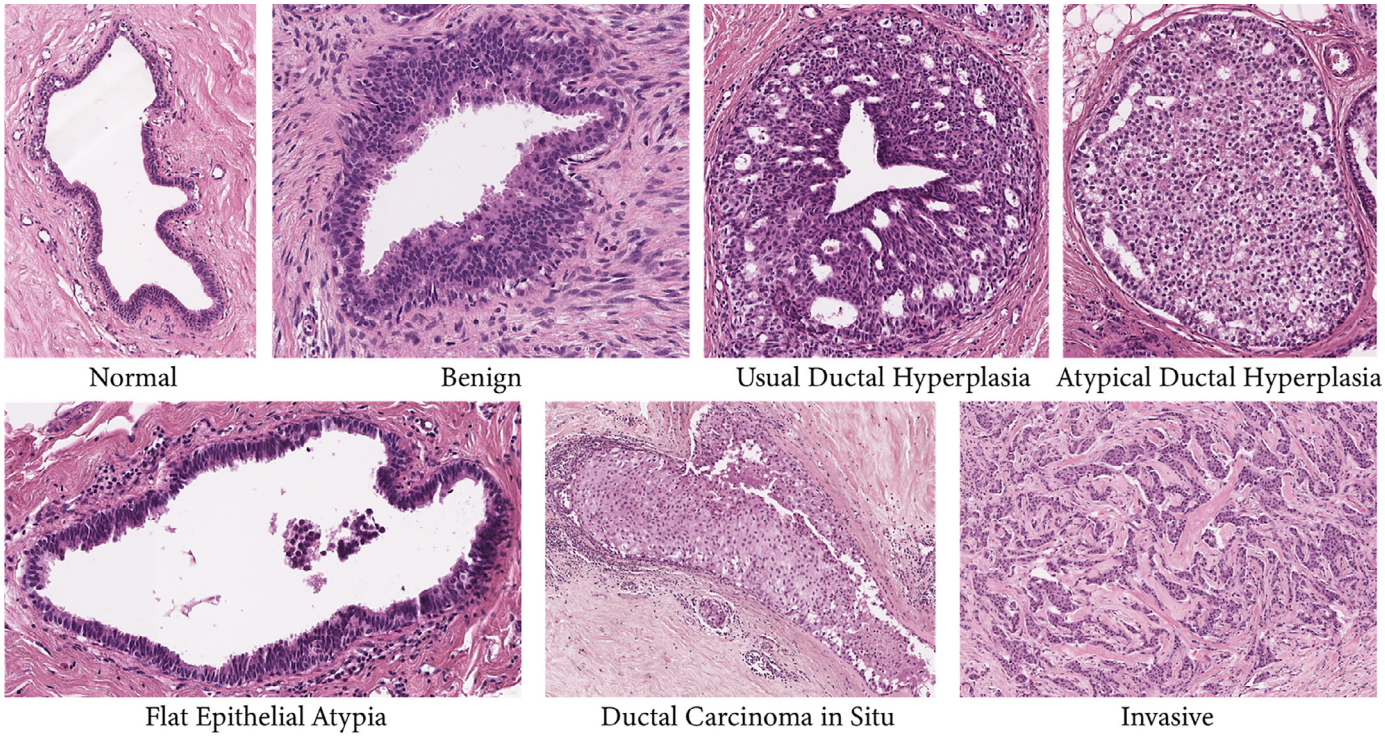

Fig. 5. Samples of class-wise tumor regions-of-interest in BRACS dataset. (Figure is best viewed in color).

5. Datasets

BRACS dataset: As part of this work, we introduce a new dataset termed as BReAst Cancer Subtyping (**BRACS**). It contains 4391 TRoIs from 325 H&E breast carcinoma WSIs. The WSIs were selected from the archives of the Department of Pathology at National Cancer Institute- IRCCS-Fondazione Pascale, Naples, Italy. They are scanned with an Aperio AT2 scanner at 0.25 μm/pixel resolution. The TRoIs were selected and annotated using QuPath (Bankhead et al., 2017) as being Normal, Benign, Usual ductal hyperplasia (UDH), Atypical Ductal Hyperplasia (ADH), Flat Epithelial Atypia (FEA), Ductal Carcinoma In Situ (DCIS), and Invasive. Fig. 5 presents sample TRoIs from all cancer subtypes in BRACS. Each TRoI was first annotated independently by three pathologists. TRoIs with disagreement were further discussed and annotated by the consensus. Note that the pathologists used the entire WSI context during annotation. Fig. 6 presents some DCIS samples in BRACS dataset, and highlight the included appearance variability. Such TRoI variability is typical in practice, and were included in BRACS to mimic the real world diagnosis. It ensures a realistic and representative evaluation set, with results readily applicable in the field.

Table 1 presents category-wise statistics of the TRoIs in BRACS. The statistics demonstrate a high variation in TRoI dimensions. We also include the statistics for the CG and TG representations constructed by our framework, which indicate a large variation in the size of the entity-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, *i.e.*, the Grand Challenge on BreAst Cancer Histology images **BACH** (Aresta et al., 2019). It 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 × 0.42 μm. Notably, BRACS presents three major advantages over BACH:

- **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 complex precancerous 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 BRACS in terms of TRoI appearances and dimensions is clinically more representative, and corresponds to a more realistic scenario of breast cancer subtyping.

6. Results

In this section, we comparatively assess the proposed method for breast cancer subtyping. First, we introduce state-of-the-art CNN and GNN baselines, and their implementation schemes. Second, we conduct ablations on BRACS to examine the impact of various components in our framework. Third, we evaluate the classification performance of our method and compare with the baselines, on BRACS and BACH datasets for different classification settings. Finally, we include a comparison of HACT-Net with three independent expert-pathologists.

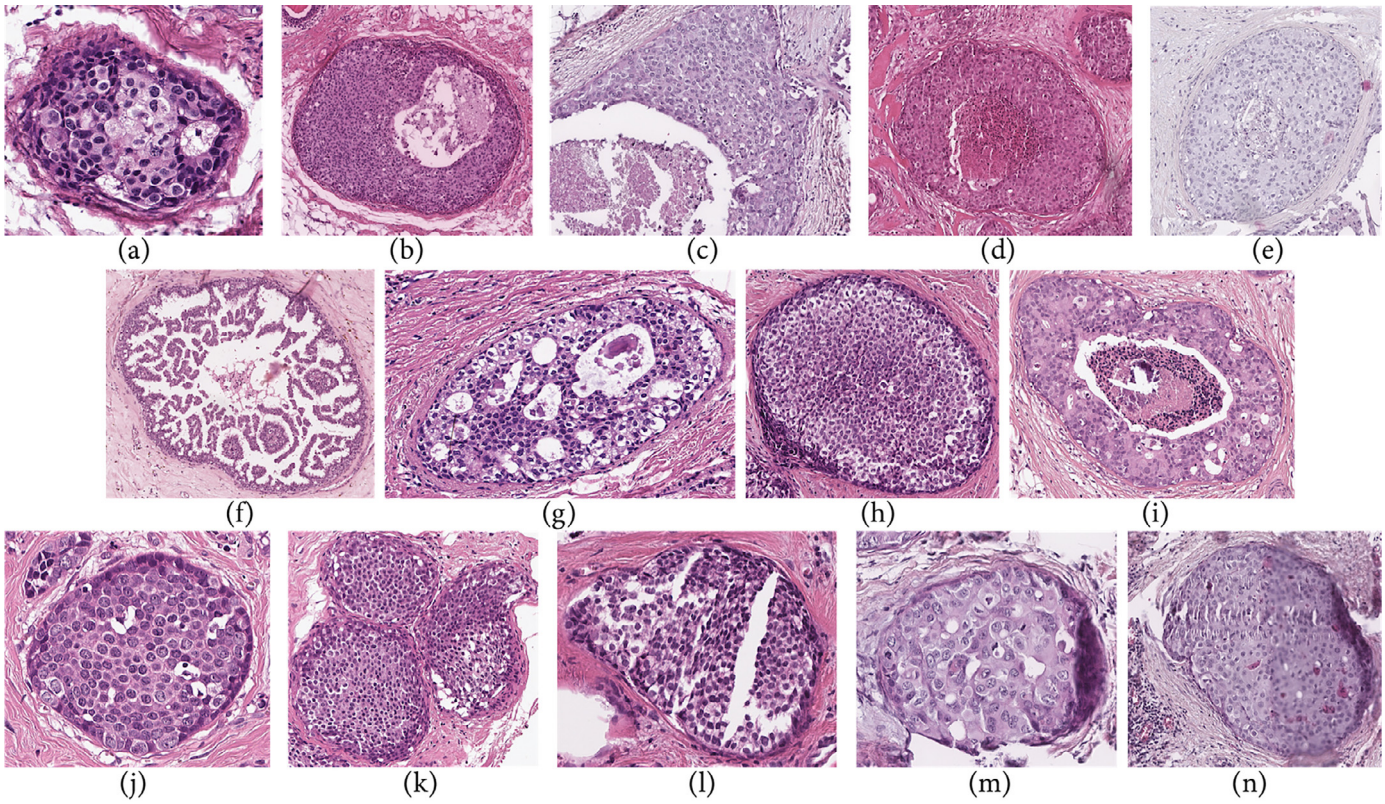

Fig. 6. Overview of the variability for DCIS category in BRACS. The samples depict variations in, (a, b, c) tumor size, (d, e) staining appearance, sub-patterns: (f) low-grade Papillary, (g) moderate-grade Cribriform, (h, i) high-grade Solid and Comedo, (j, k) number of glandular regions per TRoI, and artifacts due to tissue and slide preparation: (l) tissue-folding or tear, (m) ink stain, (n) blur. Similar variability also persists in other categories in BRACS. (Figure is best viewed in color).

Table 1
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

6.1. 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, *i.e.*, 10×, 20×, and 40×, denoted herein as CNN(10×), CNN(20×), and CNN(40×), using the same network architecture and training scheme. 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 Sirinukunwattana et al. (2018), and patch-wise predictions are aggregated using the Agg-Penultimate strategy by Mercan et al. (2019a). We use transfer learning with a ResNet-50 architecture, pre-trained on ImageNet, as the CNN backbone. Following feature extraction by ResNet-50, a two-layer MLP with 128 channels classifies the patches. To improve the classification, the ResNet-50 parameters are fine-tuned. Adam optimizer (Kingma and Ba 2015) with 10^{-3} learning rate, a batch size of 16, and a dropout of 0.2 is 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 Sirinukunwattana et al. (2018). We operate at two settings, *i.e.*, (10×+20×) and (10×+20×+40×), and denote by prepending Multi-scale CNN in front of each. The patch-wise predictions are aggregated using the Agg-Penultimate strategy by Mercan et al. (2019a). On the concatenated features from

the LSTM, we use 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 Zhou et al. (2019), 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 4.2.1. We initialize the CG nodes with hand-crafted features, employ the Adaptive GraphSage-based CGC-Net architecture, and follow the training strategy proposed by Zhou et al. (2019).

• **Patch-GNN** implements the method proposed by Aygunes et al. (2020), which is the state-of-the-art GNN method for classifying patch-graph representations of TRoIs. It incorporates local inter-patch context through a GNN to construct a graph-level features, which is then processed by an MLP to classify the TRoIs. We experiment with Patch-GNN at three scales, *i.e.*, 10×, 20×, and 40×, denoted herein as Patch−GNN(10×), Patch−GNN(20×), and Patch−GNN(40×). 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 Aygunes et al. (2020).

• **CG-GNN** is provided as a standalone CG-based learning baseline, to compare with our proposed hierarchical learning. CG − GNN uses PNA layers, an LSTM-based jumping knowledge, sum readout, and a two-layer MLP classifier. We follow the CG representation strategy as described in Section 4.2.1.

• **TG-GNN** is provided as a standalone TG-based learning baseline, to compare with our proposed hierarchical learning. TG − GNN uses the same architecture as the CG − GNN, with the node features directly initialized by H_{TG} instead of Eq. (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.*, $h_{\text{CONCAT}} = \text{CONCAT}(h_{\text{CG}}, h_{\text{TG}})$. Finally, a two-layer MLP classifies h_{CONCAT} into a cancer subtype.

6.2. Implementation

Graph representations: CG representations (Section 4.2.1) 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 4.2.2) 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 (Kingma and Ba 2015) 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 1), we evaluate the classification performance using weighted F1-score, an average weighted by the number of true instances for each class. The best weighted F1-scores on the validation set is used as the model selection criteria during the training of each 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.

Table 2
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**

Computational resources: All the experiments were conducted using PyTorch (Paszke et al., 2019) and Deep Graph Library (DGL) (Wang et al., 2019), on NVIDIA Tesla P100 GPUs and POWER8 processors.

6.3. Ablation studies

We conduct ablation to evaluate the impact of three major components of our methodology on TRoI classification performance, *i.e.*, i) node feature initialization, ii) GNN layer type, and iii) jumping knowledge technique. Each component is analyzed individually, while fixing the others. Ablations are performed on BRACS for classifying the TRoIs into 7-classes.

6.3.1. Impact of node feature initialization

The performance of GNNs eminently rely on the initial node features (Kipf and Welling 2017). In our context, we analyze the impact of initial morphological features of the nodes with the following three feature initialization schemes:

• **No morphological features:** The nodes of an entity-graph are initialized with only the spatial features. Experiments with this setting demonstrate the impact of standalone graph topology on the classification performance.

• **Hand-crafted morphological features:** The entity-graph nodes are initialized with hand-crafted morphological features as suggested by Zhou et al. (2019), *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 entity-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 2 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, including morphological features significantly improves the classification. The superiority of graphs with CNN-based

Table 3
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 4
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**

morphological features indicate the richness of morphological information acquired by CNNs, compared to hand-crafted measures.

6.3.2. *Impact of* GNN *layer type*

We investigate the impact of two state-of-the-art GNN layers, *i.e.*, GIN and PNA (Fig. 1), on the classification performance. The experiments use CNN-based node feature initialization and LSTM-based jumping knowledge. Results in Table 3 demonstrate that GNNs with PNA layers outperform GNNs with GIN layers, for all the four GNN constructions.

6.3.3. *Impact of jumping knowledge technique*

To investigate the impact of jumping knowledge, we experiment with three settings: no jumping knowledge, CONCAT-based, and LSTM-based. LSTM-based technique follows Eq. (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 4 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.

6.3.4. *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.

6.4. *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 5 tabulates the classification performance of the compared methods.

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

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, they outperform each other for Normal, Benign, UDH, ADH, and FEA categories, displaying the importance of complementary information captured by standalone TG and CG representations. Further, both HACT-Net and CONCAT − GNN provide overall superior performance compared to all CNN and GNN baselines. HACT-Net significantly outperforms CONCAT − GNN showing the significance of 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 using complementary information from CG and TG. Such complementary information is indeed best utilized by HACT-Net, with high per-class and overall classification performance. Though HACT-Net achieves the third best result for the UDH category, it uses the complementarity of CG and TG to provide better classification than TG − GNN. Moreover, the misclassified UDH samples are predominantly categorized as Benign due to the expected ambiguity between Benign and UDH classes. All the proposed GNNs often outperform all CNN baselines, establishing the potential of entity-based analysis.

Fig. 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. Fig. 8 presets 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 pairwise classes bear high pathological ambiguity and are diagnostically very challenging.

• **Setting 2: 4-class classification:** This setting categorizes TRoIs into 4-classes as per cancer risk: Normal, Non-cancerous (Benign + UDH), Precancerous (ADH + FEA), and Cancerous (DCIS + Invasive). Classification performance of CNN and GNN baselines, and HACT-Net are presented in Table 6. 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 over the single-scales. 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 the baselines. HACT-Net produces the best overall performance, with the best classification performance for Normal, Precancerous, and Cancerous categories. To highlight, HACT-Net achieves

Table 5
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	45.00±4.97	24.00±2.83	47.00±4.32	53.33±2.62	86.67±2.64	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**	39.99±3.56	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	72.15±1.35	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	60.83±3.71	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**

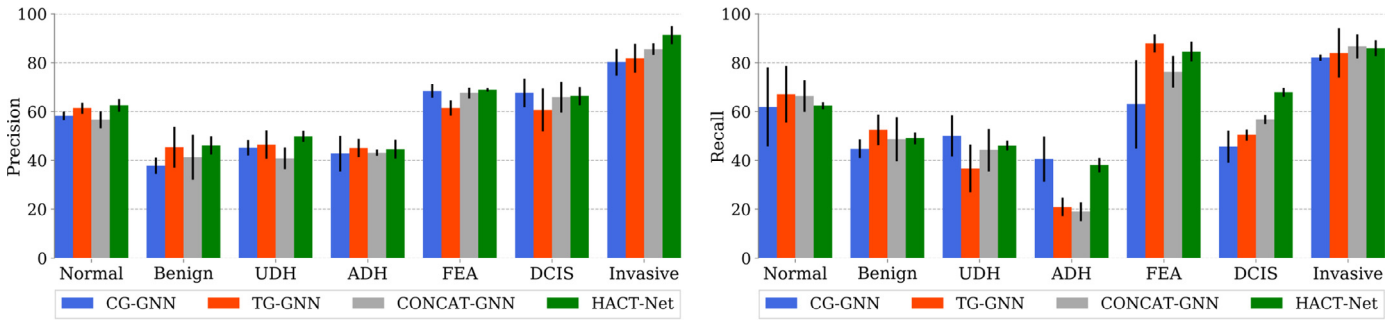

Fig. 7. Mean and standard deviation of per-class precision and recall for 7-class classification with HACT-Net. (Figure is best viewed in color).

Table 6
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	56.55±3.70	61.53±3.03	84.47±0.87	66.10±2.58
	TG-GNN	52.96±6.81	56.52±2.85	64.36±1.05	82.21±0.78	66.24±1.11
	CONCAT-GNN	54.54±1.64	**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**

≈ 66% F1-score for the diagnostically challenging Precancerous category.

• **Setting 3: Binary classifications:** In this setting, we replicate the typical decision process of a pathologist for breast cancer subtyping which follows a diagnostic decision tree as presented in Fig. 9. It is inspired by the classification scheme presented by Mercan et al. (2018). Note that such individual binary decisions are less constrained compared to multi-class classification, thus allows for better discrimination between a selected pair of classes. The binary classifiers can assist pathologists in categorizing ambiguous cases at different bifurcations of the decision tree. Table 7 presents the results for six individual binary classifications, 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.

6.4.1. 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 from additional independent pathologists. For such comparison with domain experts, we follow the evaluation protocol in Elmore et al. (2015). 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 prac-

Table 7
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**	75.71±1.68	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	76.12±0.99	80.67±0.22	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	88.88±3.86	82.61±2.15
	HACT-Net (Proposed)	**96.32±0.64**	83.63±0.73	**76.84±0.68**	**77.66±0.37**	**81.11±0.72**	**89.35±0.26**	**84.15±0.60**

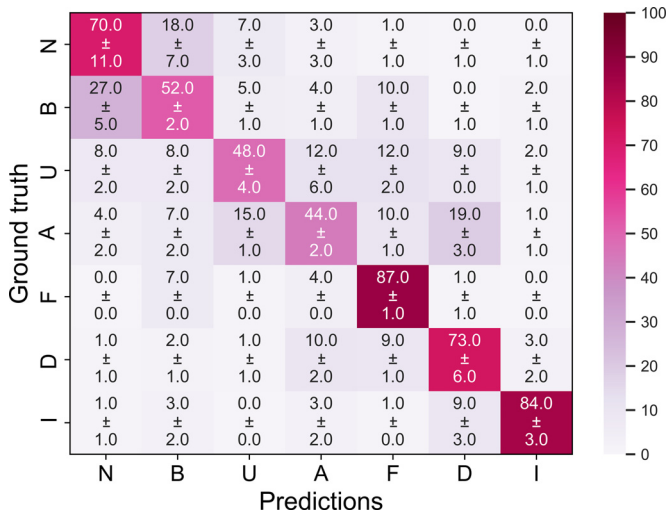

Fig. 8. Mean and standard deviation of row-normalized 7-class confusion matrix for HACT-Net.

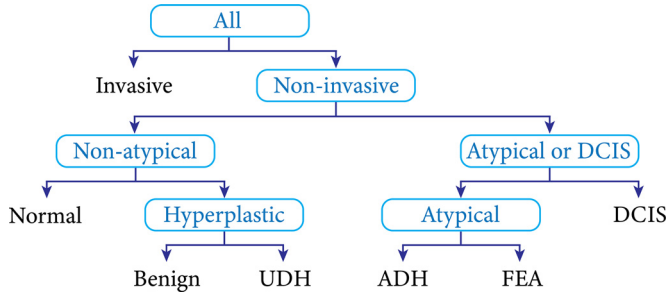

Fig. 9. Decision tree used by pathologists for breast cancer diagnosis. The 7-class classification is simplified to a series of binary decision tasks, through which the diagnosis becomes more and more specific until the leaves, *i.e.*, the 7 diagnostic decision classes, are reached.

tice 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.

The independent pathologists' annotations are compared to the ground truth, with the results shown in Table 8. We present per-class F1-scores, overall weighted F1-score, and overall weighted accuracy for each pathologist. We also include the aggregated statistics of the three pathologists for benchmarking HACT-Net with domain experts. Table 8 indicates that HACT-Net outperforms the domain experts in distinguishing TRoIs of diagnostically challenging classes, *i.e.*, atypia and hyperplasia, while yielding comparable performance for the normal and cancerous categories. Per-class standard deviations of pathologists' statistics show 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 for the 7-class classification.

To benchmark the BRACS dataset with respect to the dataset by Elmore et al. (2015), 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 Elmore et al. (2015) consists of 240 breast biopsy slides, while BRACS consists of 626 TRoI images. For the dataset by Elmore et al. (2015), 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 9 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.

6.4.2. Computational time analysis

We report computation time for processing a tumor RoI of size 1000 × 1000 pixels on a single-core POWER8 processor combined with an NVIDIA P100 GPU. Stain normalization with the Macenko method takes 0.8 seconds (CPU-only), CG generation 2.51 seconds, and TG generation 4.14 seconds. Thus, the overall computational time for transforming the RoI into HACT representation is 7.92 seconds. The superpixel extraction step can be further optimized by using fast GPU implementations, *e.g.*, as proposed by Jampani et al. (2018). Provided the HACT representation, HACT-Net renders near real-time inference by requiring 34.11 milliseconds. Additional run-time analysis is presented by Jaume et al. (2021a).

6.5. 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

Table 8
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**

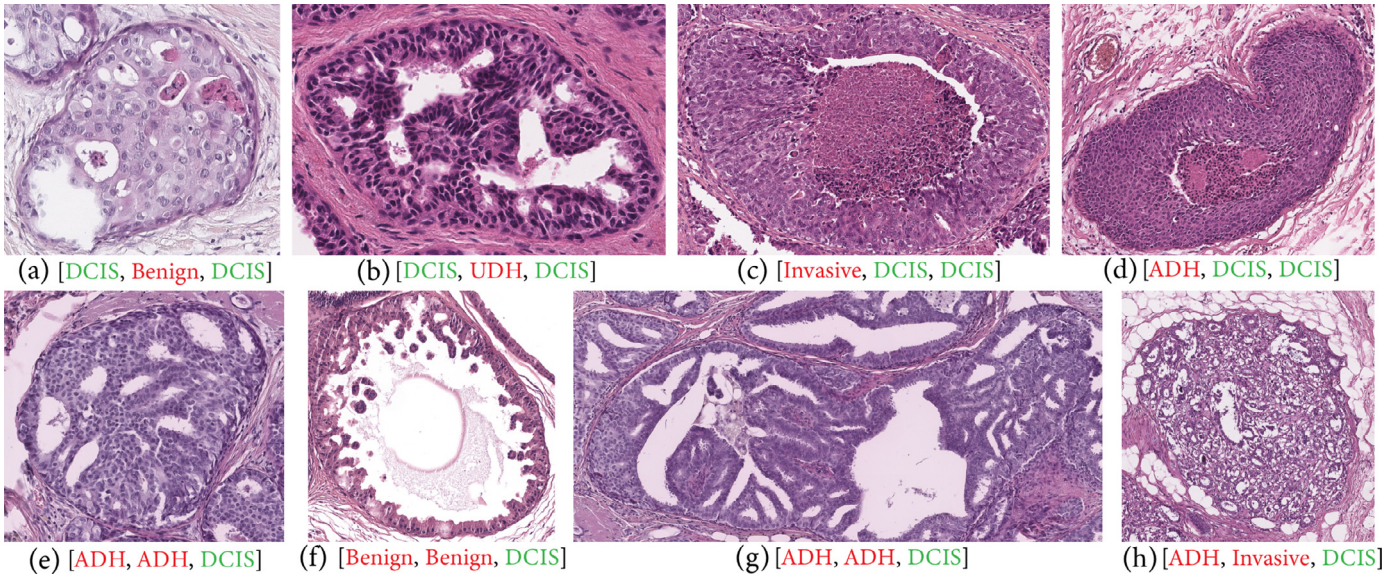

Fig. 10. Qualitative comparison of CG – GNN, TG – GNN, and HACT-Net for 7-class classification. Predictions by the classifiers are noted below each example. Red and Green denote incorrect and correct classification, respectively. (a,b) TRoIs which TG – GNN misclassifies, while CG – GNN and HACT-Net classify 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).

Table 9
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

Table 10
Accuracy of 4-class breast cancer subtyping in BACH dataset. Results are expressed in %.

	Methods	Accuracy
Ensemble networks (Aresta et al., 2018; Aresta et al., 2019)	Wang et al. (2019)	95.00
	Marami et al. (2018)	94.00
	Yang et al. (2019)	93.00
	Chennamsetty et al. (2018)	87.00
	Kwok et al. (2018)	87.00
	Brancati et al. (2018)	86.00
Single network	HACT-Net	91.00

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 6.2 are employed 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 10. Our predictions have been evaluated independently by the organizers of the BACH challenge, ensuring a fair comparison. 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 employ an ensemble strategy with multiple networks during inference.

6.6. Qualitative analysis

Qualitative assessment of a few TRoIs from the BRACS dataset using HACT-Net, CG – GNN, and TG – GNN is presented in Fig. 10. In Fig. 11, we use GraphGradCAM (Pope et al., 2019; Jaume et al., 2021b), a post-hoc gradient based feature attribution technique, to highlight the nuclei and tissue-region nodes in CG and TG, respectively, which 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 epithelium and necrotic regions in TG, while ignoring the less important stroma and lumen, cf. Figs. 11(b,h). Further, within the relevant tissue regions, HACT-Net focuses on a subset of tumorous epithelial nuclei in CG, as shown in Figs. 11(c,i). Interestingly, we observe in Figs. 11(h,i) that HACT-Net uses complementary information from the necrotic region captured by TG, but not by CG. Similar observations of HACT-Net considering the diagnostically relevant regions can be made for FEA and Benign examples shown in Figs 11(d-f&j-l). Noticeably, such feature attribution analysis of GNNs localizes and highlights the focus of deep networks in the given entity-paradigm, which is both more interpretable and more

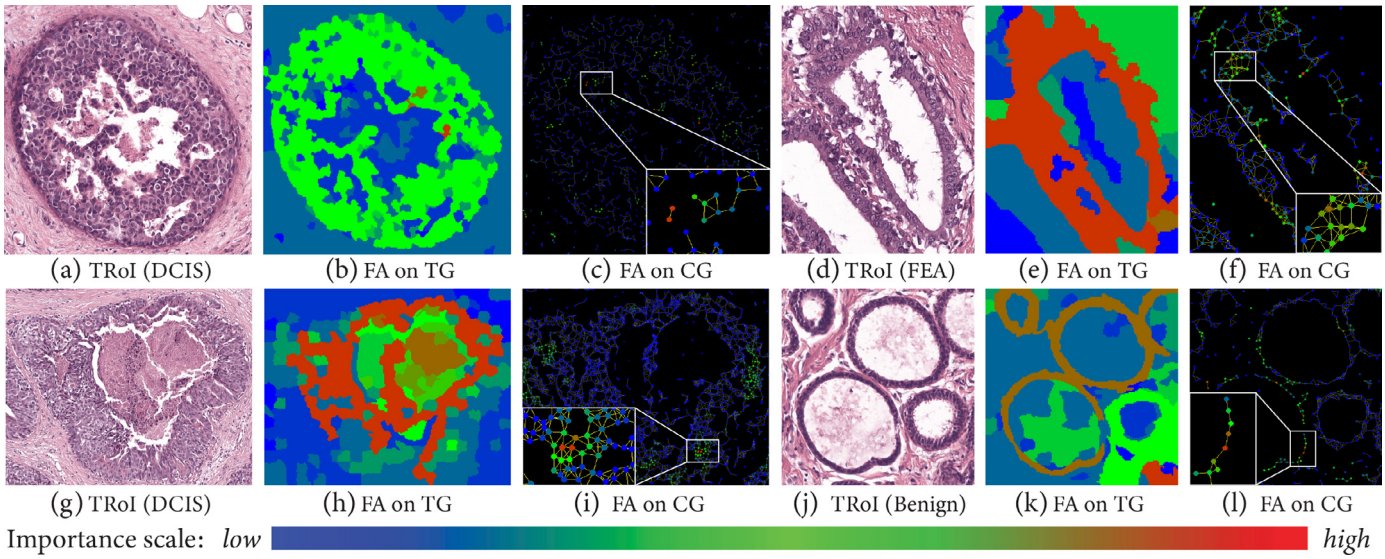

Fig. 11. Feature attribution (FA) maps of HACT-Net on TG and CG for four sample TRoIs for 7-class classification: Sample TRoIs of (a,g) DCIS, (d) FEA, and (j) Benign classes, with their corresponding feature attribution maps on (b,h,e,k) TG and (c,i,f,l) CG. (Figure is best viewed in color).

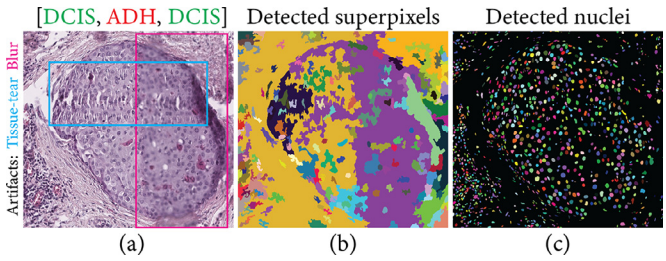

Fig. 12. (a) A DCIS sample including tissue-tear and blur artifacts. (b) Detected superpixels. (c) Detected nuclei. The classifications by CG − GNN, TG − GNN and HACT-Net are indicated, where Red and Green denote incorrect and correct classification.

explainable compared to feature attribution strategies in a pixel-paradigm (Jaume et al., 2020, 2021b). Interestingly, we also analyze the impact of tissue or slide preparation artifacts on the model performance. In Fig. 12, we present a DCIS image with tissue-tear and blur artifacts. We observe that the detected superpixels do not aptly depict the tissue in the blur region, and consequently the TG − GNN using standalone TG misclassifies it. However, the nuclei detection is less impacted by the artifact, which allows the CG to appropriately encode the cell microenvironment and correctly classify the sample. To highlight, HACT-Net utilizing the complementary information from both CG and TG compensates for the issue in TG, and correctly identifies the subtype.

7. 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 these issues. 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 tissue subtypes. 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 present 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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Pushpak Pati: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Guillaume Jaume:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Antonio Foncubierta-Rodríguez:** Resources, Data curation, Writing – review & editing. **Florinda Feroce:** Resources, Data curation. **Anna Maria Anniciello:** Resources, Data curation. **Giosue Scognamiglio:** Resources, Data curation. **Nadia Brancati:** Resources, Data curation. **Maryse Fiche:** Resources, Data curation. **Estelle Dubruc:** Resources, Data curation. **Daniel Riccio:** Resources, Data curation. **Maurizio Di Bonito:** Resources, Data curation. **Giuseppe De Pietro:** Resources, Data curation. **Gerardo Botti:** Resources, Data curation. **Jean-Philippe Thiran:** Supervision. **Maria Frucci:** Supervision. **Orcun Goksel:** Writing – review & editing, Supervision. **Maria Gabrani:** Supervision, Funding acquisition.

References

Achanta, R., et al., 2012. Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell., 34 (11), 2274–2282.

Adnan, M., et al., 2020. Representation learning of histopathology images using graph neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 4254–4261.

Allemani, C., et al., 2015. 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). The Lancet, 385 (9972), 977–1010.

Anand, D., et al., 2020. Histographs: graphs in histopathology. SPIE Medical Imaging.

Anklin, V., et al., 2021. Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs. Medical Image Computing and Computer Assisted Intervention (MICCAI).

Araujo, T., et al., 2005. Classification of breast cancer histology images using convolutional neural networks. PLoS One, 12 (6).

Aresta, G., et al., 2018. Iciar 2018 grand challenge on breast cancer histology images.

Aresta, G., et al., 2019. Bach: grand challenge on breast cancer histology images. Med. Image Anal., 56, 122–139.

Aygunes, B., et al., 2020. Graph convolutional networks for region of interest classification in breast histopathology. SPIE Medical Imaging.

Bankhead, P., et al., 2017. Qupath: open source software for digital pathology image analysis. Sci. Rep., 7 (1), 1–7.

Bardou, D., et al., 2018. Classification of breast cancer based on histology images using convolutional neural networks. IEEE Access, 6, 24680–24693.

Bejnordi, B., et al., 2017. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. J. Med. Imaging, 4 (4).

Bejnordi, B., et al., 2019. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA, 318 (22), 21992210.

Binder, T., et al., 2019. Multi-organ gland segmentation using deep learning. Front. Med.

Brancati, N., Frucci, M., Riccio, D., 2018. Multi-classification of breast cancer histology images by using a fine-tuning strategy. In: International Conference Image Analysis and Recognition (ICIAR), pp. 771–778.

Campanella, G., et al., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med., 25 (8), 130109.

Chen, R., et al., 2020. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans. Med. Imag.

Chennamsetty, S., Safwan, M., Alex, V., 2018. Classification of breast cancer histology image using ensemble of pre-trained neural networks. In: International Conference Image Analysis and Recognition (ICIAR), pp. 804–811.

Corso, G., et al., 2020. Principal neighbourhood aggregation for graph nets. Neural Information Processing Systems (NeurIPS).

Dehmamy, N., et al., 2019. Understanding the representation power of graph neural networks in learning graph topology. In: Neural Information Processing Systems (NeurIPS), pp. 15413–15423.

Demir, C., Yener, B., Gultekin, S., 2004. The cell graphs of cancer. Bioinformatics, 20, 145–151.

Deng, J., et al., 2009. Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255.

Deng, S., et al., 2020. Deep learning in digital pathology image analysis: asurvey. Front. Med.

Dwivedi, V., et al., 2020. Benchmarking graph neural networks. arXiv:2003.00982.

Elmore, J., et al., 2015. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA, 313 (11), 1122–1132.

Francis, K., Palsson, B., 1997. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. Proc Natl Acad Sci, 94 (23), 12258–12262.

Gilmer, J., et al., 2017. Neural message passing for quantum chemistry. In: International Conference on Machine Learning (ICML), pp. 1263–1272.

Gomes, D., et al., 2014. 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. Diagn Pathol, 9 (121).

Graham, S., et al., 2019. Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal., 58.

Graham, S., et al., 2019. Mild-net: minimal information loss dilated network for gland instance segmentation in colon histology images. Med. Image Anal., 52, 199–211.

Hagele, M., et al., 2020. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. Sci. Rep., 10.

Hamilton, W., et al., 2017. Inductive representation learning on large graphs. In: Neural Information Processing Systems (NeurIPS), pp. 1024–1034.

He, K., et al., 2016. Deep residual learning for image recognition. In: IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778.

Jampani, V., et al., 2018. Superpixel sampling networks. In: European Conference on Computer Vision (ECCV).

Jaume, G., et al., 2019. edgnn: a simple and powerful gnn for directed labeled graphs. In: International Conference on Learning Representations (ICLR) Workshop on Representation Learning on Graphs and Manifolds.

Jaume, G., et al., 2020. Towards explainable graph representations in digital pathology. In: International Conference on Machine Learning (ICML), Workshop on Computational Biology.

Jaume, G., et al., 2021. Histocartography: A toolkit for graph analytics in digital pathology. arXiv:2107.10073.

Jaume, G., et al., 2021. Quantifying explainers of graph neural networks in computational pathology. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Javed, S., et al., 2020. Cellular community detection for tissue phenotyping in colorectal cancer histology images. Med. Image Anal., 63.

Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR).

Kipf, T., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. In: Int. Conf. Learning Representations (ICLR).

Komura, D., Ishikawa, S., 2018. Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J., 34–42.

Kumar, N., et al., 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imag., 36 (7), 1550–1560.

Litjens, G., et al., 2017. A survey on deep learning in medical image analysis. Med. Image Anal., 42, 60–88.

Lu, M., et al., 2021. Data efficient and weakly supervised computational pathology on whole slide images. Nat Biomed Eng.

Macenko, M., et al., 2009. A method for normalizing histology slides for quantitative analysis. In: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1107–1110.

Madabhushi, A., Lee, G., 2016. Image analysis and machine learning in digital pathology: challenges and opportunities. Med. Image Anal., 33, 170–175.

Marami, B., et al., 2018. Ensemble network for region identification in breast histopathology slides. In: International Conference Image Analysis and Recognition (ICIAR), pp. 861–868.

Mehta, S., et al., 2018. Learning to segment breast biopsy whole slide images. In: IEEE Winter Conference on Applications of Computer Vision (WACV).

Mercan, C., et al., 2019. From patch-level to roi-level deep feature representations for breast histopathology classification. SPIE Medical Imaging.

Mercan, E., et al., 2018. Automated diagnosis of breast cancer and pre-invasive lesions on digital whole slide images. In: International Conference on Pattern Recognition Applications and Methods (ICPRAM).

Mercan, E., et al., 2019. Assessment of machine learning of breast pathology structures for automated differentiation of breast cancer and high-risk proliferative lesions. JAMA Netw Open, 2.

Morris, C., et al., 2018. Weisfeiler and leman go neural: Higher-order graph neural networks. Association for the Advancement of Artificial Intelligence (AAAI), 33.

Mukhopadhyay, S., et al., 2017. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). Am J Surg Pathol, 42 (1), 39–52.

Parwani, A., 2019. Next generation diagnostic pathology: use of digital pathology and artificial intelligence tools to augment a pathological diagnosis. Diagn Pathol, 14 (138).

Paszke, A., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Neural Information Processing Systems (NeurIPS), pp. 8024–8035.

Pati, P., et al., 2018. Deep positive-unlabeled learning for region of interest localization in breast tissue images. SPIE Medical Imaging.

Pati, P., et al., 2020. Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification. Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on GRaphs in biomedicAl Image anaLysis.

Pati, P., et al., 2021. Reducing annotation effort in digital pathology: aco-representation learning framework for classification tasks. Med. Image Anal., 67.

Pinckaers, H., van Ginneken, B., Litjens, G., 2020. Streaming convolutional neural networks for end-to-end learning with multi-megapixel images. IEEE Trans. Med. Imag., 39 (5), 1306–1315.

Pope, P., et al., 2019. Explainability methods for graph convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10764–10773.

Potjer, F., 1996. Region adjacency graphs and connected morphological operators. Mathematical Morphology and its Applications to Image and Signal Processing. Computational Imaging and Vision, 5, 11118.

Roy, K., et al., 2019. Patch-based system for classification of breast histology images using deep learning. Comput Med Imaging Graph, 71, 90–103.

Shaban, M., et al., 2020. Context-aware convolutional neural network for grading of colorectal cancer histology images. IEEE Trans. Med. Imag., 39, 2395–2405.

Sharma, H., et al., 2015. A review of graph-based methods for image analysis in digital histopathology. Diagn Pathol.

Sharma, H., et al., 2016. Cell nuclei attributed relational graphs for efficient representation and classification of gastric cancer in digital histopathology. SPIE Medical Imaging, 9791.

Sharma, H., et al., 2017. 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, pp. 61–66.

Siegel, R., Miller, K., Jemal, A., 2020. Cancer statistics, 2020. CA: A Cancer Journal for Clinicians, 70, 7–30.

Sirinukunwattana, K., et al., 2018. Improving whole slide segmentation through visual context - a systematic study. Medical Image Computing and Computer Assisted Intervention (MICCAI), 11071.

Spanhol, F., et al., 2016. A dataset for breast cancer histopathological image classification. IEEE Trans. Biomed. Eng., 63 (7), 1455–1462.

Srinidhi, C., et al., 2021. Deep neural network models for computational histopathology: A survey. Med. Image Anal., 67.

Stanisavljevic, M., et al., 2018. A fast and scalable pipeline for stain normalization of whole-slide images in histopathology. In: European Conference on Computer Vision (ECCV) Workshops.
Sung, H., et al., 2021. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. clinicians 1–41.
Sureka, M., et al., 2020. Visualization for histopathology images using graph convolutional neural networks. In: IEEE International Conference on Bioinformatics and Bioengineering (BIBE).
Tellez, D., et al., 2019. Neural image compression for gigapixel histopathology image analysis. IEEE Trans. Pattern Anal. Mach. Intell., 58.
Tellez, D., et al., 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Med. Image Anal., 58.
Velickovic, P., et al., 2018. Graph attention networks. In: International Conference on Learning Representations (ICLR).
Verma, R., et al., 2020. Multi-organ nuclei segmentation and classification challenge. IEEE Trans. Med. Imag.
Veta, M., et al., 2014. Breast cancer histopathology image analysis: areview. IEEE Trans. Biomed. Eng. 1400–1411.
Wang, M., et al., 2019. Deep graph library: Towards efficient and scalable deep learning on graphs. CoRR, abs/1909.01315.
Weisfeiler, B., Lehman, A.A., 1968. A reduction of a graph to a canonical form and an algebra arising during this reduction. In: Nauchno-Technicheskaya Informatsia, 2, pp. 12–16.
Xu, K., et al., 2018. Representation learning on graphs with jumping knowledge networks. In: International Conference on Machine Learning (ICML).
Xu, K., et al., 2019. How powerful are graph neural networks? In: International Conference on Learning Representations (ICLR).
Yan, R., et al., 2020. Breast cancer histopathological image classification using a hybrid deep neural network. Methods, 173, 52–60.
Yao, J., et al., 2021. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med. Image Anal., 65 (473).
Zhao, Y., et al., 2020. 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), pp. 4837–4846.
Zhou, Y., et al., 2019. CGC-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: IEEE International Conference on Computer Vision (ICCV) Workshops.
Zhu, X., et al., 2017. Wsisa: Making survival prediction from whole slide pathology images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).