Medical Image Analysis 90 (2023) 102936

Contents lists available at ScienceDirect

Medical Image Analysis

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

Multi-cell type and multi-level graph aggregation network for cancer grading in pathology images

Syed Farhan Abbas [a], Trinh Thi Le Vuong [a], Kyungeun Kim [b], Boram Song [b], Jin Tae Kwak [a,*]

[a] School of Electrical Engineering, Korea University, Seoul 02841, Republic of Korea
[b] Department of Pathology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul 03181, Republic of Korea

ARTICLE INFO

Keywords:
Cancer grading
Graph neural network
Cell graphs
Graph aggregation

ABSTRACT

In pathology, cancer grading is crucial for patient management and treatment. Recent deep learning methods, based upon convolutional neural networks (CNNs), have shown great potential for automated and accurate cancer diagnosis. However, these do not explicitly utilize tissue/cellular composition, and thus difficult to incorporate the existing knowledge of cancer pathology. In this study, we propose a multi-cell type and multi-level graph aggregation network (MMGA-Net) for cancer grading. Given a pathology image, MMGA-Net constructs multiple cell graphs at multiple levels to represent intra- and inter-cell type relationships and to incorporate global and local cell-to-cell interactions. In addition, it extracts tissue contextual information using a CNN. Then, the tissue and cellular information are fused to predict a cancer grade. The experimental results on two types of cancer datasets demonstrate the effectiveness of MMGA-Net, outperforming other competing models. The results also suggest that the information fusion of multiple cell types and multiple levels via graphs is critical for improved pathology image analysis.

1. Introduction

Cancer is the leading cause of death worldwide and accounted for 10 million deaths in 2020 (Sung et al., 2021). Early diagnosis and detection of cancer can improve survival rates and reduce death rates. However, traditional cancer pathology, which requires a manual examination of biopsied or resected tissue samples upon staining, suffers from low-throughput and large inter- and intra-observer variations (Elmore et al., 2015; Mahmood et al., 2020). The consistent increase in the incidence of cancer may contribute to the increase in diagnostic errors and a decrease in the quality of pathology services in clinics. Therefore, alternative methods that can improve the accuracy, throughput, and reliability of cancer pathology are needed.

In recent years, computational pathology has shown to be effective in processing and analyzing digitized pathology images (Niazi et al., 2019), holding great potential for facilitating improved cancer pathology today. Recent advances in computational pathology are largely attributable to the availability of large pathology image datasets (Bulten et al., 2022; Gamper et al., 2019) and deep learning, in particular convolutional neural networks (Vit). Nuclear feature extraction has been successfully applied to several pathology tasks such as nuclei segmentation (Doan et al., 2022; Kumar et al., 2017), tissue segmentation (Mehta et al., 2018a; Mercan et al., 2019), nuclei classification (Doan et al., 2022; Graham et al., 2019), and tumor detection (Bejnordi et al., 2017; Pati et al., 2021) and staging (Le Vuong et al., 2021; Mercan et al., 2019). Although CNNs have shown their ability to process and analyze pathology images, there exist several critical issues inherent to the nature of CNNs and pathology images. First, CNNs mostly operate on per image basis. Each image has a pre-determined, fixed size and fixed field-of-view, and thus the analysis of CNNs is, by and large, confined to the size and resolution of the image. Second, pathology images are enormous; for instance, a single whole-slide image (WSI), in general, possesses an order of gigapixels, which overwhelms the capacity of CNNs. For this reason, WSIs are often represented as a bag of smaller images (or patches), and the information from numerous images is aggregated to conduct WSI-level tasks. Third, the aggregation of information from numerous images is non-trivial. A majority voting scheme (Roy et al., 2019), multiple instance learning (Sudharshan et al., 2019), and recurrent neural networks (RNNs) (Yan et al., 2020) are often utilized, but these may ignore the original spatial location and relationship among images. Last, pathology assessment is performed based on various histological objects, including glands, cells, nuclei, etc., but the operation of CNNs mostly focuses on neighboring pixels, not such objects. This makes it hard to incorporate the existing knowledge of cancer pathology, leading to poor interpretability and explainability of CNNs and their decisions on pathology images.

* Corresponding author.
E-mail address: jkwak@korea.ac.kr (J.T. Kwak).

https://doi.org/10.1016/j.media.2023.102936
Received 16 November 2022; Received in revised form 30 May 2023; Accepted 16 August 2023
Available online 25 August 2023

In computational pathology, there exists another line of research based on graph theory that could resolve the shortcomings of the current CNN approaches. A pathology image can be described as a set of histological objects such as cells. Using the histological objects, we can form a graph in which nodes and edges characterize individual histological objects and the interaction/relationship among them, respectively. In this manner, we can exploit both cellular morphology and cell-to-cell interactions for tissue microenvironment and disease status, which is beneficial for pathology image analysis for several reasons. For example, it becomes easier to incorporate the existing pathology knowledge of cells and tissues as well as requires less computational cost, i.e., scalable to WSIs. It also enables the integration of the relationship among proximal/distant histological objects into a computational pathology system. Earlier works extracted and exploited graph-based hand-crafted features that are computed using the degree of a node, the number and ratio of nodes and edges, the length and weight of edges, and the shortest and longest path in a graph (Doyle et al., 2012; Nguyen et al., 2014). Recently, graph neural networks (GNNs) has been applied to several pathology image tasks, including cancer grading (Zhou et al., 2019) and tissue sub-typing (Jaume et al., 2021; Pati et al., 2022). In GNN-based methods, some construct graphs using tissue components (e.g., epithelium, stroma, lumen, etc.) (Anklin et al., 2021; Pati et al., 2022), but cell graphs are most widely used (Jaume et al., 2021; Pati et al., 2022). Although there exist several cell types (or tissue components) in a pathology image, none of the prior GNN-based methods sought to analyze them separately, largely ignoring the interaction among the cells of an individual type. Moreover, most of the GNN-based methods construct and utilize a single graph. Due to the abundance of cells, the interaction that is exploited and learned via graph learning tends to be limited to a local structure (Jaume et al., 2021). A hierarchical graph representation, so-called HACT, that utilizes both a tissue graph and a cell graph has been proposed for a comprehensive understanding of tissue structures (Pati et al., 2022). However, HACT may not properly reflect the overall tissue structure since it builds a tissue graph based on fragmented tissue segments; for instance, a single glandular structure is split into irregular tissue segments of varying sizes and shapes. Hence, a GNN method that can make use of both global and local tissue structures efficiently and effectively is needed to facilitate an improved graph representation of tissues. Herein, we propose a **M**ulti-cell type and **M**ulti-level **G**raph **A**ggregation **Net**etwork (MMGA-Net) for cancer grading in pathology images (Fig. 1). Given a pathology image, we construct multiple cell graphs by utilizing an advanced nuclei segmentation and classification algorithm: one graph for the entire cell types and a separate graph per cell type. Multi-cell type graphs are biologically inspired since each cell type has its own biological, histological, and functional characteristics. These cell graphs are also constructed at multiple levels to represent global and local cellular interactions and to mimic the pathology review process where tissue samples are examined under different magnifications from glandular formation and distribution to cell morphology. To represent the global cellular interaction, cell graphs are constructed for an entire pathology image. As for the local cellular interaction, cell graphs are constructed for small local regions, generating a multitude of local cell graphs. We aggregate these local graphs in a way that only important nodes and local regions are preserved irrespective of the order of nodes and local regions. In addition, we utilize a CNN to incorporate tissue-level information. MMGA-Net is composed of three major components. The first component learns the relationship among cells via GNNs. The second component extracts tissue contextual information using a CNN. The third component fuses tissue contextual information with cellular information from multi-level and multi-cell type graphs to conduct cancer grading. We evaluate MMGA-Net on two types of cancer datasets, including colorectal cancer and gastric cancer. The experimental results demonstrate that MMGA-Net outperforms other existing CNN- and GNN-based methods.

In summary, the major contributions of this paper are as follows:

- **Multi-cell type and multi-level graphs:** We introduce multi-cell type and multi-level graphs, which are inspired by tissue biology and pathology review process, for pathology image analysis. This allows us to study the global and local cellular interactions among the entire cells and cells of a specific type. To the best of our knowledge, this is the first GNN-based attempt to construct and utilize multi-cell type graphs in pathology image analysis.
- **Tissue and cellular information aggregation mechanism:** We introduce an efficient and effective three-stage aggregation mechanism for multi-cell type and multi-level graphs. In the first stage, we aggregate a set of local graphs via attention mechanism. In the second stage, we use deep sets (Zaheer et al., 2017) to discover feature representations that are invariant to the order of the cells in the global and local graphs. In the third stage, we exploit an entropy weighting scheme to aggregate multiple levels of tissue and cellular information and to make an aggregated prediction on a pathology image.
- **Evaluation on multi-organ data:** We systematically evaluate the proposed methodology for cancer grading on data from multiple organs such as colorectal and gastric cancer. To the best of our knowledge, this is the first attempt to apply a GNN-based method for gastric cancer grading. For both cancer types, the proposed method outperforms several CNN- and GNN-based methods.

2. Related work

2.1. Cancer classification in computational pathology

In computational pathology, there has been substantial effort to develop automated and robust tools for cancer classification. Nowadays, CNN-based methods are predominant and outperform other approaches built based upon hand-crafted features (LeCun et al., 2015; Van der Laak et al., 2021). Many CNN-based methods adopt a single-path or -stream architecture to conduct cancer classification in pathology images, including breast cancer (Araújo et al., 2017), prostate cancer (Arvaniti et al., 2018), and lung cancer (Coudray et al., 2018). To further improve the performance of cancer classification, more advanced CNN approaches have been proposed. For instance, Le Trinh et al. (2021) develops an end-to-end multi-scale CNN approach where multi-scale features are encoded as a binary code to discover multi-scale patterns and are used to conduct cancer grading in colorectal and prostate tissues. Bejnordi et al. (2017) utilizes a stacked CNN structure to incorporate large contextual information from tissues to conduct the classification of breast carcinomas.

Tellez et al. (2019) adopts neural image compression to compress a pathology image via unsupervised learning and employs a CNN to identify tumor metastasis in the breast and to classify 9 tissue types in the rectum. Multi-task learning that simultaneously learns multiple related tasks is another popular approach. For example, Mehta et al. (2018b) jointly conducts breast tissue segmentation and classification by using a modified U-Net. Le Vuong et al. (2021) formulates cancer grading as both categorical and ordinal classification problems and simultaneously conducts both classification tasks for cancer grading in prostate and colorectal tissues. Other approaches include the integration of multiple instance learning into Vit (Sudharshan et al., 2019) and the combination of Vit and RNNs (Yan et al., 2020) for cancer classification in pathology image analysis.

2.2. Graph representation and deep learning in computational pathology

Graph representation of a pathology image permits a histological object-based analysis that gives distinct advantages over pixel-based analysis utilized by Vit (Pati et al., 2022). It has shown to be promising in several applications in computational pathology, including cancer classification (Anand et al., 2020; Pati et al., 2022; Zhou et al., 2019), survival prediction (Chen et al., 2020), and tissue segmentation (Anklin

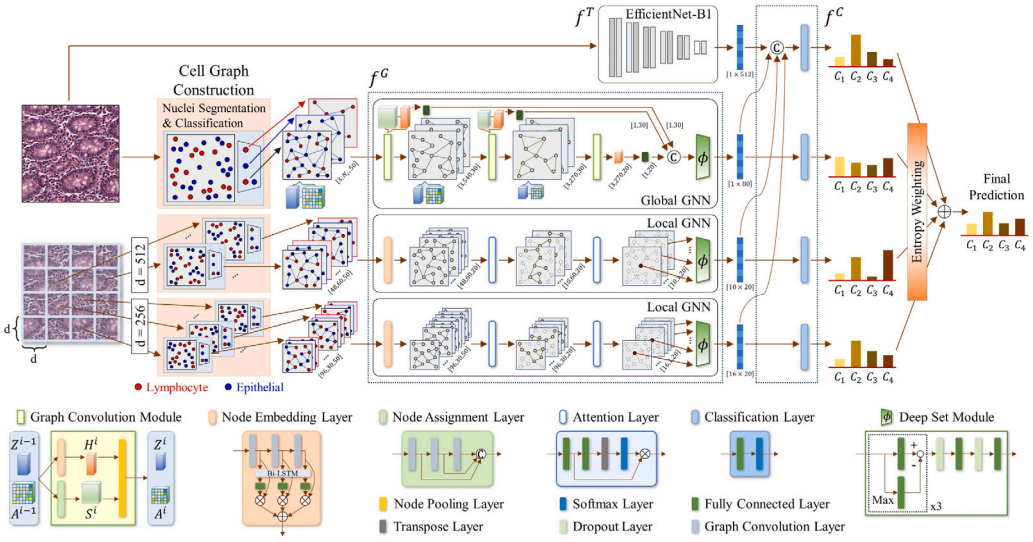

Fig. 1. Overview of MMGA-Net.

et al., 2021). Graphs are often constructed by using cells (or nuclei) where each cell is characterized by hand-crafted or deep learning-based features and edges are established based on the Euclidean distance between cells. For example, Anand et al. (2020) detects nuclei, extracts nuclei features using both hand-crafted features (color, gray level co-occurrence matrix, and the number of neighboring nuclei) and deep learning-based features, forms edges by thresholding the Euclidean distance among cells, and utilizes graph convolutions to conduct breast cancer classification. Similarly, Zhou et al. (2019) constructs a cell graph and uses a GNN for colorectal cancer classification. In a cell graph, each cell is characterized by 17 nuclear descriptors that are known to be discriminative for nuclei classification. Edges are formed by using the k-nearest neighbor (kNN) algorithm and the Euclidean distance among nuclei. In the GNN, intermediate outputs are concatenated, forming multi-level feature representations, and used for cancer classification. Chen et al. (2020) also utilizes a cell graph and a GNN for survival outcome prediction. To further improve the prediction performance, the output of GNN is combined with deep learning-based features from a pathology image and genomic features from molecular profiles. Moreover, for an improved tissue representation, Pati et al. (2022) proposes a hierarchical cell-to-tissue (HACT) graph representation for breast cancer classification that exploits both cells and tissue components. Cell and tissue graphs are separately constructed and combined via an assignment module that assigns cells to the corresponding tissue components by using their spatial location. In addition, Javed et al. (2020) employs both cell-level and patch-level graphs to detect cellular communities, which could lead to the discovery of distinct tissue phenotypes in colorectal cancer.

3. Methodology

3.1. Problem formulation

Suppose that we are given a pathology image $I \in \mathbb{N}^3$ with N_c cells of differing types. Let $\mathcal{G}$ be a set of N_m multi-level cell graphs $\left\{G^{(m)} \mid m = 1, \ldots, N_m\right\}$ where $G^{(m)} = \left(V^{(m)}, E^{(m)}\right)$ is the mth level cell graph with $N_c^{(m)} \leq N_c$ cells, $V^{(m)}$ denotes a set of nodes $\left\{v_i^{(m)} \mid i = 1, \ldots, N_c^{(m)}\right\}$, and $E^{(m)}$ represents a set of edges $\left\{e_{v_i,v_j}^{(m)} \mid v_i^{(m)}, v_j^{(m)} \in V^{(m)} \text{ and } v_i^{(m)} \neq v_j^{(m)}\right\}$, respectively. Each node $v_i^{(m)} \in V^{(m)}$ represents a cell in I that is quantified by a d-dimensional feature vector $b_{v_i} \in \mathbb{R}^d$. An adjacency matrix $A^{(m)} \in \mathbb{R}^{n \times n}$ of the graph $G^{(m)}$ shows the association between nodes, i.e., cells. It simply tells whether there exists an edge between any pair of nodes. If $A_{i,j}^{(m)} > 0$, $e_{v_i,v_j}^{(m)} \in E^{(m)}$. Otherwise, $e_{v_i,v_j}^{(m)} \notin E^{(m)}$.

The objective of this study is to learn a mapping function $f : \mathbb{N}^3 \longrightarrow \mathbb{N}^1$ such that $\forall I \in \mathbb{N}^3$, $f(I) = y \in \{0, 1, \ldots, C-1\}$ where C is the cardinality of class labels. The function f consists of three kinds of sub-functions, including f^G, f^T, and f^C. f^G is a set of functions $\left\{f^{G^{(m)}} \mid m = 1, \ldots, N_m\right\}$, of which each $f^{G^{(m)}}$ is built based upon a GNN and maps the mth level graph $G^{(m)}$ to an embedding vector $b^{G^{(m)}} \in \mathbb{R}^{1 \times m_{G^{(m)}}}$ using cellular interactions among $\left\{v_i^{(m)} \mid i = 1, \ldots, N_c^{(m)}\right\}$. f^T and f^C are built based on convolutional neural networks. F^T receives I and extracts another embedding vector $b^T \in \mathbb{R}^{1 \times m_T}$ that provides tissue-level information. F^C takes feature vectors from f^G and f^T and utilizes them to make a prediction. Hence, the objective can be formulated as minimizing a loss function $\mathcal{L}(f(I;\theta))$ where $\theta = \left\{\theta_{G^{(1)}}, \theta_{G^{(2)}}, \ldots, \theta_{G^{(m)}}, \theta_C, \theta_S\right\}$ denotes trainable parameters.

3.2. Cell graph construction

Given the pathology image I, we construct a cell graph G by (1) identifying cells using an advanced nuclear instance segmentation and classification algorithm, (2) extracting nuclear features from each cell, (3) sampling a subset of cells to reduce redundancy, and improve the efficiency of the cell graph, and (4) computing edge weights in the graph. The overall diagram for the cell graph construction is shown in Fig. 2.

3.2.1. Nuclei segmentation and classification

We adopt SONNET (Doan et al., 2022) to identify cells of different types in the pathology image I. SONNET contains a single encoder and three decoders. The encoder extracts the high-dimensional feature representation. The three decoders generate a nuclear foreground map, a nuclear type map, and a nuclear ordinal regression map. During training, the nuclear ordinal regression map is used to identify the most challenging regions, i.e., pixels near nuclear boundaries, and to direct the network to focus on these pixels for improved nuclear segmentation and classification. SONNET is trained on GLySAC dataset (Doan et al., 2022) with three types of nuclei, including epithelial, lymphocyte, and miscellaneous. Thus, SONNET identifies individual cells in the pathology image I and classifies them into three types.

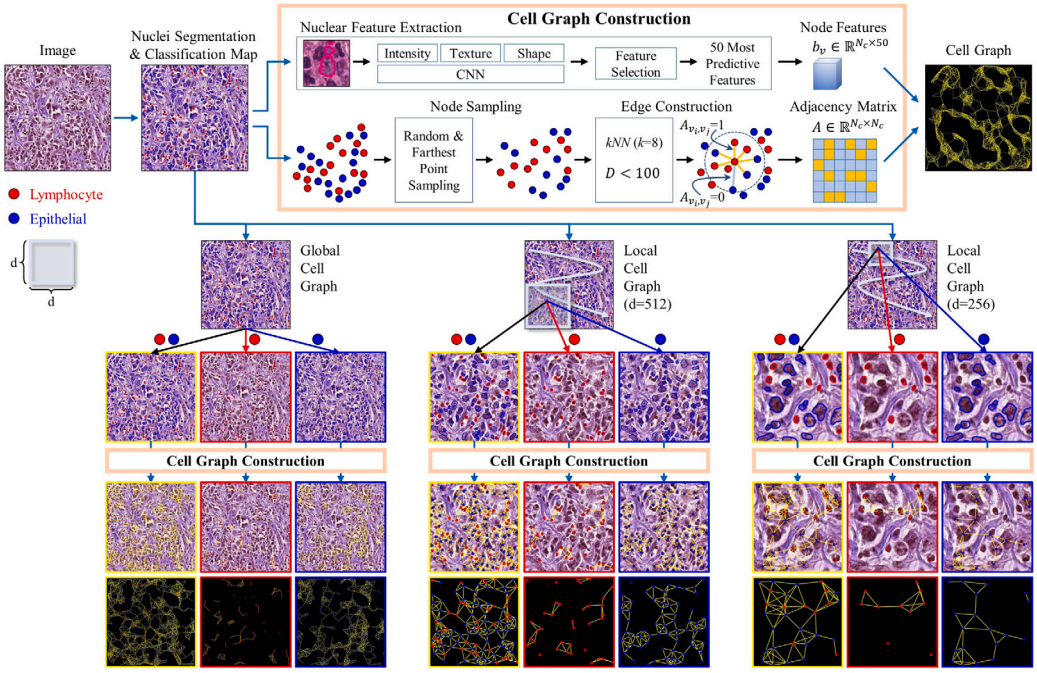

Fig. 2. Overview of multi-cell type and multi-level graph construction.

3.2.2. Nuclear feature extraction

For each of the identified cells (or nuclei), denoted by v_i, we extract a feature vector $b_{v_i} \in \mathbb{R}^d$ that describes its shape, texture, and appearance ($d = 50$). Specifically, we obtain 1280 features from the last convolutional layer of SONNET's encoder using the ROI Align technique (He et al., 2017). We also incorporate 17 cellular/nuclear descriptors, including 5 intensity-based features (mean and standard deviation of nuclei intensity, skewness of nuclei intensity, mean entropy of nuclei intensity, and the average difference in the intensity of foreground and background pixels), 4 texture-based features (dissimilarity, homogeneity, energy, and angular second moment of GLCM), and 8 shape-based features (eccentricity, area, maximum and minimum length of axis, perimeter, solidity, orientation, and centroid coordinates). Provided with the 1297 features, we utilize random forest to choose 50 features, i.e., $d = 50$, that are most predictive for cancer types. Utilizing all nuclei in each image of the training set of the colorectal dataset, we train random forest for cancer classification and compute the importance of each feature. Sorting all the features by their importance, we select the top-50 features for further investigation. The 50 selected features include 7 cellular/nuclear descriptors and 33 CNN-driven features. The list of the selected features is available in Supplementary Table 1 and 2.

3.2.3. Graph node sampling

Following Zhou et al. (2019), we utilize a subset of cells to construct a graph, which could reduce the redundancy among cells with similar characteristics and computational complexity. Two sampling strategies are employed. One is random sampling where cells within a pathology image are randomly selected with a probability of 0.2. The other is farthest point sampling in which 40% of the cells are selected iteratively. Given an initial set of the selected cells, each time a new cell is added to the selected set if it has the farthest distance from all the cells in the set. The union of the two selected sets of cells serves as nodes of a cell graph G. However, the selection is likely to be sensitive to the starting set. So, the first approach provides extra cells at random. By using these two strategies, we aim to select >50% of the cells in an image that is dispersed and diverse enough.

3.2.4. Graph edge construction

Upon completion of graph node sampling, we define an edge e_{v_i,v_j} for a pair of neighboring cells or nodes (v_i, v_j) to represent cell-to-cell interactions. To encourage the interaction among neighboring cells, we limit the maximum degree of a node to 8 and the maximum Euclidean distance between any pair of nodes to 100. These can be summarized as an adjacency matrix A as follows: $A_{v_i,v_j} = 1$ if $v_j \in kNN(v_i)$ and $D(v_i, v_j) < 100$ and, otherwise, $A_{v_i,v_j} = 0$ where D denotes the Euclidean distance.

3.2.5. Multi-cell type and multi-level graph construction

For each pathology image I, we construct three cell graphs $G^{(1)}$, $G^{(2)}$, and $G^{(3)}$ where $G^{(1)}$ is a global graph and $G^{(2)}$ and $G^{(3)}$ are local graphs (Fig. 2). The global graph $G^{(1)}$ is constructed for the entire pathology image I. It includes three kinds of graphs, a graph for epithelial cells, a graph for lymphocytes, and a graph for both epithelial cells and lymphocytes. The first and second graphs are for epithelial cells and lymphocytes, respectively, which are used to examine the cellular interactions of each cell type, and the third graph is for cellular interactions among all the cells, including both epithelial cells and lymphocytes, in the image I. To construct the two local graphs $G^{(2)}$ and $G^{(3)}$, we slide a rectangular window of size $d \times d$ pixels where d is set to 512 and 256 by heuristics, respectively, throughout the pathology image I with a stride of 256. For each local region of size $d \times d$ pixels, we construct three cell graphs: one for epithelial cells, one for lymphocytes, and one for both epithelial cells and lymphocytes, similar to the construction of the global graph $G^{(1)}$. Therefore, both $G^{(2)}$ and $G^{(3)}$ contain multitude of three types of cell graphs.

3.3. Network architecture

The overall architecture of MMGA-Net is described in Fig. 1. MMGA-Net includes three components that are corresponding to f^G, f^T, and f^C. The first component (f^G) includes three branches ($f^{G^{(1)}}$, $f^{G^{(2)}}$ and $f^{G^{(3)}}$) that are built based upon a GNN to utilize and learn cell-to-cell interactions in the pathology image I (Hence, $N_m = 3$). The first branch ($f^{G^{(1)}}$) utilizes a global cell graph to obtain the overall cellular interactions. The second and third branches ($f^{G^{(2)}}$ and $f^{G^{(3)}}$)

exploit local cell graphs to learn regional cellular interactions. The next component (f^T) employs a CNN to extract the overall tissue-level information in the pathology image I. The last component (f^C) combines and processes the outputs of f^G and f^T to provide a class label for the pathology image I. Hence, we analyze various aspects of a pathology image at multiple levels from tissue to cell characteristics, from global to local cell-to-cell interactions, and from whole-cell to cell-type specific interactions.

For $f^{G^{(1)}}$, $f^{G^{(2)}}$, and $f^{G^{(3)}}$, the pathology image I is converted into graphs $G^{(1)}$, $G^{(2)}$, and $G^{(3)}$, respectively. $f^{G^{(1)}}$ is composed of a series of three graph convolution modules (ψ_1, ψ_2, and ψ_3) and a deep set module ϕ where ψ_l contains a node embedding layer, a node assignment layer, and a node pooling layer. For $f^{G^{(2)}}$ and $f^{G^{(3)}}$, we adopt a node embedding layer, two attention layers, and ϕ. $f^{G^{(1)}}$, $f^{G^{(2)}}$, and $f^{G^{(3)}}$ produce an embedding vector $f^G(I) = \left[f^{G^{(m)}}\left(G^{(m)}\right)\right] = \left[b_{G^{(m)}} \in \mathbb{R}^{1\times m_{G^{(m)}}}\right]$, $m = 1, 2, 3$. Moreover, f^T utilizes EfficientNet-B1 (Tan and Le, 2019) except the last FC layer to generate an embedding vector $f^T(I) = b^T \in \mathbb{R}^{1\times m_T}$. The four feature vectors $\left\{b^{G^{(1)}}, b^{G^{(2)}}, b^{G^{(3)}}, b^T\right\}$ are fed into f^C that contains multiple classification layers, of which each consists of a fully connected (FC) layer and a softmax layer. Each of the first three feature vectors from f^G is separately fed into a classification layer. The four feature vectors are concatenated and fed into another classifier layer to predict a class label.

3.3.1. Graph convolution module

Inspired by Zhou et al. (2019), the lth graph convolution module ψ_l consists of a node embedding layer, a node assignment layer, and a node pooling layer. Given the node features $Z^{(l-1)} \in \mathbb{R}^{n_{l-1}\times d_{l-1}}$ and adjacency matrix $A^{(l-1)} \in \mathbb{R}^{n_{l-1}\times n_{l-1}}$ from ψ^{l-1}, the node assignment layer generates the assignment matrix $S^{(l)}$ where $S^{(l)}_{i,j} \in \mathbb{R}^{n_{l-1}\times d_l}$ denotes the probability that the ith input node is assigned to the jth hidden node and the node embedding layer extracts the embedding matrices $H^{(l)} \in \mathbb{R}^{n_{l-1}\times n_l}$, which represent the hidden representation of the input nodes. Given $H^{(l)}$ and $S^{(l)}$, the node pooling layer clusters the nodes and produces a coarsened graph with $n_{l+1} < n_l$ nodes.

The node assignment layer conducts a series of three graph convolutions and obtains three embeddings $\left\{h^{(1)}, h^{(2)}, h^{(3)}\right\}$, including the output of each graph convolution, and aggregates them via concatenation, i.e., aggregating three-hop neighbor's information. The node embedding layer also performs a series of three graph convolutions and conducts a weighted sum of the three embedding $\left\{h^{(1)}, h^{(2)}, h^{(3)}\right\}$ as $\sum_{i=1}^{3} \alpha^{(i)} h^{(i)}$ where $\alpha^{(i)}$ is a weight for the ith embedding. To compute these weights, the three embeddings $\left\{h^{(1)}, h^{(2)}, h^{(3)}\right\}$ are fed into a bi-directional LSTM, the forward and backward embeddings are concatenated and go through a FC layer, generating three outputs $\left\{o^{(1)}, o^{(2)}, o^{(3)}\right\}$, and a softmax function is applied to these outputs $\alpha^{(i)} = \frac{\exp\left(o^{(i)}\right)}{\sum_{j=1}^{3}\exp\left(o^{(j)}\right)}$. Provided with $H^{(l)}$ and $S^{(l)}$, the node pooling layer generates new node features $Z^{(l+1)}$ and a new adjacency matrix $A^{(l+1)}$ as follows: $Z^{(l+1)} = {S^{(l)}}^T H^{(l)} \in \mathbb{R}^{n_{l+1}\times d_{l+1}}$ and $A^{(l+1)} = {S^{(l)}}^T A^{(l)} S^{(l)} \in \mathbb{R}^{n_{l+1}\times n_{l+1}}$.

The first two graph convolution modules have an identical structure but the last module misses the node assignment layer and node pooling layer. From the three modules, we collect the embedding matrices $\left\{H^{(1)}, H^{(2)}, H^{(3)}\right\}$, conduct a max operation for each, generating embedding vectors, and concatenate them. The concatenated embedding vectors are fed into a deep set module ϕ to produce a permutation invariant embedding vector.

3.3.2. Deep set module

A deep set module ϕ includes a series of three deep set layers, followed by a series of a Dropout layer, an FC layer, an ELU activation layer, a Dropout layer, and an FC layer. Dropout layers are implemented using the probability of 0.5. A deep set layer (Zaheer et al., 2017) is defined as follows:

$$o = \sigma\left(h\Lambda - 1\,\mathrm{maxpool}\left(h\right)\Gamma\right) \tag{1}$$

where $h \in \mathbb{R}^{n\times d}$ is a set of input embedding vectors, $o \in \mathbb{R}^{n\times d'}$ is a set of permutation invariant output embedding vectors, σ is an exponential linear unit (ELU) activation function, Λ, $\Gamma \in \mathbb{R}^{D\times D'}$ are learnable parameters, $1 = [1, \ldots, 1]^T \in \mathbb{R}^n$ is a column vector, and $\mathrm{maxpool}(\cdot)$ is a column-wise max operation.

3.3.3. Attention layer

For $f^{G^{(2)}}$ and $f^{G^{(3)}}$, two attention layers (Ilse et al., 2018) are utilized, of which each employs a series of two FC layers, a transpose layer, and a softmax layer. These two attention layers are used to compute weights associated with local graphs and nodes. Using the weights, the first attention layer chooses the most important k local graphs and the second attention layer selects the single most representative node.

3.4. Entropy-weighted inference

MMGA-Net provides four predictions at multiple levels. During the testing phase, we make the final prediction by combining these four predictions via an entropy weighting method. The entropy weighting scheme is utilized to determine a weight for the ith prediction given by:

$$E_i = -\frac{\sum_{j=1}^{n} p_{ij}\ln(p_{ij})}{\ln(n)} \tag{2}$$

$$\omega_i = \frac{1-E_i}{\sum_{i=1}^{m}(1-E_i)} \tag{3}$$

where n is the number of classes, E_i is the entropy of the ith prediction, p_{ij} is the probability of the class j in the ith prediction, ω_i is the weight for the ith prediction, m is the number of predictions ($m = 4$). The range of E_i is [0, 1]. The larger E_i is, the greater uncertainty the ith prediction has, and thus, the smaller weight it takes.

3.5. Datasets

To examine the effectiveness of the proposed method, we employ two different types of cancer datasets, including the colorectal cancer dataset and the gastric cancer dataset. Table 1 shows the details of the colorectal and gastric cancer datasets. The colorectal cancer dataset (Le Vuong et al., 2021) is publicly available.[1] It includes two sets of tissue patches that were digitized using two scanners. The first set contains 6 colorectal tissue microarrays (TMAs) from 340 patients and 3 WSIs from 3 patients that were scanned at 40x magnification using an Aperio digital slide scanner (Leica Biosystems) with a pixel resolution of 0.2465 μm × 0.2465 μm and 0.2518 μm × 0.2518 μm, respectively. The second set comprises 45 WSIs from 45 patients that were scanned at 40x magnification using a NanoZoomer digital slide scanner (Hamamatsu Photonics K.K.) with a pixel resolution of 0.2253 μm × 0.2253 μm. In the first set, there are 9863 image patches of size 1024 × 1024 pixels, which are divided into training, validation, and test data (C-Test-I). The second set, designated as C-Test-II, includes 110 260 image patches of size 1144 × 1144 pixels that are resized to 1024 × 1024 pixels. Upon histologic review, each of these image patches was classified into benign (BN), well-differentiated cancer (WD), moderately differentiated cancer (MD), and poorly differentiated cancer (PD).

The gastric cancer dataset contains 96 de-identified WSIs from 96 patients collected between 2016 and 2020 from Kangbuk Samsung Hospital (IRB No. 2021-04-035). The digitized images were taken at 40x magnification using an Aperio digital slide scanner (Leica Biosystems).

[1] https://github.com/QuIIL/KBSMC_colon_cancer_grading_dataset.

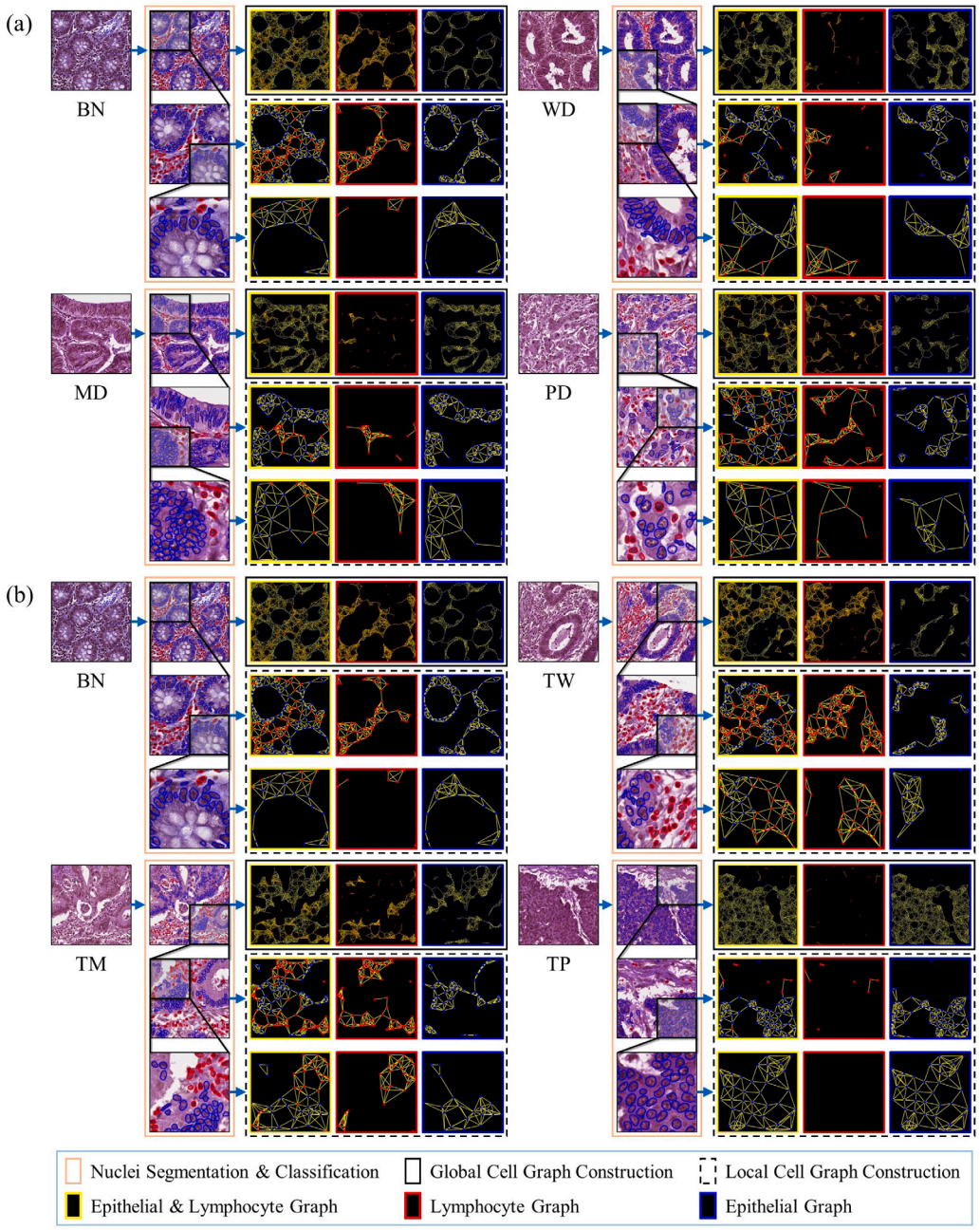

Fig. 3. Exemplary tissue images and global and local graphs. (a) Colorectal cancer and (b) gastric cancer.

The size of a WSI is ~100,000 × ~80,000 with 0.2635 μm × 0.2635 μm pixel spacing. Each WSI was examined by experienced pathologists (K. Kim and B. Song) to identify distinct tissue regions, including BN, tubular well-differentiated adenocarcinoma (TW), tubular moderately-differentiated adenocarcinoma (TM), and tubular poorly- differentiated adenocarcinoma (TP). Using these tissue regions, we generated image patches of size 1024 × 1024 pixels that are split into training, validation, and test data (G-Test). Some exemplary images along with multi-cell and multi-level graph are shown in Fig. 3.

3.6. Comparative models

3.6.1. CNNs

A number of plain CNN models are employed in this study, including DenseNet-121 (Huang et al., 2017), EfficientNet-B0, -B1, and -B2 (Tan and Le, 2019), and ResNet-34, -50, and -101 (He et al., 2016). Moreover, a multi-task CNN approach, built based upon EfficientNet-B0, is adopted (Le Vuong et al., 2021). It proposes to re-formulate cancer grading as categorical and ordinal classification problems and to jointly conduct the two classifications for cancer grading. As for ordinal classification, it uses (1) a mean absolute error (MAE) and (2) a mean square error (MSE) as a loss function. To further improve the learning capability of the model, it also proposes so called ordinal cross-entropy loss that converts the output of the ordinal classification into probability measures and computes cross-entropy loss. The combination of the ordinal classification loss and ordinal cross-entropy gives rise to two models, i.e., $\mathcal{M}_{MAE-CE_O}$ and $\mathcal{M}_{MSE-CE_O}$.

3.6.2. ViT

Transformers were first introduced in natural language processing by Vaswani et al. (2017) and have achieved remarkable success in numerous applications. They were later adapted for vision tasks, resulting

Table 1
Details of colorectal and gastric cancer datasets.

Colorectal cancer					Gastric cancer			
Class	Training	Validation	C-Test-I	C-Test-II	Class	Training	Validation	G-Test
BN	773	374	453	27 986	BN	150 063	29 790	26 221
WD	1866	264	192	8394	TW	14 162	8809	7197
MD	2997	370	738	61 985	TM	20 808	9510	9892
PD	1397	234	205	11 895	TP	27 597	9464	14 386

in Vision Transformers (ViTs) (Dosovitskiy et al., 2020). ViT has been applied to various medical imaging tasks, including pathological diagnosis applications, such as Chen et al. (2022) and Mari et al. (2022). In ViT, an input image is divided into several patches and passed through a series of self-attention layers, enabling the model to learn global dependencies and meaningful representations/embeddings. These embeddings are then processed by an MLP head for image classification.

3.6.3. CGC-Net

CGC-Net (Zhou et al., 2019) is a GNN-based model that is specifically designed to conduct colorectal cancer grading. It includes three stages of graph convolution and graph pooling modules. To construct a cell graph, cells are identified by using a nuclear instance map obtained by SONNET, and edges are formed by using kNN and the Euclidean distance among cells. In a cell graph, each cell is quantified using several morphological and texture features. Node embedding features from the three stages of CGC-Net are concatenated and used to predict cancer grades.

3.6.4. HACT-Net

HACT-Net (Pati et al., 2022) is another GNN-based model that exploits both a cell graph and a tissue graph to take into account the hierarchical relationship between cells and tissue components. A cell graph is constructed using HoVer-Net (Graham et al., 2019) for cell segmentation as well as a kNN-based strategy for edge configuration. As for a tissue graph, tissue components are first identified by using superpixels, k-means clustering, and iterative merging strategy. Using the tissue components, a tissue graph is constructed by forming edges between adjacent tissue components following Potjer (1996). To obtain hierarchical cell-to-tissue graph representation, HACT-Net assigns cell-to-tissue components according to their spatial locations.

3.6.5. Pathomic fusion

Pathomic fusion (Chen et al., 2020) is a multi-model fusion model that utilizes both pathology images and genomic information for cancer diagnosis and prognosis. Pathomic fusion has three streams for model fusion: a CNN for WSIs, a GNN for a cell graph, and a feed-forward network for the genomic profile. This generates three sets of multimodal features that are fused via Kronecker Product and gating-based attention. VGG19 is adopted as the architecture of CNN. For the GNN, nuclei are identified by SONNET and used to construct a cell graph using kNN and the Euclidean distance among nuclei. A set of manual and unsupervised cell features are employed to quantify each nucleus. The manual features include morphological and texture features. The unsupervised features are extracted using contrastive predictive coding (CPC) scheme (Henaff, 2020). For our study, we only utilize the first two streams with CNN and GNN since we do not have the genomics profile of the two datasets under consideration.

3.7. Training schemes and evaluation metric

3.7.1. Training strategy

MMGA-Net is implemented using PyTorch (Paszke et al., 2017) alongside PyTorch geometric deep learning package (Fey and Lenssen, 2019). The model is trained for 40 epochs, with three different random seeds, batch size of 16, Adam optimization with default parameter values ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1.0e^{-8}$) and an initial learning rate of $1.0e^{-4}$, which decreases to $1.0e^{-5}$ after 20 epochs. In MMGA-Net, three sets of trainable weights W^G, W^T, and W^C that are corresponding to three branches f^G, f^T, and f^C, respectively. W^T is initialized with the pre-trained weights on ImageNet. Cross-entropy loss is employed to optimize these weights W^G, W^T, and W^C.

Moreover, plain CNN models are initialized using pre-trained weights on ImageNet adopted from Pytorch Library (Paszke et al., 2017) and optimized by setting the learning rate to $1.0e^{-5}$ and using a cosine annealing scheduler, Adam optimizer, and cross-entropy loss. The number of epochs and batch size are set to 30 and 16, respectively. During the training phase, several augmentations are applied as follows: (1) random horizontal flip, (2) random vertical flip, (3) affine transformation, and (4) random contrast in a range of [−20, 20]. All augmentations are adopted from Aleju library.[2] ViT is also initialized by per-trained weights on Imagenet, with hyperparameters such as a patch-size of 64 × 64 pixels and a number of heads of 8, and a number of Transformer layers to be 8. All images are downscaled to 512 × 512 pixels for efficient processing while retaining sufficient resolution for multiclass classification. Global average pooling is adopted to produce the final feature vector, which is used to obtain the logit. Then, class probabilities are calculated using the SoftMax operation. The two multi-task CNN models ($\mathcal{M}_{MAE-CE_O}$ and $\mathcal{M}_{MSE-CE_O}$) and three GNN models (CGC-Net, HACT-Net, and Pathomic Fusion) are optimized by following the training strategy of the original papers. All the image patches are resized, if necessary, to 1024 × 1024 pixels.

3.7.2. Evaluation metrics

We employ three evaluation metrics to assess the performance of MMGA-Net and other comparative models, including accuracy (Acc), macro average F1-score (F1), and quadratic weighted kappa (κ_w) (Cohen, 1968). Acc is the ratio of the number of correctly classified instances over total number of instances. F1 is computed as the harmonic mean of the average precision and recall. κ_w is calculated as $\kappa_w = 1 - \frac{\sum_i^n \sum_j^n w_{ij} c_{ij}}{\sum_i^n \sum_j^n w_{ij} g_{ij}}$ where $w_{ij} = \frac{(i-j)^2}{(n-1)^2}$ and c_{ij} and g_{ij} are the predicted and expected proportions for the predicted class i and ground truth class j, respectively.

4. Results

4.1. Classification results on colorectal cancer tissues

Table 2 summarizes the results of colorectal cancer grading by MMGA-Net in comparison to other comparative methods. MMGA-Net obtained 88.55% Acc, 0.865 F1, and 0.952 κ_w in C-Test-I and 80.92% Acc, 0.751 F1, and 0.890 κ_w in C-Test-II. In comparison to other competing models, MMGA-Net consistently obtained superior performance regardless of the type of test datasets and evaluation metrics except κ_w in C-Test-II. Among other competing models, the multi-task CNN models ($\mathcal{M}_{MAE-CE_O}$ and $\mathcal{M}_{MSE-CE_O}$) showed better performance than others. Excluding these multi-task CNN models, GNN-based models, in general, demonstrated better (or comparable) performance than plain CNN-based models. HACT-Net, a GNN-based model that utilizes both a cell graph and a tissue graph, achieved better performance

[2] https://github.com/aleju/imgaug.

Table 2
Comparative results for colorectal and gastric cancer grading.

Model	C-Test-I		
	Acc (%)	F1	κ_w
DenseNet-121 (Huang et al., 2017)	81.37 ± 1.717	0.730 ± 0.034	0.868 ± 0.013
ResNet-34 (He et al., 2016)	86.26 ± 1.362	0.819 ± 0.010	0.914 ± 0.009
ResNet-50 (He et al., 2016)	84.61 ± 0.738	0.814 ± 0.036	0.929 ± 0.005
ResNet-101 (He et al., 2016)	83.20 ± 1.827	0.780 ± 0.025	0.896 ± 0.018
EfficientNet-B0 (Tan and Le, 2019)	85.72 ± 0.724	0.807 ± 0.010	0.929 ± 0.008
EfficientNet-B1 (Tan and Le, 2019)	86.71 ± 0.317	0.830 ± 0.004	0.939 ± 0.005
EfficientNet-B2 (Tan and Le, 2019)	85.91 ± 1.874	0.834 ± 0.024	0.939 ± 0.017
ViT (Dosovitskiy et al., 2020)	84.29 ± 1.207	0.815 ± 0.013	0.927 ± 0.014
$\mathcal{M}_{MAE-CE_o}$ (Le Vuong et al., 2021)	87.73 ± 0.617	0.844 ± 0.013	0.941 ± 0.004
$\mathcal{M}_{MAE-CE_o}$ (Le Vuong et al., 2021)	88.44 ± 0.537	0.859 ± 0.011	0.942 ± 0.007
CGC-Net (Zhou et al., 2019)	85.40 ± 1.980	0.791 ± 0.028	0.920 ± 0.016
Pathomic Fusion (Chen et al., 2020)	86.19 ± 1.628	0.821 ± 0.024	0.936 ± 0.020
HACT-Net (Pati et al., 2022)	86.95 ± 1.375	0.835 ± 0.013	0.939 ± 0.017
MMGA-Net (ours)	**88.55 ± 0.710**	**0.865 ± 0.010**	**0.952 ± 0.007**
Model	C-Test-II		
	Acc (%)	F1	κ_w
DenseNet-121 (Huang et al., 2017)	71.19 ± 1.809	0.656 ± 0.029	0.827 ± 0.021
ResNet-34 (He et al., 2016)	74.05 ± 1.860	0.676 ± 0.014	0.851 ± 0.009
ResNet-50 (He et al., 2016)	72.91 ± 0.874	0.684 ± 0.009	0.856 ± 0.007
ResNet-101 (He et al., 2016)	70.05 ± 1.463	0.660 ± 0.010	0.839 ± 0.013
EfficientNet-B0 (Tan and Le, 2019)	75.37 ± 0.919	0.698 ± 0.013	0.856 ± 0.019
EfficientNet-B1 (Tan and Le, 2019)	76.63 ± 1.081	0.709 ± 0.015	0.863 ± 0.018
EfficientNet-B2 (Tan and Le, 2019)	75.72 ± 1.956	0.699 ± 0.0187	0.863 ± 0.010
ViT (Dosovitskiy et al., 2020)	76.47 ± 1.890	0.703 ± 0.017	0.860 ± 0.009
$\mathcal{M}_{MAE-CE_o}$ (Le Vuong et al., 2021)	80.35 ± 0.941	0.745 ± 0.019	0.892 ± 0.009
$\mathcal{M}_{MAE-CE_o}$ (Le Vuong et al., 2021)	79.17 ± 0.884	0.738 ± 0.015	0.875 ± 0.008
CGC-Net (Zhou et al., 2019)	75.35 ± 2.567	0.692 ± 0.012	0.855 ± 0.023
Pathomic Fusion (Chen et al., 2020)	77.20 ± 1.593	0.710 ± 0.020	0.859 ± 0.017
HACT-Net (Pati et al., 2022)	78.30 ± 1.870	0.729 ± 0.019	0.871 ± 0.020
MMGA-Net (ours)	**80.92 ± 1.229**	**0.751 ± 0.015**	**0.890 ± 0.008**
Model	G-Test		
	Acc (%)	F1	κ_w
DenseNet-121 (Huang et al., 2017)	72.91 ± 1.421	0.720 ± 0.034	0.818 ± 0.024
ResNet-34 (He et al., 2016)	83.85 ± 0.922	0.983 ± 0.007	0.897 ± 0.014
ResNet-50 (He et al., 2016)	82.28 ± 0.981	0.897 ± 0.007	0.885 ± 0.017
ResNet-101 (He et al., 2016)	81.05 ± 1.021	0.750 ± 0.014	0.822 ± 0.016
EfficientNet-B0 (Tan and Le, 2019)	84.13 ± 0.776	0.891 ± 0.006	0.901 ± 0.009
EfficientNet-B1 (Tan and Le, 2019)	83.92 ± 0.580	0.785 ± 0.008	0.894 ± 0.017
EfficientNet-B2 (Tan and Le, 2019)	83.48 ± 0.781	0.780 ± 0.018	0.929 ± 0.014
ViT (Dosovitskiy et al., 2020)	83.27 ± 1.027	0.781 ± 0.019	0.918 ± 0.011
$\mathcal{M}_{MAE-CE_o}$ (Le Vuong et al., 2021)	83.70 ± 0.385	0.784 ± 0.007	0.929 ± 0.012
$\mathcal{M}_{MAE-CE_o}$ (Le Vuong et al., 2021)	84.39 ± 1.804	0.791 ± 0.012	0.931 ± 0.019
CGC-Net (Zhou et al., 2019)	84.70 ± 1.499	0.790 ± 0.014	0.910 ± 0.016
Pathomic Fusion (Chen et al., 2020)	84.81 ± 0.795	0.799 ± 0.012	0.902 ± 0.008
HACT-Net (Pati et al., 2022)	85.87 ± 1.801	0.808 ± 0.007	0.921 ± 0.008
MMGA-Net (ours)	**87.32 ± 0.384**	**0.834 ± 0.007**	**0.936 ± 0.010**

among GNN-based models. CGC-Net was inferior to other GNN-based models. Moreover, regarding plain CNN-based models, EfficientNet-B1 produced the best results. DenseNet-121 and ResNet-101, however, were the two worst models.

In a head-to-head comparison between C-Test-I and C-Test-II, there was a consistent performance drop for all the models under consideration; approximately 8.00% ∼ 12.00% in Acc, 0.07 ∼ 0.14 in F1, and 0.04 ∼ 0.08 in κ_w. This is likely due to the difference between the two datasets as described in Section 3.5. C-Test-II was acquired using a different digital scanner and at a different time from the training and validation datasets for model training, whereas C-Test-I was obtained from the same patient population using the same digital scanner. As mentioned above, MMGA-Net, in general, outperformed other models on C-Test-II, which indicates the superior robustness of the model to the domain shift due to digital scanners. Among other models, $\mathcal{M}_{MAE-CE_O}$ demonstrated a lower performance drop than others; however, $\mathcal{M}_{MSE-CE_O}$ produced a larger performance drop than GNN-based models. GNN-based models, in general, demonstrated lower performance drop than plain CNN-based models. Within the GNN-based models, CGC-Net obtained the poorest performance. As for plain CNN-based models, the decrease in performance was more or less the same for all models. EfficientNet-B1 showed the lowest performance drop than other plain CNN-based models.

4.2. Classification results on gastric cancer tissues

Classification results on the gastric cancer dataset are available in Table 2. Overall, similar observations with the results on the colorectal cancer datasets were made on the gastric cancer dataset. MMGA-Net achieved 87.32% Acc, 0.834 F1, and 0.936 κ_w, outperforming other competing models. Among other competing models, the results were inconsistent with those in C-Test-I and C-Test-II. HACT-Net was second to MMGA-Net. GNN-based models were superior to plain and multi-task CNN-based models. CGC-Net was poorer than other GNN-based models. Among plain CNN-based models, EfficientNet-B1 obtained the best results. DenseNet-121 was the worst model in gastric cancer grading. These results suggest that our approach is not specific to a particular type of cancer or dataset and applies to other types of cancers or diseases.

Table 3
Results for ablation experiments on multi-level graphs.

Model	C-Test-I		
	Acc (%)	F1	κ_w
$f^{G^{(1)}}$	86.45 ± 0.007	0.819 ± 0.011	0.933 ± 0.009
$f^{G^{(2)}} + f^{G^{(3)}}$	86.15 ± 1.170	0.825 ± 0.013	0.937 ± 0.011
f^{T}	86.71 ± 0.317	0.830 ± 0.004	0.939 ± 0.005
$f^{G^{(1)}} + f^{T}$	87.83 ± 0.354	0.855 ± 0.011	0.947 ± 0.013
$f^{G^{(2)}} + f^{G^{(3)}} + f^{T}$	87.09 ± 0.718	0.835 ± 0.013	0.942 ± 0.008
$f^{G^{(1)}} + f^{G^{(2)}} + f^{G^{(3)}}$	86.92 ± 1.047	0.831 ± 0.008	0.940 ± 0.007
MMGA-Net (ours)	**88.55 ± 0.710**	**0.868 ± 0.010**	**0.950 ± 0.007**
Model	C-Test-II		
	Acc (%)	F1	κ_w
$f^{G^{(1)}}$	76.60 ± 1.364	0.705 ± 0.012	0.858 ± 0.017
$f^{G^{(2)}} + f^{G^{(3)}}$	76.16 ± 1.373	0.708 ± 0.012	0.862 ± 0.017
f^{T}	76.63 ± 1.081	0.709 ± 1.015	0.863 ± 0.018
$f^{G^{(1)}} + f^{T}$	79.20 ± 1.455	0.738 ± 0.009	0.886 ± 0.007
$f^{G^{(2)}} + f^{G^{(3)}} + f^{T}$	77.12 ± 1.094	0.712 ± 0.013	0.868 ± 0.014
$f^{G^{(1)}} + f^{G^{(2)}} + f^{G^{(3)}}$	78.31 ± 1.428	0.726 ± 0.014	0.870 ± 0.012
MMGA-Net (ours)	**80.92 ± 1.229**	**0.756 ± 0.013**	**0.887 ± 0.009**
Model	G-Test		
	Acc (%)	F1	κ_w
$f^{G^{(1)}}$	84.07 ± 1.059	0.792 ± 0.018	0.899 ± 0.015
$f^{G^{(2)}} + f^{G^{(3)}}$	83.84 ± 1.338	0.790 ± 0.010	0.893 ± 0.010
f^{T}	83.92 ± 0.580	0.785 ± 0.008	0.894 ± 0.017
$f^{G^{(1)}} + f^{T}$	86.51 ± 1.091	0.820 ± 0.011	0.928 ± 0.009
$f^{G^{(2)}} + f^{G^{(3)}} + f^{T}$	84.22 ± 1.617	0.794 ± 0.018	0.903 ± 0.008
$f^{G^{(1)}} + f^{G^{(2)}} + f^{G^{(3)}}$	84.67 ± 1.354	0.796 ± 0.016	0.915 ± 0.015
MMGA-Net (ours)	**87.32 ± 0.384**	**0.840 ± 0.009**	**0.935 ± 0.010**

Table 4
Results for ablation experiments on multi-cell type graphs.

Model	C-Test-I		
	Acc (%)	F1	κ_w
$f^{G_E^{(1)}}$	82.25 ± 1.002	0.746 ± 0.015	0.880 ± 0.018
$f^{G_L^{(1)}}$	78.09 ± 1.409	0.719 ± 0.007	0.840 ± 0.017
$f^{G_{E+L}^{(1)}}$	84.62 ± 1.808	0.811 ± 0.009	0.931 ± 0.008
$f^{G^{(1)}}$	**86.45 ± 0.007**	**0.819 ± 0.011**	**0.933 ± 0.009**
Model	C-Test-II		
	Acc (%)	F1	κ_w
$f^{G_E^{(1)}}$	71.86 ± 1.804	0.670 ± 0.017	0.843 ± 0.019
$f^{G_L^{(1)}}$	64.98 ± 1.857	0.610 ± 0.017	0.775 ± 0.021
$f^{G_{E+L}^{(1)}}$	72.75 ± 1.510	0.673 ± 0.012	0.855 ± 0.013
$f^{G^{(1)}}$	**76.60 ± 1.364**	**0.705 ± 0.012**	**0.858 ± 0.017**
Model	G-Test		
	Acc (%)	F1	κ_w
$f^{G_E^{(1)}}$	80.71 ± 1.581	0.755 ± 0.019	0.853 ± 0.016
$f^{G_L^{(1)}}$	72.91 ± 1.149	0.727 ± 0.023	0.809 ± 0.017
$f^{G_{E+L}^{(1)}}$	81.60 ± 1.502	0.760 ± 0.010	0.855 ± 0.014
$f^{G^{(1)}}$	**84.07 ± 1.059**	**0.792 ± 0.018**	**0.899 ± 0.015**

4.3. Ablation studies

We conducted exhaustive ablation experiments on MMGA-Net to further assess the classification results and to gain insights into the model. Specifically, the ablation experiments sought to analyze the effect of (1) multi-level graphs, (2) multi-cell type graphs, and (3) aggregation mechanisms on MMGA-Net.

4.3.1. Role of multi-level graphs in cancer grading

MMGA-Net analyzed tissues at multiple levels, including a tissue-level (f^T) and two cellular levels, i.e., global and local cell-to-cell interactions ($f^{G^{(1)}}$ and $f^{G^{(2)}} + f^{G^{(3)}}$). We examined the effect of each of these levels and their combinations. The detailed results are described in Table 3. Using the information from a single level only, there was a consistent performance drop as compared to MMGA-Net; for instance, ≥2.06% Acc, ≥4.32% Acc, and ≥3.25% Acc for C-Test-I, C-Test-II, and G-Test, respectively. However, the global graph ($f^{G^{(1)}}$) and local graphs ($f^{G^{(2)}} + f^{G^{(3)}}$) alone were comparable to other GNN-based models (CGC-Net, Pathomic Fusion, and HACT-Net). The combination of the information from any two levels consistently provided an increase in the performance, which of each is superior to most of the CNN- and GNN-based models (Tables 2 and 3). Among the tissue-level information and cellular level graphs, there was a larger synergy between the global cell-to-cell interaction ($f^{G^{(1)}}$) and tissue-level information (f^T) in comparison to other combinations, including the combination among cell graphs ($f^{G^{(1)}}$, $f^{G^{(2)}}$, $f^{G^{(3)}}$) and the combination between local graphs ($f^{G^{(2)}}$, $f^{G^{(3)}}$) and tissue-level information (f^T).

4.3.2. Role of multi-cell type graphs in cancer grading

MMGA-Net utilizes three types of cell graphs, i.e., graphs for epithelial cells only (f^{G_E}), lymphocytes only (f^{G_L}), and both epithelial cells and lymphocytes ($f^{G_{E+L}}$). To investigate the role of different types of cell graphs in MMGA-Net, we conducted cancer grading using each type of cell graph in $f^{G^{(1)}}$. Table 4 demonstrates the results of the ablation experiments on multi-cell type graphs. Among $f^{G_E^{(1)}}$, $f^{G_L^{(1)}}$, and $f^{G_{E+L}^{(1)}}$, $f^{G_{E+L}^{(1)}}$ achieved the best performance, suggesting that it is advantageous to construct a single cell graph for multiple cell types in comparison to a cell graph using a single cell type. However, $f^{G^{(1)}}$ substantially outperforms $f^{G_E^{(1)}}$, $f^{G_L^{(1)}}$, and $f^{G_{E+L}^{(1)}}$; for example, an Acc of ≥1.83%, ≥3.85%, and ≥2.47% in C-Test-I, C-Test-II, and G-Test, respectively. These results clearly demonstrate that the construction and integration of multi-cell type graphs aid in analyzing pathology images and improving cancer grading.

4.3.3. Role of aggregation mechanism in cancer grading

MMGA-Net aggregates multiple tissues and cellular information to predict a cancer grade. To explore the effectiveness of the aggregation mechanism in MMGA-Net, we replaced the entropy weighting scheme with (1) hard (majority) voting and (2) soft voting. Hard voting takes the majority of the class labels that are predicted by the four tissue and cellular levels. Soft voting weights the vote by the probability of each class over the four predictions and takes the class label with the highest value. As shown in Table 5, the results show that the entropy weighting scheme was superior to two other aggregation methods regardless of the test datasets.

Furthermore, we examined the effect of invariant feature representations on cancer grading. For each of MMGA-Net, a global graph ($f^{G^{(1)}}$), and local graphs ($f^{G^{(2)}+G^{(3)}}$), we performed cancer grading without deep set modules (ϕ). Table 6 demonstrates the experimental results with and without ϕ. The addition of the deep set modules consistently improved the classification performance for the three datasets. These indicate that the feature representations that are invariant to the order of nodes and cells in both global and local graphs play a vital role in graph-based pathology image analysis.

5. Discussion

GNNs have been shown to be effective for pathology image analysis. The primary advantage of GNNs is the ability to exploit characteristics of individual histological objects, including glands, cells, etc., (as a node) and their relationships (as an edge). This is in line with the current practice of pathology where pathologists assess various kinds of histological objects and their arrangements and distributions to make a definitive decision. For cancer diagnosis, various handcrafted features have been adopted to extract and quantify individual histological objects and their relationships in conventional computational pathology. CNN-based approaches, in general, do not explicitly identify and utilize individual histological objects; they rather process tissue images as a whole. Some proposed a hybrid approach where both conventional approaches and CNN-based methods are employed

Table 5
Results for ablation experiments on aggregation methods.

Method	C-Test-I		
	Acc (%)	F1	κ_w
Hard voting	87.30 ± 1.130	0.840 ± 0.013	0.941 ± 0.018
Soft voting	87.99 ± 1.821	0.856 ± 0.019	0.950 ± 0.017
Entropy weighting (ours)	**88.55 ± 0.710**	**0.865 ± 0.010**	**0.952 ± 0.007**
Method	C-Test-II		
	Acc (%)	F1	κ_w
Hard voting	78.73 ± 2.01	0.730 ± 0.019	0.881 ± 0.015
Soft voting	79.88 ± 1.117	0.746 ± 0.009	0.890 ± 0.016
Entropy weighting (ours)	**80.92 ± 1.229**	**0.751 ± 0.015**	**0.890 ± 0.008**
Method	G-Test		
	Acc (%)	F1	κ_w
Hard voting	86.15 ± 1.321	0.810 ± 0.015	0.929 ± 0.016
Soft voting	86.78 ± 0.961	0.831 ± 0.011	0.932 ± 0.012
Entropy weighting (ours)	**87.32 ± 0.384**	**0.834 ± 0.007**	**0.936 ± 0.010**

Table 6
Results for ablation experiments on deep set modules.

Model	C-Test-I		
	Acc (%)	F1	κ_w
$f^{G^{(1)}}$ w.o. ϕ	85.69 ± 1.500	0.819 ± 0.010	0.931 ± 0.010
$f^{G^{(1)}}$	86.45 ± 0.007	0.819 ± 0.011	0.933 ± 0.009
$f^{G^{(2)}} + f^{G^{(3)}}$ w.o. ϕ	84.58 ± 0.931	0.815 ± 0.012	0.931 ± 0.014
$f^{G^{(2)}} + f^{G^{(3)}}$	86.15 ± 1.170	0.825 ± 0.013	0.937 ± 0.011
MMGA-Net w.o. ϕ	86.24 ± 1.161	0.820 ± 0.016	0.935 ± 0.011
MMGA-Net (ours)	**88.55 ± 0.710**	**0.868 ± 0.010**	**0.950 ± 0.007**
Model	C-Test-II		
	Acc (%)	F1	κ_w
$f^{G^{(1)}}$ w.o. ϕ	74.29 ± 1.500	0.686 ± 0.010	0.847 ± 0.010
$f^{G^{(1)}}$	76.60 ± 1.364	0.705 ± 0.012	0.858 ± 0.017
$f^{G^{(2)}} + f^{G^{(3)}}$ w.o. ϕ	72.89 ± 0.910	0.682 ± 0.013	0.842 ± 0.014
$f^{G^{(2)}} + f^{G^{(3)}}$	76.16 ± 1.373	0.708 ± 0.012	0.862 ± 0.017
MMGA-Net w.o. ϕ	76.26 ± 1.070	0.707 ± 0.012	0.868 ± 0.014
MMGA-Net (ours)	**80.92 ± 1.229**	**0.756 ± 0.013**	**0.887 ± 0.009**
Model	G-Test		
	Acc (%)	F1	κ_w
$f^{G^{(1)}}$ w.o. ϕ	83.25 ± 1.500	0.780 ± 0.010	0.891 ± 0.010
$f^{G^{(1)}}$	84.07 ± 1.059	0.792 ± 0.018	0.899 ± 0.015
$f^{G^{(2)}} + f^{G^{(3)}}$ w.o. ϕ	83.89 ± 1.334	0.779 ± 0.012	0.891 ± 0.013
$f^{G^{(2)}} + f^{G^{(3)}}$	83.84 ± 1.338	0.790 ± 0.010	0.893 ± 0.010
MMGA-Net w.o. ϕ	83.85 ± 2.030	0.790 ± 0.016	0.900 ± 0.019
MMGA-Net (ours)	**87.32 ± 0.384**	**0.840 ± 0.009**	**0.935 ± 0.010**

to extract hand-crafted features and CNN-driven features. MMGA-Net utilizes both CNN and GNNs to characterize tissue images. For GNNs, we quantify each nucleus by using 17 hand-crafted features and 1280 CNN-based features. Among 1297 features, we select 50 features that are most predictive for cancer types. In this regard, MMGA-Net can be considered a type of hybrid approach for cancer grading.

In pathology image analysis, tissue images are often processed and analyzed at multiple scales, resolutions, and levels. The effectiveness of such an approach has already been proved in literature (Kwak and Hewitt, 2017; Le Trinh et al., 2021). However, prior works of GNNs have mainly focused on representing and utilizing cell-to-cell interactions at a single level. In Pati et al. (2022), a hierarchical graph representation, which employs both a tissue graph and a cell graph, demonstrated a substantial performance gain in breast cancer subtyping. MMGA-Net builds cell graphs at multiple levels, including a global graph and two levels of local graphs, outperforming the other competing models. Similar results were also found for prostate cancer classification (Supplementary Table 3 and 4). The superior performance of MMGA-Net on colorectal, gastric, and prostate cancer datasets further emphasizes the importance of multilevel analysis in GNNs.

There exist various types of cells in tissues, which of each has unique functional and histological meanings. For instance, most tumors originate from epithelial cells, which lose cell-to-cell adhesion and cellular polarity in tumors (Coradini et al., 2011). A lymphocyte is a type of white blood cell that is an essential part of an immune system. Tumor-infiltrating lymphocytes (TILs) are, in particular, significantly associated with tumor prognosis (Elomaa et al., 2022; Kurozumi et al., 2019). However, none of the previous GNN-based approaches has differentiated and used differing cell types in constructing cell graphs, i.e., ignoring the functional and histological differences among them. MMGA-Net builds three types of cell graphs using epithelial cells only, lymphocytes only, and both epithelial cells and lymphocytes. In this manner, MMGA-Net exploits both intra- and inter-cell type interactions, leading to an improved representation and characterization of tissue/cellular structures for cancer diagnosis.

There are several limitations to this study. First, we built and tested MMGA-Net and other competing methods using image patches similar to previous studies (Chen et al., 2020; Pati et al., 2022; Zhou et al., 2019). A majority of image patches were obtained from WSIs; however, MMGA-Net has not been evaluated at the WSI level. In the follow-up study, we will extend MMGA-Net to directly process and analyze WSIs for cancer diagnosis. Second, we generated two sets of local graphs by setting the size of local regions to 512 × 512 pixels and 256 × 256 pixels, respectively. Varying the size, numerous local graphs can be built. Further optimizing the size and number of local regions for the graph construction will help to improve the graph representation of histological objects and pathology images. Third, two cell types, including epithelial cells and lymphocytes, were utilized in this study. There are several cell types, which of each has implications for cancer diagnosis. Other cell types can be integrated into GNNs to further extend our approach. Fourth, several nuclear features were selected to characterize individual cells before graph construction and graph-based learning as done in other studies (Chen et al., 2020; Pati et al., 2022; Zhou et al., 2019). Although such nuclear features have been known to be related to cancer diagnosis, the effect of each of the nuclear features on the graph representation and learning is unclear. From our observations, all three types of nuclear features (intensity-, texture-, and shape-based features) are substantially contributable to cancer diagnosis (see Supplementary Table 1 and 2). Fifth, we extracted tissue-level features by CNN and combined them with cell-level features that were obtained from GNNs at multiple scales. Similar to cell-level features, tissue-level features can be extracted by GNNs. We can further extend our method by incorporating tissue-level features obtained by GNNs and investigating the relationship between tissue- and cell-level features for cancer grading. Sixth, two sampling strategies were employed in our study to build cellular graphs. As we include >50% of cells that are dispersed in an image, there is still a chance that we miss some of the important cells for cancer grading. However, the experimental results were consistent across different datasets, suggesting that the two sampling strategies can include many of such important cells that are necessary for cancer grading. Even though we miss some important cells, it was the multi-cell graphs at multiple scales that can make proper decisions as shown in the experiments. Seventh, we utilized three types of cell graphs ($f^{G^{(1)}}$, $f^{G^{(2)}}$, $f^{G^{(3)}}$) to analyze tissue at multiple scales. The global cell graph ($f^{G^{(1)}}$) is, in particular, designed to investigate the global cell-to-cell interactions, i.e., long-range cell interactions, to analyze the macrostructure of tissues; however, it has not been assessed whether the global cell graph did exploit the long-range cell interactions and macrostructure of tissues in our study. The future study will entail further investigation and interpretation of the global cell graphs and cell interactions with respect to macrostructure of tissues. Eighth, we employed multiple cell type graphs to conduct cancer classification by fusing the prediction results of the cell graphs, which can be considered as a late fusion approach. However, there are other alternative ways to combine different cell type information such as incorporating cell type predictions into the nuclear features, i.e., an early fusion method, which is left for future research Last, a single gastric cancer dataset was employed. Additional datasets and experiments on them will further strengthen our findings on GNNs and gastric cancer diagnosis.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jin Tae Kwak reports a relationship with National Research Foundation of Korea that includes: funding grants.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant (No. 2021R1A2C2014557 and No. 2021R1A4A1031864). The authors would also like to thank Nhu Nhat Tan Doan for technical assistance and support.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2023.102936.

References

Anand, D., Gadiya, S., Sethi, A., 2020. Histographs: graphs in histopathology. In: Medical Imaging 2020: Digital Pathology, Vol. 11320. International Society for Optics and Photonics, p. 113200O.

Anklin, V., Pati, P., Jaume, G., Bozorgtabar, B., Foncubierta-Rodriguez, A., Thiran, J.-P., Sibony, M., Gabrani, M., Goksel, O., 2021. Learning whole-slide segmentation from inexact and incomplete labels using tissue graphs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 636–646.

Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., Polónia, A., Campilho, A., 2017. Classification of breast cancer histology images using convolutional neural networks. PLoS One 12 (6), e0177544.

Arvaniti, E., Fricker, K.S., Moret, M., Rupp, N., Hermanns, T., Fankhauser, C., Wey, N., Wild, P.J., Rueschoff, J.H., Claassen, M., 2018. Automated gleason grading of prostate cancer tissue microarrays via deep learning. Sci. Rep. 8 (1), 1–11.

Bejnordi, B.E., Zuidhof, G., Balkenhol, M., Hermsen, M., Bult, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J., 2017. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. J. Med. Imaging 4 (4), 044504.

Bulten, W., Kartasalo, K., Chen, P.-H.C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., Steiner, D.F., van Boven, H., Vink, R., et al., 2022. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the PANDA challenge. Nat. Med. 28 (1), 154–163.

Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F., 2022. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16144–16155.

Chen, R.J., Lu, M.Y., Wang, J., Williamson, D.F., Rodig, S.J., Lindeman, N.I., Mahmood, F., 2020. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans. Med. Imaging.

Cohen, J., 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol. Bull. 70 (4), 213.

Coradini, D., Casarsa, C., Oriana, S., 2011. Epithelial cell polarity and tumorigenesis: new perspectives for cancer detection and treatment. Acta Pharmacol. Sin. 32 (5), 552–564.

Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A., 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat. Med. 24 (10), 1559–1567.

Doan, T.N.N., Song, B., Le Vuong, T.T., Kim, K., Kwak, J.T., 2022. SONNET: A self-guided ordinal regression neural network for segmentation and classification of nuclei in large-scale multi-tissue histology images. IEEE J. Biomed. Health Inf..

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Doyle, S., Feldman, M.D., Shih, N., Tomaszewski, J., Madabhushi, A., 2012. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. BMC Bioinform. 13 (1), 1–15.

Elmore, J.G., Longton, G.M., Carney, P.A., Geller, B.M., Onega, T., Tosteson, A.N., Nelson, H.D., Pepe, M.S., Allison, K.H., Schnitt, S.J., et al., 2015. Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA 313 (11), 1122–1132.

Elomaa, H., Ahtiainen, M., Väyrynen, S.A., Ogino, S., Nowak, J.A., Friman, M., Helminen, O., Wirta, E.-V., Seppälä, T.T., Böhm, J., et al., 2022. Prognostic significance of spatial and density analysis of T lymphocytes in colorectal cancer. Br. J. Cancer 1–10.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch geometric. arXiv preprint arXiv:1903.02428.

Gamper, J., Alemi Koohbanani, N., Benet, K., Khuram, A., Rajpoot, N., 2019. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In: European Congress on Digital Pathology. Springer, pp. 11–19.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Henaff, O., 2020. Data-efficient image recognition with contrastive predictive coding. In: International Conference on Machine Learning. PMLR, pp. 4182–4192.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708.

Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning. In: International Conference on Machine Learning. PMLR, pp. 2127–2136.

Jaume, G., Pati, P., Bozorgtabar, B., Foncubierta, A., Anniciello, A.M., Feroce, F., Rau, T., Thiran, J.-P., Gabrani, M., Goksel, O., 2021. Quantifying explainers of graph neural networks in computational pathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8106–8116.

Javed, S., Mahmood, A., Fraz, M.M., Koohbanani, N.A., Benes, K., Tsang, Y.-W., Hewitt, K., Epstein, D., Snead, D., Rajpoot, N., 2020. Cellular community detection for tissue phenotyping in colorectal cancer histology images. Med. Image Anal. 63, 101696.

Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A., 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging 36 (7), 1550–1560.

Kurozumi, S., Inoue, K., Matsumoto, H., Fujii, T., Horiguchi, J., Oyama, T., Kurosumi, M., Shirabe, K., 2019. Prognostic utility of tumor-infiltrating lymphocytes in residual tumor after neoadjuvant chemotherapy with trastuzumab for HER2-positive breast cancer. Sci. Rep. 9 (1), 1–8.

Kwak, J.T., Hewitt, S.M., 2017. Multiview boosting digital pathology analysis of prostate cancer. Comput. Methods Programs Biomed. 142, 91–99.

Van der Laak, J., Litjens, G., Ciompi, F., 2021. Deep learning in histopathology: the path to the clinic. Nat. Med. 27 (5), 775–784.

Le Trinh, T.T., Song, B., Kim, K., Cho, Y.M., Kwak, J.T., 2021. Multi-scale binary pattern encoding network for cancer classification in pathology images. IEEE J. Biomed. Health Inf..

Le Vuong, T.T., Kim, K., Song, B., Kwak, J.T., 2021. Joint categorical and ordinal learning for cancer grading in pathology images. Med. Image Anal. 73, 102206.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Mahmood, H., Shaban, M., Indave, B., Santos-Silva, A., Rajpoot, N., Khurram, S., 2020. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review. Oral Oncol. 110, 104885.

Mari, C.R., Gonzalez, D.V., Bou-Balust, E., 2022. Multi-scale transformer-based feature combination for image retrieval. In: 2022 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3166–3170.

Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J., Shapiro, L., 2018a. Learning to segment breast biopsy whole slide images. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 663–672.

Mehta, S., Mercan, E., Bartlett, J., Weaver, D., Elmore, J.G., Shapiro, L., 2018b. Y-Net: joint segmentation and classification for diagnosis of breast biopsy images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 893–901.

Mercan, C., Aksoy, S., Mercan, E., Shapiro, L.G., Weaver, D.L., Elmore, J.G., 2019. From patch-level to ROI-level deep feature representations for breast histopathology classification. In: Medical Imaging 2019: Digital Pathology, Vol. 10956. SPIE, pp. 86–93.

Nguyen, K., Sarkar, A., Jain, A.K., 2014. Prostate cancer grading: use of graph cut and spatial arrangement of nuclei. IEEE Trans. Med. Imaging 33 (12), 2254–2270.

Niazi, M.K.K., Parwani, A.V., Gurcan, M.N., 2019. Digital pathology and artificial intelligence. Lancet Oncol. 20 (5), e253–e261.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in pytorch.

Pati, P., Foncubierta-Rodríguez, A., Goksel, O., Gabrani, M., 2021. Reducing annotation effort in digital pathology: A co-representation learning framework for classification tasks. Med. Image Anal. 67, 101859.

Pati, P., Jaume, G., Foncubierta-Rodríguez, A., Feroce, F., Anniciello, A.M., Scognamiglio, G., Brancati, N., Fiche, M., Dubruc, E., Riccio, D., et al., 2022. Hierarchical graph representations in digital pathology. Med. Image Anal. 75, 102264.

Potjer, F.K., 1996. Region adjacency graphs and connected morphological operators. In: Mathematical Morphology and Its Applications to Image and Signal Processing. Springer, pp. 111–118.

Roy, K., Banik, D., Bhattacharjee, D., Nasipuri, M., 2019. Patch-based system for classification of breast histology images using deep learning. Comput. Med. Imaging Graph. 71, 90–103.

Sudharshan, P., Petitjean, C., Spanhol, F., Oliveira, L.E., Heutte, L., Honeine, P., 2019. Multiple instance learning for histopathological breast cancer image classification. Expert Syst. Appl. 117, 103–111.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clin. 71 (3), 209–249.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.

Tellez, D., Litjens, G., van der Laak, J., Ciompi, F., 2019. Neural image compression for gigapixel histopathology image analysis. IEEE Trans. Pattern Anal. Mach. Intell. 43 (2), 567–578.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., Rao, X., Zheng, C., Zhang, F., 2020. Breast cancer histopathological image classification using a hybrid deep neural network. Methods 173, 52–60.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R.R., Smola, A.J., 2017. Deep sets. Adv. Neural Inf. Process. Syst. 30.

Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.-A., Rajpoot, N., 2019. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.