# Differentiation of pancreatic ductal adenocarcinoma and chronic pancreatitis using graph neural networks on histopathology and collagen fiber features

Bin Li [a,b], Michael S. Nelson [a], Omid Savari [c], Agnes G. Loeffler [d], Kevin W. Eliceiri [a,b,e,*]

[a] Department of Biomedical Engineering, University of Wisconsin-Madison, Madison 53706, WI, USA
[b] Morgridge Institute for Research, Madison 53705, WI, USA
[c] Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh 15213, PA, USA
[d] Department of Pathology, MetroHealth Medical Center, Cleveland 44109, OH, USA
[e] Department of Medical Physics, University of Wisconsin-Madison, Madison 53706, WI, USA

## ARTICLE INFO

## ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal human cancers. However, the symptoms and radiographic appearance of chronic pancreatitis (CP) mimics that of PDAC, and sometimes the 2 entities can also be difficult to differentiate microscopically. The need for accurate differentiation of PDAC and CP has become a major topic in pancreatic pathology. These 2 diseases can present similar histomorphological features, such as excessive deposition of fibrotic stroma in the tissue microenvironment and inflammatory cell infiltration. In this paper, we present a quantitative analysis pipeline empowered by graph neural networks (GNN) capable of automatic detection and differentiation of PDAC and CP in human histological specimens. Modeling histological images as graphs and deploying graph convolutions can enable the capture of histomorphological features at different scales, ranging from nuclear size to the organization of ducts. The analysis pipeline combines image features computed from co-registered hematoxylin and eosin (H&E) images and Second-Harmonic Generation (SHG) microscopy images, with the SHG images enabling the extraction of collagen fiber morphological features. Evaluating the analysis pipeline on a human tissue micro-array dataset consisting of 786 cores and a tissue region dataset consisting of 268 images, it attained 86.4% accuracy with an average area under the curve (AUC) of 0.954 and 88.9% accuracy with an average AUC of 0.957, respectively. Moreover, incorporating topological features of collagen fibers computed from SHG images into the model further increases the classification accuracy on the tissue region dataset to 91.3% with an average AUC of 0.962, suggesting that collagen characteristics are diagnostic features in PDAC and CP detection and differentiation.

## Introduction

Pancreatic cancer is one of the most lethal malignancies with a poor 5-year survival rate of around 5%–9% that has remained almost stagnant since the 1960s.[1,2] More than 85% of pancreatic cancers are pancreatic ductal adenocarcinoma (PDAC), which originates from pancreatic ductal epithelium in the pancreatic head.[2] Chronic pancreatitis (CP), though an entirely different disease with a distinct prognosis compared to PDAC, can present symptoms and a radiographic appearance similar to PDAC. The 2 entities can be difficult to differentiate microscopically, as the pathologist sometimes has only a small amount of tissue recovered by fine needle aspiration or, more rarely, needle biopsy, for examination.[3–7] Therefore, differentiating PDAC and CP is a challenge in pathology where the

consequence of misdiagnosis can be severe due to the rapid progression of PDAC and the high frequency of distant metastases.[8–10]

In clinics, histopathological assessment follows computed tomography, MRI or ultrasound to ensure the accurate diagnosis of PDAC and CP.[3–5,11,12] CP can present similar histomorphological features such as inflammatory infiltration, dense stroma, angulated glands, cytologic atypia, and tumor-like duct organization that mimic PDAC. The difficulty in distinguishing PDAC from CP is further aggravated by the fact that PDAC may induce CP in the surrounding pancreatic tissue, posing additional challenges to tissue sampling and diagnosis.[3–7]

One important hallmark of the tissue microenvironment for both PDAC and CP is a dense desmoplastic stroma, characterized by the increased deposition of fibrillar collagen. In PDAC, the extensive desmoplasia present

in the tumor may contribute to resistance to radiotherapy and hinder drug delivery.[13–15] Besides PDAC, evidence in a variety of cancer types has also shown that stromal properties are important factors for disease diagnosis, cancer progression, and tumor response to therapy.[16–21] In differentiating PDAC and CP, the collagen fiber topology at the stromal–epithelial interfaces has been shown to be a statistically significant discriminating feature.[22] Quantification of the size, shape, and patterns of collagen fibers show that these factors might impact tissue stiffness and are associated with an increased risk of cancer progression.[18,23] Many of the findings are derived from studies using label-free collagen-sensitive imaging modalities such as second-harmonic generation microscopy (SHG) and polarization-based optical microscopy, which can enable the quantification of collagen fibers and desmoplasia.[24–30]

The advent of digital slide scanners and whole slide imaging (WSI) has given rise to the opportunity to integrate quantitative algorithms into clinical workflows.[31–34] The potential for WSI-based computational analysis is increasing due to the recent success of deep neural networks in computer vision and language modeling.[35,36] Deep learning has already demonstrated state-of-the-art performance in tasks such as cancer detection, tissue segmentation, disease prognostication, and spatial omics analysis.[37–39]

In this paper, we present a computational workflow for detecting and differentiating histological samples of PDAC and CP, empowered by advances in artificial intelligence and collagen-targeted tissue imaging. Our method is motivated by the observation that discriminative histological features of PDAC and CP span multiple fields and multiple magnifications, from cell morphology (e.g., variation in epithelial nuclear sizes) to duct organization, and both diseases are characterized by dense stroma. Thus, a model with the ability to learn image features across different scales and incorporate collagen-based image features would be favored.

The analysis pipeline includes a deep learning model built on graph neural networks (GNNs)[40,41] that can be trained with manually classified, coarsely annotated regions of varying sizes, along with a region proposal algorithm that generates candidate regions in unannotated slides. Built on the canonical scheme of WSI data processing, the proposed method extracts local features from image patches but further models the patch-to-patch interactions by constructing graphs from the patch sets and utilizing graph convolutions.[41,42] The use of graph models for analyzing tissue structure was first suggested in Prewitt and Wu.[43] More recently, graph convolution and deep learning have become the building blocks of many state-of-the-art tissue analysis workflows. GNNs increase the expressivity of the model by letting the information flow between adjacent image patches, thus, capturing histomorphological features that span multiple patches.[44–46] This more closely mimics histopathological examinations conducted by pathologists where relationships between various tissue features are integral to diagnosis. Moreover, the analysis workflow incorporates multimodal image analysis where image features from hematoxylin and eosin (H&E) images and morphological features of collagen fiber from SHG images are registered and used as the input of the GNN.

The proposed GNN-based method was evaluated on a human tissue micro-arrays (TMAs) dataset and a human tissue section, both consisting of PDAC, CP, and normal pancreas tissue samples. The resulting model outperforms the widely used multiple-instance learning (MIL) framework,[47,48] on both datasets. Furthermore, we demonstrated that incorporating collagen-based features extracted from SHG images leads to higher classification accuracy than using brightfield H&E features alone confirming the diagnostic potential of utilizing characteristic collagen topology in differentiating PDAC, CP, and normal pancreas tissue.

## Material and methods

### Sample sets and pathology annotations

#### Human tissue microarrays

Seven H&E stained formalin-fixed paraffin-embedded human pancreas TMAs constructed from 481 cases were purchased from US Biomax (Rockville, MD, USA), the slide IDs used were: BIC14011b (24 cases), HPan-Ade180Sur-01 (90 cases), HPanAde170Sur-01 (90 cases), HPan-Ade120Sur-01 (90 cases), PA2081b (96 cases), BBS14011 (48 cases), and PA485 (43 cases). The TMAs consist of cores from 1.0 to 1.5 mm in diameter and were reviewed by 2 US Biomax pathologists. All slides were scanned with an Aperio CS2 (Leica Biosystems, Wetzlar, Germany) at $20\times$ magnification, with a pixel size of 0.504 μm/pixel. All PDAC (380 cores), chronic pancreatitis (193 cores), and normal (213 cores) TMA cores from the slides were used for analysis and image classification. The TMA form factor was selected to gain access to a large number and broad range of samples that would otherwise be difficult to obtain with traditional histopathology sections, with the tradeoff being smaller tissue regions to sample.

#### Human tissue sections

24 H&E-stained tissue section slides were obtained from Metro Health Medical Center (Cleveland, OH, USA) with IRB approval. The tissue sections were obtained from surgical specimens, including Whipple procedures and distal pancreatectomies, and from autopsies. The slide set contains tissue sections from 12 cases of PDAC and 12 cases of chronic pancreatitis. Slides were reviewed and annotated by 2 pathologists (Omid Savari and Agnes G. Loeffler). For the PDAC slides, roughly 6 regions containing PDAC, 6 regions containing chronic pancreatitis, and 3 regions containing histologically normal pancreas were coarsely marked with rectangles; for chronic pancreatitis slides, on average 3 regions containing chronic pancreatitis and 6 regions containing normal-appearing pancreas were coarsely marked with rectangles. Note that since the regions were coarsely annotated, a PDAC region might include chronic pancreatitis sub-regions, and a PDAC or a chronic pancreatitis region might contain sub-regions with normal tissues. The annotations were made using Aperio ImageScope (12.4.3) software (Leica Biosystems, Wetzlar, Germany) on brightfield images scanned by an Aperio CS2 digital pathology slide scanner (Leica Biosystems, Wetzlar, Germany) at $20\times$ magnification, with a pixel size of 0.504 μm/pixel. The annotation process produced a total of 70 PDAC regions, 112 chronic pancreatitis regions, and 86 normal regions.

### Second-harmonic generation microscopy imaging of samples

Annotated regions from the tissue section slides were imaged using a custom-built multiphoton laser-scanning microscope (LSM) that has been previously described.[49] In brief, a Tsunami Ti:Sapphire laser (Spectra-Physics, Santa Clara, CA, USA) tuned to 800 nm, with a pulse length of approximately 100 f. was focused on the sample by a 20X/0.75NA air immersion objective (Nikon, Melville, NY, USA). The SHG signal was collected in the forward direction, filtered by a narrow bandpass filter centered at $400 \pm 10$ nm (ThorLabs, Newton, NJ, USA), onto a H7422-40P photomultiplier tube (Hamamatsu, Hamamatsu, Japan). The resulting pixel size of the SHG image is 0.509 μm/pixel. The scanning process and data acquisition signal were controlled by our custom laser-scanning software, OpenScan. The acquisition workflow utilizes the open-source software QuPath,[50] Micro-Manager, and Pycro-Manager.[51,52]

### Graph neural networks

### Graph construction

*Image registration, pre-processing, and normalization.* Brightfield images of individual TMA cores and annotated regions were cropped from the Aperio scans. Brightfield images of annotated regions in the tissue section set were registered with the corresponding SHG images using the automatic H&E to SHG registration algorithm described in Keikhosravi et al.[53] Image pairs that failed to be registered by the algorithm were further registered using the method described in Pielawski et al[54] or manually registered by corresponding landmarks in Fiji.[55]

Each brightfield image was tiled into patches with a size of $256 \times 256$ pixels (without overlap); patches were only extracted within tissue regions. The tiles were then normalized using the H&E image normalization

algorithm described in Macenko et al[56] with batch processing. Every batch contained 256 patches sampled from all patches. The normalization algorithm then treated the 256 patches as a single input image and performed normalization. Running the algorithm in batches of patches helped stabilize the normalization by evenly sampling the whole dataset to include more types of tissue components while keeping the number of pixels relatively small. It helps avoid the generation of unbalanced pixel values in a single TMA or annotated region that might contain insufficient variety.

Registered SHG images of annotated regions in the tissue section set were normalized by stretching the pixel values between 5 and 99.5 percentile to the full dynamic range. The images were then tiled into patches with a size of $256 \times 256$ pixels (without overlap), resulting in SHG image patches with pixel-wise correspondences to the brightfield image patches.

*Patch feature extraction.* A feature vector with a shape of $512 \times 1$ was then extracted from every brightfield patch using the ImageNet pretrained ResNet18.[57] For each SHG patch, collagen fiber-specific features such as width, length, and angle were extracted using CurveAlign and CT-FIRE.[58] Fiber segments, alignment coefficients, numbers of fibers, circular mean and variance of fiber angles, mean and variance of fiber lengths, mean and variance of fiber straightness, and mean and variance of fiber widths were finally computed and formed a feature vector with a shape of $10 \times 1$ for each patch. The features were normalized to a range from 0 to 1.

*Constructing graphs from images.* The image of a TMA core or an annotated region was treated as a single graph, and the patches tiled from the image were treated as nodes of the graph. Each patch (node) was connected with the nearest neighboring patches (if they existed) with edge weights equal to the cosine similarity between the feature vectors of the 2 nodes. The information about a tile (a node feature) consisted of the corresponding $512 \times 1$ brightfield feature vector in the case of the brightfield graph or the corresponding $10 \times 1$ SHG feature vector in the case of the SHG graph. A brightfield-SHG graph was constructed by concatenating the brightfield and SHG feature vectors as the node features (with a vector size of $522 \times 1$).

### Graph convolutional network architecture

*Feature pre-processing layer.* The node features were first processed by a fully connected neural network layer followed by a parameterized ReLU (PReLU).[59] For the brightfield graph and brightfield-SHG graph, the fully connected layer produced an output vector with a size of $512 \times 1$. For the SHG graph, the fully connected layer produced an output vector with a size of $32 \times 1$. The processed features were then fed to the following graph convolution layers.

*Graph convolution network (GCN).* Let V be the set of nodes and $\mathbf{h}_v$ where $v \in \mathcal{V}$ be the node feature vector processed using the procedures described above. The updated note feature $\mathbf{h}'_v$ after a 1-hop graph convolution operation[41] is given by

$$\mathbf{h}'_v = \text{PReLU} \left( \frac{1}{\|\mathcal{N}(v)\|} \sum_{u \in \mathcal{N}(v)} \Theta \mathbf{h}_u + \mathbf{b} \right), \forall v \in \mathcal{V} \tag{1}$$

where $\Theta$ is a weight matrix and b is a bias vector, and $u \in \mathcal{N}(v)$ denotes the neighbors of node $v$ (including $v$ itself). Three 1-hop graph convolution layers are stacked and the update $\mathbf{h}'_v$ of each layer is passed to the next graph convolution layer. The resulting embedding $\mathbf{h}_o$ then represents information up to 3-hop neighbors from the node. Besides, another type of graph convolution operation was evaluated in the design. This alternative model used a single-layer 3-hop simplified graph convolution (SGC) operator that directly aggregates information in the 3-hop neighborhood of the nodes without non-linearity between each hop.[42] This operation was more memory and computationally efficient and empirically showed comparable

performance to the standard graph convolution. $\mathbf{h}_o$ was concatenated with $\mathbf{h}_v$ to form a new node embedding as output.

*Readout layer.* The output node embeddings of each graph were averaged to produce a graph embedding. The graph embedding was scored by a linear classification head followed by a sigmoid function to produce a graph prediction with the same shape as the graph label representation using one-hot encoding. The overall analysis pipeline, including the network training procedure, is shown in Fig. 1. The objective function with respect to a dataset containing $N$ image-label pairs can be written as:

$$\{\theta, \mathbf{W}\} = \arg \min_{\theta, \mathbf{W}} \frac{1}{N} \sum_{i=0}^{N-1} \mathcal{L}(\mathbf{W}f(\text{G}_i; \theta), \mathbf{y}_i) \tag{2}$$

where L is the loss function, $\theta$ is the set of weights parameterizing the graph convolution network $f$ described above, $\mathbf{W}$ is the weight matrix of a linear layer, and $\text{G}_i$ and $\mathbf{y}_i$ are a graph constructed from an image and its class label, respectively.

*Class coding and loss function.* PDAC can induce CP in surrounding areas and areas with a mixture of the 2 are also often observed. However, when both CP features and PDAC features are present in the same region, the sample will be labeled as PDAC. Therefore, we used distributed coding instead of 1-hot coding for class labeling to account for the potential co-existence of both classes in a sample. Specifically, a normal region or core is labeled as [0,0], a CP region or core is labeled as [0,1], and a PDAC region or core is labeled as [1,1]. The final class label is obtained using the following rule: a prediction is regarded as normal when both digits are 0, CP when the first digit is 0 and the second digit is 1, and PDAC when the first digit is 1 (regardless of the second digit). The loss function is multi-label binary cross-entropy loss. The optimal thresholds for the 2 digits are obtained by finding the threshold leading to the best true-positive to false-positive ratio in the validation set for each digit.

### Network training specifics

The networks were optimized with stochastic gradient descent with the Adam solver and the initial learning rate was set to 0.0001.[60] The learning rate was scheduled with cosine annealing without warm restart.[61] The batch size was 1. For the TMA dataset, the cores were randomly separated into 4 equal-size chunks and a 4-fold cross validation was carried out. For the tissue section dataset, the slides were first randomly separated into 4 equal-size sets, and the annotated areas were then cropped from the slides to form 4 sets for a 4-fold cross validation. Cross-validations were repeated 3 times for each dataset, each time with a new random split.

### Region proposal in whole slide images

A region proposal pipeline was used to identify candidate regions in WSIs of tissue sections. This pipeline took coarsely annotated target regions as inputs and searched for similar regions in WSIs based on the distribution of patches in the annotated regions. The overall process can be seen as a task of slide region retrieval for some annotated regions as queries.

Firstly, patches with a size of $256 \times 256$ were extracted from the training WSIs of the tissue section set without overlap, and the patch feature vectors were extracted by a ResNet18 pre-trained on ImageNet. A k-means clusterer $\mathcal{K}$ was trained on the feature vectors using mini-batch k-means.[62] Secondly, the same patch extraction and feature extraction processes were applied to the annotated regions of PDAC and chronic pancreatitis, and the feature vectors of the patches were assigned to the closest centroid of $\mathcal{K}$. The distribution of centroid assignments was computed and represented in a vector $\mathbf{q}$. The idea of learning a dictionary of pattern basis is frequently used in fine-grained image classification.[63] Thirdly, a sliding window of size 8 patches $\times$ 8 patches was used to scan through
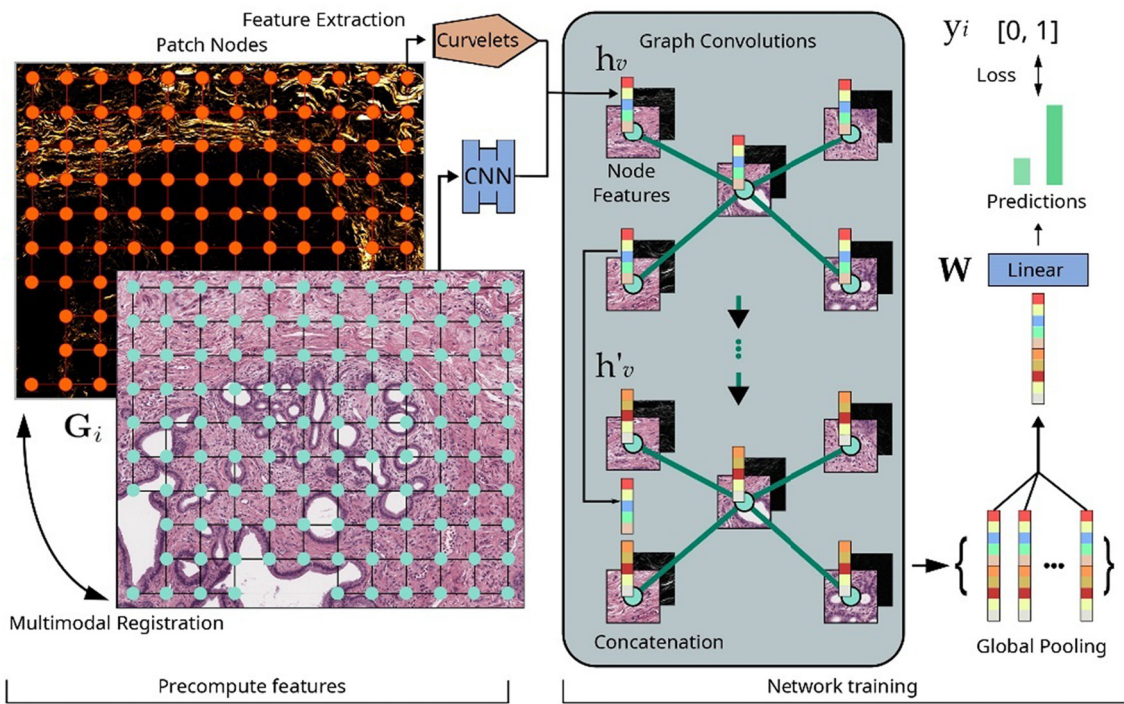
**Fig. 1.** Analysis pipeline including multimodal image registration, patch extraction, feature computing, graph construction, and GNN training.

the WSIs, and the distribution of patch feature vector assignments regarding the centroids of $\mathcal{K}$ within the window was computed per sliding step. The distribution of the assignments was denoted as $\mathbf{p}_{ij}$, where $i, j$ are the coordinate of the window center. The cosine similarity between $\mathbf{p}_{ij}$ and $\mathbf{q}$ was computed and placed at $(i, j)$. A similarity map was produced after the window slides through the entire WSI. Lastly, a multi-threshold Otsu's method was applied to the similarity map to separate the similarity values into 3 groups corresponding to background, low-similarity, and high-similarity

areas.[64] The high similarity regions were processed by a compact watershed algorithm that grouped local high-similarity areas into superpixels.[65] The bounding boxes of the superpixels were used as the region proposals. The region proposal algorithm is summarized in Algorithm 1 (Supplementary material).

The region proposals generated from testing slides were converted into graphs using the procedure described above and scored by the trained GNN for predictions. The overall workflow is illustrated in Fig. 2.
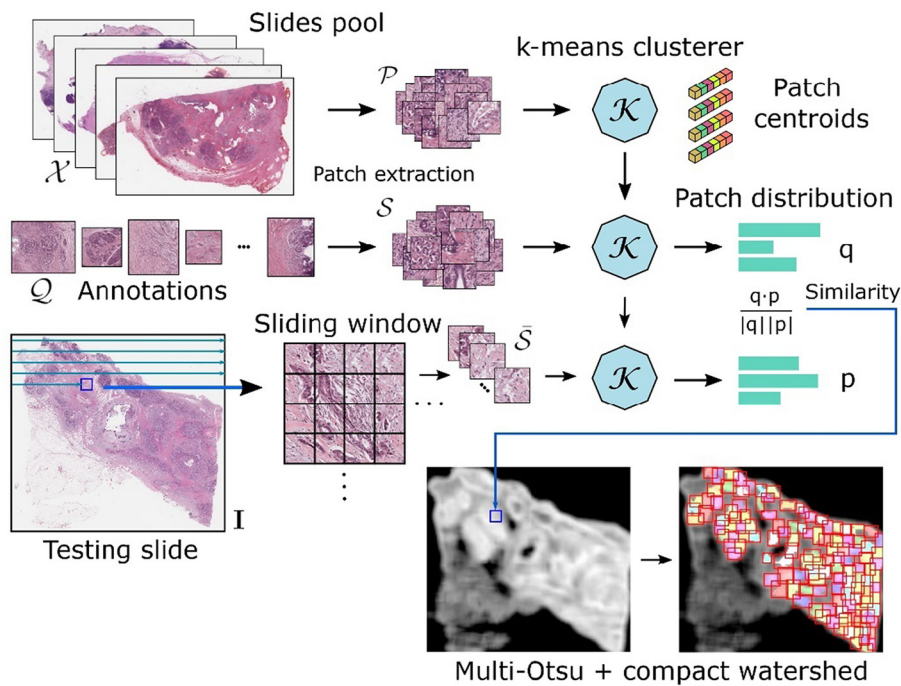


**Fig. 2.** Generation of region proposals from a WSI. Patches are extracted from WSIs and their feature vectors are computed and used to train a k-means clusterer. The trained k-means clusterer is used to predict the cluster assignments of input patches extracted from an ROI. The similarity of the assignment distributions between annotated ROIs and a sliding window is then computed.
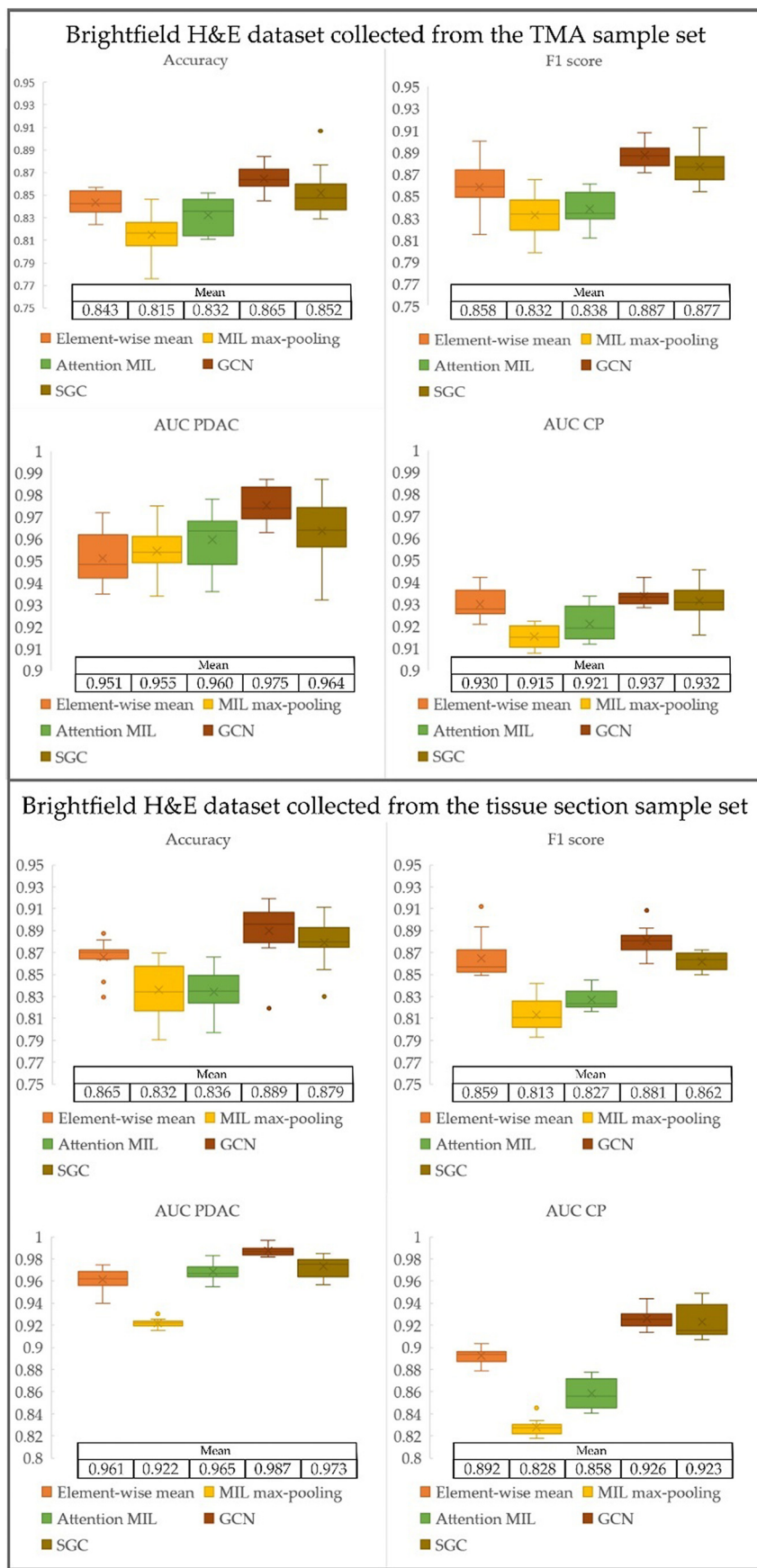
**Fig. 3.** Classification performance on brightfield datasets. Element-wise mean: apply element-wise mean-pooling on the node embeddings, followed by a fully connected layer. MIL max-pooling: score each node embedding with a fully connected layer, and the resulting scores are processed by a max-pooling operator. Attention MIL: attention-based MIL networks.[46] GCN: graph convolution networks.[41] SGC: simplified graph convolution networks.[42]

## Results

*Classification of TMA dataset*

*Baselines*

The proposed GNN-based model was compared to 3 other baseline methods, including a simple model with multiple-layer perceptron (MLP) followed by element-wise mean-pooling, max-pooling MIL, and a recent attention-based MIL model.[47] The max-pooling and mean-pooling MIL used the standard assumption of MIL by treating the patches of a TMA core as a bag of instances. The embeddings of the patches of a core were element-wise pooled using mean-pooling (element-wise averaging) to form a bag embedding, and a linear classification (fully connected layer) then scored the bag embedding to reach a core-level prediction. In max-pooling MIL, a linear classification head was used to score each patch embedding, and the resulting scores were averaged to produce the final score. The attention-based MIL used attention-based pooling on the embeddings of the patches to form a bag embedding by computing a weighted sum of the patch features.

*Evaluation metrics*

Average accuracy, average F1 score, and area under the curve (AUC) were used to compare the classification performance. The mean scores were computed from a 4-fold cross validation following the scheme described in the Materials and Methods section. The classification performance of the models on brightfield images of the TMA dataset is summarized in Fig. 3 (top panel). Both GCN and SGC outperform patch-wise models, including MLP and MIL. GCN shows the best classification performance with an accuracy of 86.5%, an F1 score of 0.887, and an average AUC of 0.954.

*Classification of tissue section dataset*

*ROI classification performance*

The same evaluation process was conducted on the brightfield images of the tissue section dataset with results summarized in Fig. 3 (bottom panel). The trend of the classification performance of different methods is consistent with the TMA dataset. GCN achieves an accuracy of 88.9%, an F1 score of 0.881, and an average AUC of 0.957, outperforming MIL. We give a brief explanation of the difference in model performance in the discussion section.

*Region proposal and whole slide scoring*

The described WSI region proposal method was evaluated on the whole tissue dataset. Region proposals were generated and scored following the scheme described in the Materials and Methods section. Representative
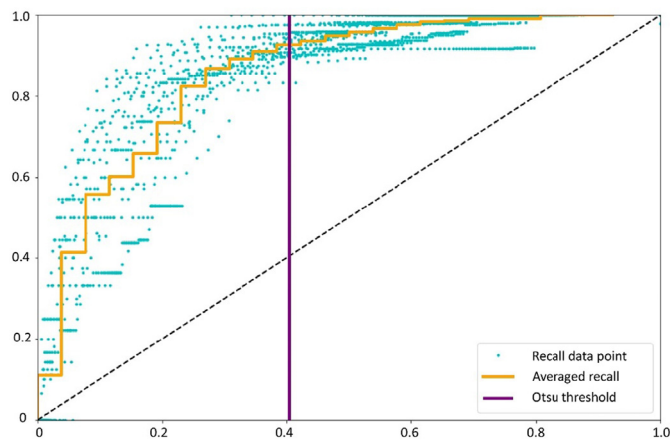


**Fig. 5.** Recall against the coverage ratio. Vertical axis: Recall; Horizontal axis: Coverage ratio. The coverage ratio is binned to 25 discrete values. The threshold returned by multiOtsu's method is marked in purple.

results are shown in Fig. 4. The effectiveness of the region proposal method was also quantitatively evaluated. The purpose of the region proposal step is to extract candidate regions from unannotated WSIs that cover positive regions as much as possible while keeping the total number of regions needed to be scored by the downstream classification model small.

Fig. 5 plots the recall against the ratio of the area in the WSI covered by region proposals. A region proposal is considered true positive if it overlaps with any of the annotations (PDAC and chronic pancreatitis). False negatives are defined as the total area of the annotated regions that are not covered by any region proposals divided by the average size of region proposals. Intuitively, decreasing the threshold that selects for high similarity areas results in more region proposals; a threshold of 0 leads to a set of region proposals that covers the whole WSI with a recall of 1. The ratio of area covered by the region proposal and the corresponding threshold returned by multi-Otsu's method is marked with a purple line in the figure. A set of good region proposals should have a high recall of disease-positive region coverage while keeping the overall tissue coverage relatively small. Overall, the region proposals achieve an average recall of 0.909 while covering 40.1% of the tissue regions.

*Effects of incorporating collagen features*

*Classification using only collagen features extracted from SHG*

The same GNN model was used on the SHG images collected from the tissue section dataset. The node features of the graph were collagen-based
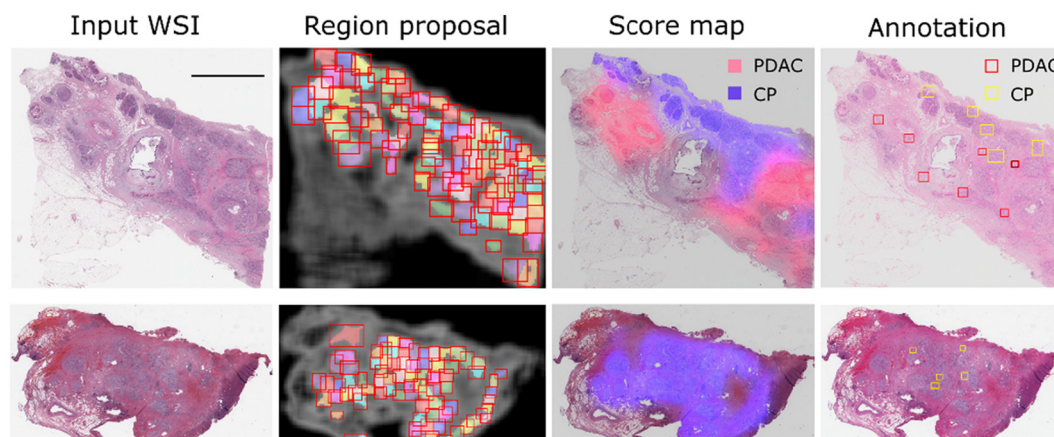


**Fig. 4.** Two examples of region proposal generation and detection results. Top row: a PDAC slide. Bottom row: a CP slide. First column: Input WSI; Second column: Region proposals; Third column: Detection map; Last column: Annotations (coarsely annotated by a pathologist). The scale bar is 6 mm.
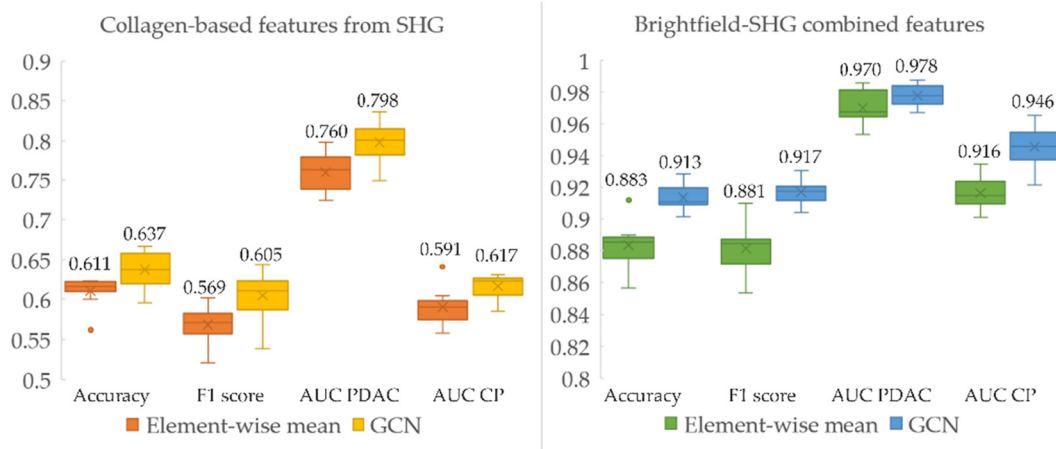
**Fig. 6.** Classification performance on SHG/Brightfield-SHG dataset collected from the tissue section sample set.

features extracted from the SHG images using CurveAlign/CT-FIRE. The classification results are summarized in Fig. 6, showing that collagen-based features alone are diagnostic and the trained model is significantly better than a random classifier.

*Classification using combined features from brightfield and SHG*

To further demonstrate the diagnostic value of collagen-based features, brightfield-SHG graphs were constructed by concatenating the brightfield feature vector and the SHG feature vector as node features to train the GNN model. The SHG image and brightfield H&E image of an annotated region were first co-registered with pixel-wise correspondences such that the tiled patches of the 2 remained co-registered. The resulting concatenated feature vector contained brightfield features and SHG features of the

same tissue area. Representative co-registered brightfield H&E images and SHG images of the three classes are shown in Fig. 7.

The evaluation results suggest that combining the brightfield H&E features with collagen-based features leads to improved performance for both the complex GNN-based model and the simple MLP (Fig. 6). Specifically, the GNN model using combined features achieves an accuracy of 91.3% while the accuracy using brightfield features alone is 88.9% (Fig. 6).

*Feature interpretability*

*ResNet extracted H&E features*

The node features of H&E patches were directly computed from ImageNet pre-trained ResNet18, as suggested in Lu et al[66]; this could be an efficient and practical way to extract features from WSI patches. To
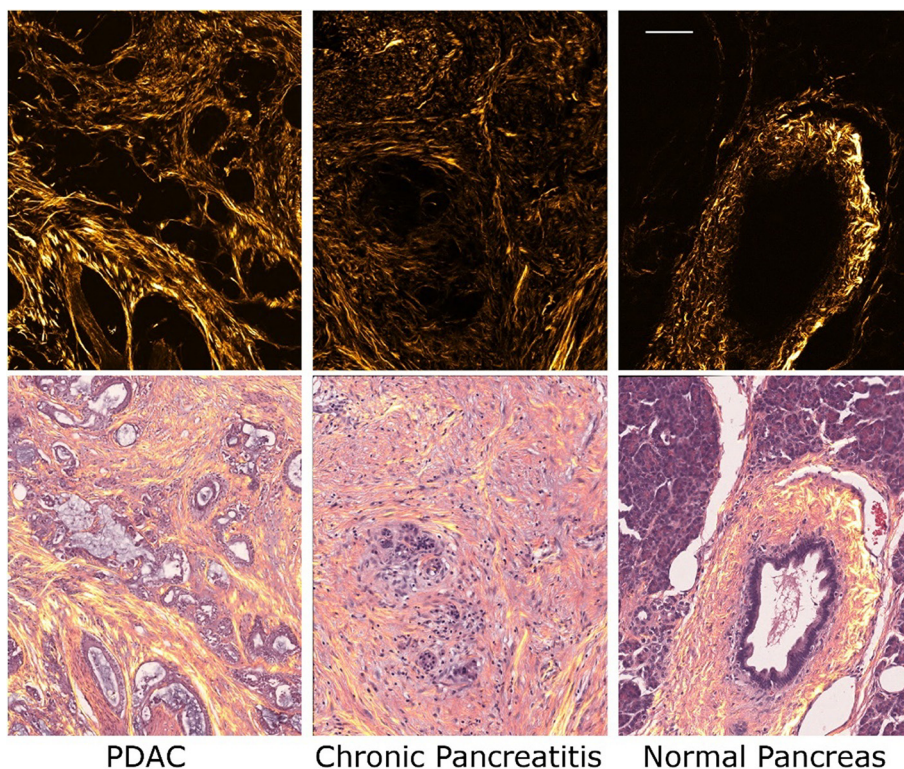


**Fig. 7.** Representative annotated regions of PDAC, chronic pancreatitis, and interlobular duct in the background of normal-appearing acinar tissue (normal pancreatic tissue). First row: SHG images. Second row: Overlay view of registered SHG and corresponding H&E images. The scale bar is 100 μm.
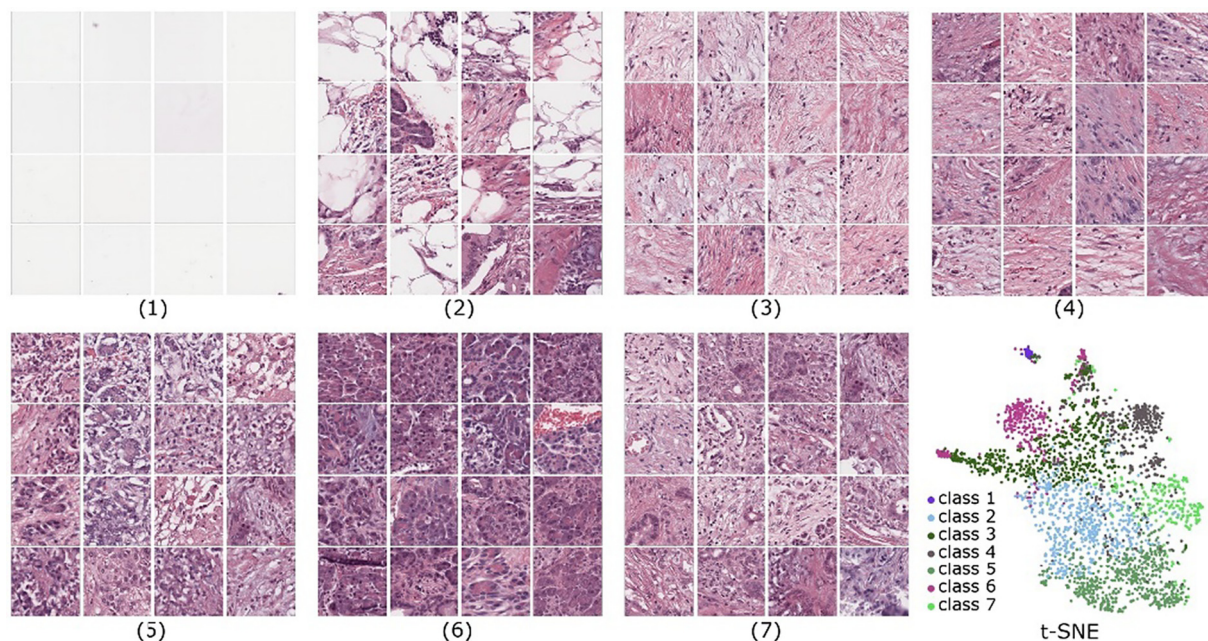
**Fig. 8.** Top-16 patches closest to the k-means centroid for each cluster (1)–(7) and the 2-dimensional t-SNE embedded plot of the clusters. class (1): Blank areas (hollow intra-gland areas); class (2): Mostly adipose tissue; (3–4): Stroma and collagen without any obviously malignant cells; class (5): Mostly chronic inflammation, sometimes in a collagenous stroma and sometimes admixed with acinar tissue; class (6): Benign tissue, including acini, islets of Langerhans, and ducts; class (7): Cancer-appearing duct parts with disorganized collagen, some contains stroma with acute inflammation or chronic inflammation.

validate that the computed features are discriminative, the top 16 patches closest to each k-means centroids are shown in Fig. 8 in the case of $k = 7$. The feature vectors of patches are also embedded in 2 dimensions using t-SNE[67] and labeled with k-means clustering results in the figure. The choice of k = 7 was experimentally determined. Our pathologist validated the following: the patches cluster of class 1 corresponds to blank areas; most of the patches in class 2 have adipose tissue; patches in class 3 and class 4 contain stroma and collagen without any obviously malignant cells; patches in class 5 mostly contain chronic inflammation, sometimes in a collagenous stroma and sometimes admixed with acinar tissue; class 6 contains benign tissue, including acini, islets of Langerhans, and ducts; patches in class 7 contain cancer-appearing duct parts with disorganized collagen, and some contain stroma with acute inflammation or chronic inflammation.

### Collagen-based features

In a previous study where collagen-based features were used for a simple binary linear classifier on PDAC and CP, the importance of each collagen-based feature was evaluated separately.[22] Here, we show an alternative way to quantify the importance of the input feature in the setting of the GNN model. We removed one of the input features (for width, angle, straightness, length, both the mean and variance were removed) and
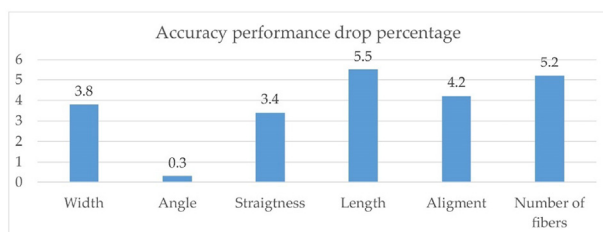


**Fig. 9.** Performance drop after removing one of the features. The performance drop is measured by averaging all the metrics. A larger performance drop (higher value in the chart) indicates that the removed feature uniquely captures information that the other feature do not and has higher impact in the discriminative process (e.g., the length of the collagen fibers).

trained the same GNN to perform classification, and observed the performance percentage drop, as shown in Fig. 9. A more significant performance drop after the removal suggests that the removed features are more important.

### Discussion and future work

*Importance of graph models in pancreas histomorphology*

The results show that modeling WSI regions as graphs is beneficial for building a computational classifier for the differential diagnosis of PDAC and CP. A potential reason for the accuracy improvement is that the diagnostic features of PDAC and CP span multiple patches and are difficult to capture using patch-wise representations. In the histomorphology of pancreatic cancer, tumor regions are defined as malignant epithelium arranged in atypical ducts, clusters of cells, or individual cells infiltrating into the surrounding tissue. The diagnostic criteria range from nuclear to architectural, including large variations in epithelial nuclear size, incomplete epithelium around ducts, angular and branching ducts, disorganized duct distribution, and ducts appearing in architecturally aberrant locations. Many of these features overlap with those of CP, including the increasing deposition of fibrillar collagen and desmoplastic stroma in the tissue microenvironment. These issues make obtaining expressive features from the image data in both patch-wise methods and MIL-based methods a challenge. Even though many MIL-based methods use bag-of-patch representations and learn a joint representation of all image patches, they could still perform incompetently as the spatial relations of the patches are not modeled and the features to aggregate in MIL operators are restricted to the field-of-view of a single patch.[47,48] Though patches extracted from multiple magnifications can be used in MIL frameworks, the receptive field is still fixed for each magnification, and the spatial relation between patches in the same magnification is not considered (See Fig. 10).

By contrast, GNNs model the patch-to-patch interactions by deploying message passing between adjacent patches and updating the node features to encode information from the neighboring patches. The updated node features can eventually incorporate features up to $k$ hops away, determined by the number of graph convolution layers, forming representations that
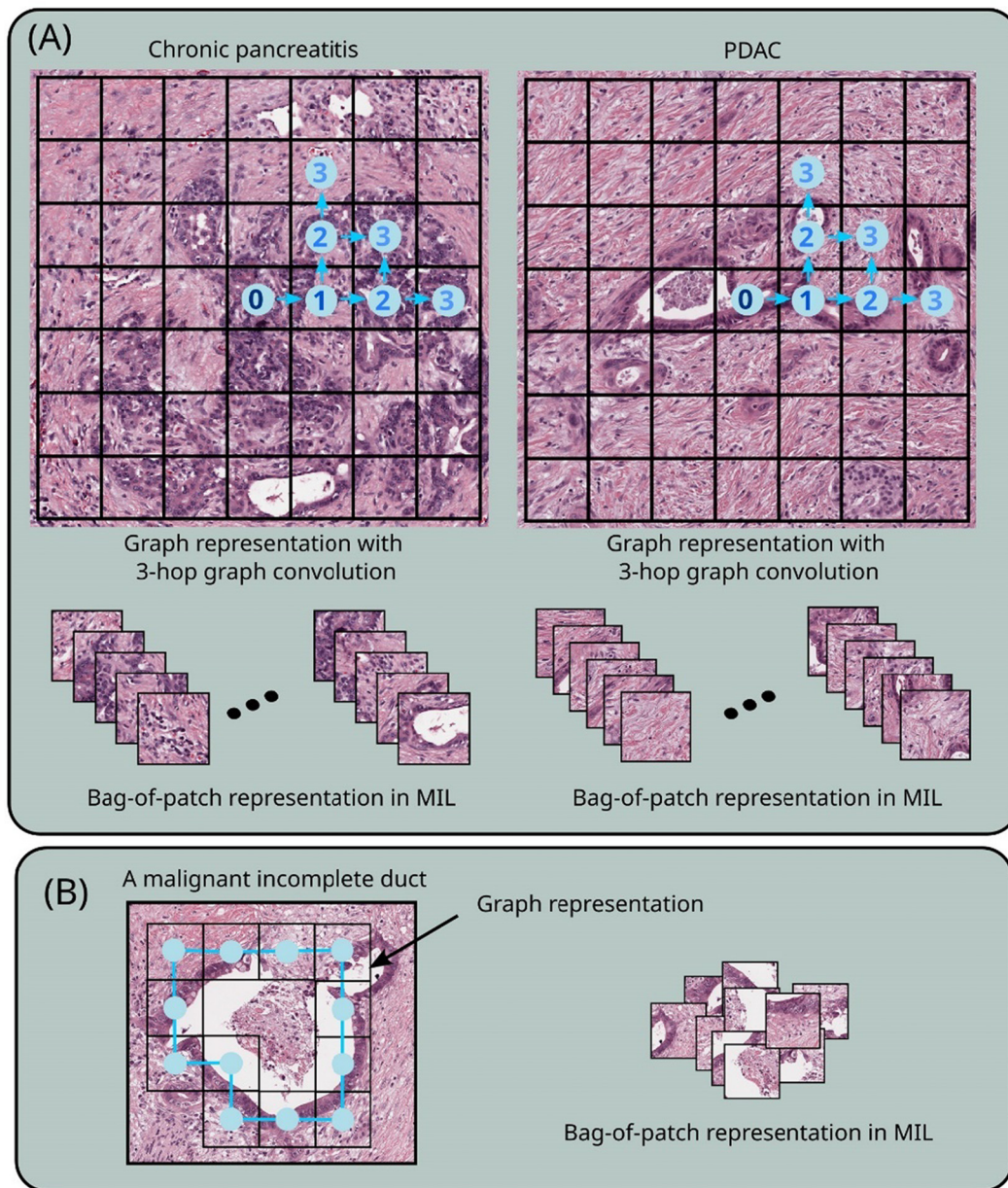
**Fig. 10.** (A) Comparison of feature extraction by graph representation in GNN and bag-of-patch representation in MIL. (B) An incomplete duct of PDAC (one hallmark of malignancy) could be under-represented in MIL because the shape of the whole duct is lost in the bag-of-patch representation.

correspond to a larger field of view. For example, the GNN could potentially learn the features of duct organization or an incomplete duct by propagating the information of a tile with disrupted epithelium to the adjacent tiles that form a duct object, as illustrated in Fig. 10. Therefore, compared to MIL-based models, a GNN-based model could capture the spatial arrangement and relationship of malignant cells to each other, to the stroma, and to other tissue elements. It is interesting that this kind of carry-over or comparison from field to field begins to approximate the manner in which pathologists evaluate or "read" slides in the clinical setting.

### Construction of graphs from histopathological images

One important step in using graph models in WSI analysis is the construction of graphs from images. The proposed method follows a nearest neighbor scheme where each patch is connected to its non-empty adjacent patches. Though easy to implement, this method does not exploit the nested structure of tissue components and redundant texture-like information that

is often spatially continuous. More complex approaches such as multiscale quadtree approximation[68] and constructing refined graphs based on the results of patch-wise spatial agglomeration[69,70] could potentially remove redundancies and result in graphs that encode the tissue structure and relations of components in more efficient ways.

### Model the interaction between histopathology modalities

Examination of H&E stained slides is the standard clinical histopathological approach to assess surgically excised pancreatic tissue. Special modalities such as histochemistry or immunohistochemistry are seldom used to differentiate between benign (CP) and malignant tissue. However, the research applications of multimodal histopathology have been rapidly advancing with promising results that could benefit clinical histopathology.

Given our prior knowledge of the importance of fibrillar collagen in the tissue microenvironment of PDAC and CP,[22] we incorporated collagen-based features extracted from SHG images, superimposed on the original
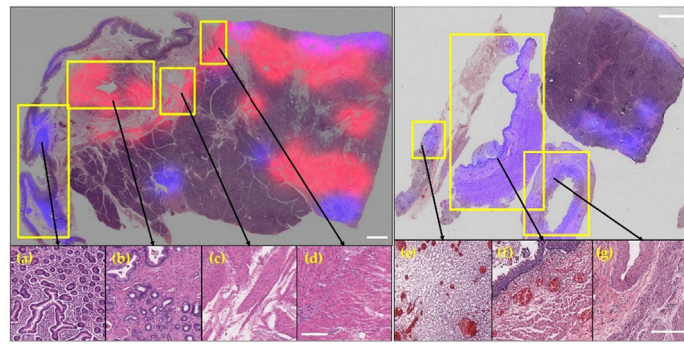
**Fig. 11.** Out-of-distribution regions can lead to false-positive detection results. Representative out-of-distribution regions are marked with yellow boxes. Left panel: a slide that contains malignant ducts, CP regions, and smooth muscles. Right panel: a slide from autopsy that contains CP regions as well as tissue pieces from other organs. (a): small intestinal mucosa. (b): normal pancreatic bile duct (near the ampulla of Vater). (c)–(d): small intestinal smooth muscle. (e): stomach. (f): esophagus. (g): gallbladder. The scale bar in the top row is 2 mm, and the scale bar in the bottom row is 200 μm.

H&E images, into the analysis. Even with the collagen fiber features alone, the GNN model achieves classification accuracy significantly better than a random classifier, showing that the features are discriminative for differentiating PDAC from CP. In the multimodal analysis pipeline, the feature vector pairs of registered image patches from both imaging modalities are concatenated to form the input to the model. After combining the collagen-based features from SHG images with the features from the corresponding H&E images, the downstream GNN achieves better accuracy in classification, as expected. To fully leverage the strength of GNNs, one could treat the features of another modality also as graph nodes and build a graph, while also connecting the nodes of 2 modalities that are paired spatially. Note that the inter-modality edge weights can be learned separately from the weights that connect the nodes spatially, forming an architecture. This could grant GNNs the flexibility to let information flow and interact between modalities.

Furthermore, recent works using image synthesis have shown that cross-modality image-to-image translation can be used to map images of one imaging modality to another.[75–76] A similar approach has been applied to synthesizing an SHG image from standard H&E images.[48] Future exploration will build on this method with necessary model generalizability improvement to generate high-quality SHG-like synthetic images from brightfield images of H&E-stained slides and apply CurveAlign/CT-FIRE to extract the topological features of collagen fibers. The advantage of such a method is that the morphological characteristics of the collagen fibers can be derived from H&E images alone, and thus avoid the use of laser scanning SHG imaging, which is not commonly available in the clinic. Another future direction is to incorporate information from patient demographics and radiographics that could provide more systematic views of the patient's condition.

*Out-of-distribution samples*

The quality of tissue sampling is critical to both visual inspection and computational analysis in histopathology. For PDAC and CP detection, low-quality tissue sampling or slide preparation could lead to inadequate tissue regions that fail to support an accurate diagnosis. For example, if the pancreas lesion is not sampled to get adequate epithelium in fine-needle aspiration, the cytopathologist would not be able to make a diagnosis of adenocarcinoma. In computational analysis, tissue sampling could affect the distribution of training samples. Thus, if the tissue components included in the training set are not exclusive, the resulted classifier can encounter out-of-distribution tissue regions and give over-confident false-positive results.

In our experimental setting, the pancreas tissue sections are annotated coarsely with PDAC, CP, or normal regions. However, when applying the classifier to all candidate regions extracted from the entire tissue on a slide, the classifier predicts some false-positive regions (PDAC or CP) with high confidence. These were areas of smooth muscle (duodenal muscularis propria), or tissue from another organ entirely (mostly from autopsy cases). These were not annotated by the pathologist because they were intuitively not "regions of interest." For the classifier, however, these were clearly out-of-distribution samples that had not been previously labeled and were therefore erroneously predicted as malignant (Fig. 11). This could be related to a known artifact of discriminative classifiers including deep neural networks.[71] The discriminative distribution density saturates at 1 even for the case where the input value to the last Sigmoid layer is arbitrary large which could be caused by an out-of-distribution sample that is very far away from any of the correct class distributions. Therefore, to facilitate the real-world applications of deep learning models in assisting clinical histopathology, it is important to incorporate uncertainty estimation into the classifier[72] or construct generative classifiers where out-of-distribution inputs could be detected.[73] The uncertainty estimation results can be used to build an active learning workflow where uncertain inputs are presented to pathologists for further annotations.[74] The classifier can then be fine-tuned and learn to make correct predictions on these inputs, enhancing the accuracy for tissue types that are not well represented in the existing training set.

## Conclusion

In summary, we developed a GNN-based analysis pipeline for detecting and differentiating PDAC and CP in histological images. To our knowledge, this is the first DL and GNN-based computational workflow for the differentiation of PDAC and CP in histopathology. Evaluation results indicate that modeling histological images as graphs significantly improves classification accuracy. Meanwhile, conventional analysis models such as MIL can be overly simple for cases where diagnostic features span multiple patches. Furthermore, the incorporation of collagen fiber-based features increased the accuracy of the classifier, highlighting once again the potential diagnostic value of collagen characteristics and the potential clinical benefits of utilizing advanced imaging modalities.

## Author contributions

Study concept and design (B.L., M.S.N., A.G.L. and K.W.E.). Analysis pipeline development, imaging experiment, and data quantification (B.L. and M.S.N.). Sample collection and pathological review (O.S. and A.G.L.). Manuscript organizing and writing (B.L., M.S.N., A.G.L. and K.W.E.).

## Funding

## Ethics approval and consent to participate

Tissue sections were used under MetroHealth IRB CR00000311.

## Data availability statement

Access to the image datasets used in this study and the code is available at https://github.com/uw-loci/gnn-pccp.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgments

We acknowledge members of the Eliceiri and Loeffler labs and Dr. Yin Li for useful input and discussion.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2022.100158.

## References

1. Hidalgo M. Pancreatic cancer. N Engl J Med 2010;362(17):1605–1617.
2. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol 2019;10(1):10.
3. Bellizzi AM, Frankel WL. Pancreatic pathology: a practical review. Lab Med 2009;40(7):417–426.
4. Mostafa ME, Erbarut-Seven I, Pehlivanoglu B, Adsay V. Pathologic classification of "pancreatic cancers": current concepts and challenges. Chin Clin Oncol 2017;6(6):59.
5. Mihaljevic A, Esposito I, Friess H, Kleeff J. Molecular biology, models, and histopathology of chronic pancreatitis and pancreatic cancer. Eur Surg 2009;41(6):250–267.
6. Klöppel G.. Chronic pancreatitis, pseudotumors and other tumor-like lesions. Mod Pathol 2007;20(1):S113–S131.
7. Reddy R. Mass-forming chronic pancreatitis: diagnostic performance of PET/CT. World J Nucl Med 2022 Sep;21(03):239–243.
8. Dal Molin M, Zhang M, De Wilde RF, et al. Very long-term survival following resection for pancreatic cancer is not explained by commonly mutated genes: results of whole-exome sequencing analysis. Clin Cancer Res 2015;21(8):1944–1950.
9. Strobel O, Neoptolemos J, J¨ager D, Bu¨chler MW.. Optimizing the outcomes of pancreatic cancer surgery. Nat Rev Clin Oncol 2019;16(1):11–26.
10. Yamamoto T, Yagi S, Kinoshita H, et al. Long-term survival after resection of pancreatic cancer: a single-center retrospective analysis. World J Gastroenterol: WJG 2015;21(1):262.
11. Esposito I, Hruban RH, Verbeke C, et al. Guidelines on the histopathology of chronic pancreatitis. Recommendations from the working group for the international consensus guidelines for chronic pancreatitis in collaboration with the International Association of Pancreatology, the American Pancreatic Association, the Japan Pancreas Society, and the European Pancreatic Club. Pancreatology 2020;20(4):586–593.
12. Dimastromatteo J, Brentnall T, Kelly KA. Imaging in pancreatic disease. Nat Rev Gastroenterol Hepatol 2017;14(2):97-109.
13. Sarantis P, Koustas E, Papadimitropoulou A, Papavassiliou AG, Karamouzis MV. Pancreatic ductal adenocarcinoma: treatment hurdles, tumor microenvironment and immunotherapy. World J Gastrointest Oncol 2020;12(2):173.
14. Truong LH, Pauklin S. Pancreatic cancer microenvironment and cellular composition: current understandings and therapeutic approaches. Cancers 2021;13(19):5028.
15. Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. Nat Rev Gastroenterol Hepatol 2020;17(8):487–505.
16. Conklin MW, Eickhoff JC, Riching KM, et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am J Pathol 2011;178(3):1221–1232.
17. Nadiarnykh O, LaComb RB, Brewer MA, Campagnola PJ. Alterations of the extracellular matrix in ovarian cancer studied by Second Harmonic Generation imaging microscopy. BMC Cancer 2010;10(1):1-14.
18. Drifka CR, Loeffler AG, Mathewson K, et al. Highly aligned stromal collagen is a negative prognostic factor following pancreatic ductal adenocarcinoma resection. Oncotarget 2016;7(46):76197.
19. Hanley CJ, Noble F, Ward M, et al. A subset of myofibroblastic cancer-associated fibroblasts regulate collagen fiber elongation, which is prognostic in multiple cancers. Oncotarget 2016;7(5):6159.
20. Best SL, Liu Y, Keikhosravi A, et al. Collagen organization of renal cell carcinoma differs between low and high grade tumors. BMC Cancer 2019;19(1):1–8.
21. Zunder SM, Gelderblom H, Tollenaar RA, Mesker WE. The significance of stromal collagen organization in cancer tissue: An in-depth discussion of literature. Crit Rev Oncol Hematol 2020;151, 102907.
22. Drifka CR, Tod J, Loeffler AG, et al. Periductal stromal collagen topology of pancreatic ductal adenocarcinoma differs from that of normal and chronic pancreatitis. Mod Pathol 2015;28(11):1470–1480.
23. Grizzi F, Fiorino S, Qehajaj D, et al. Computer-aided assessment of the extra-cellular matrix during pancreatic carcinogenesis: a pilot study. J Transl Med 2019;17(1):1–9.
24. Han M, Giese G, Bille JF. Second harmonic generation imaging of collagen fibrils in cornea and sclera. Opt Express 2005;13(15):5791–5797.
25. Strupler M, Pena AM, Hernest M, et al. Second harmonic imaging and scoring of collagen in fibrotic tissues. Opt Express 2007;15(7):4054–4065.
26. Keikhosravi A, Bredfeldt JS, Sagar AK, Eliceiri KW. Second-harmonic generation imaging of cancer. Methods Cell Biol 2014;123:531–546.
27. Chen X, Nadiarynkh O, Plotnikov S, Campagnola PJ. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. Nat Protoc 2012;7(4):654–669.
28. Bredfeldt JS, Liu Y, Pehlke CA, et al. Computational segmentation of collagen fibers from secondharmonic generation images of breast cancer. J Biomed Opt 2014;19(1), 016007.
29. Valkenburg KC, De Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. Nat Rev Clin Oncol 2018;15(6):366–381.
30. Zhang J, Liu J. Tumor stroma as targets for cancer therapy. Pharmacol Ther 2013;137(2):200–215.
31. Pantanowitz L, Valenstein PN, Evans AJ, et al. Review of the current state of whole slide imaging in pathology. J Pathol Inform 2011:2.
32. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. Annu Rev Pathol Mech Dis 2013;8:331–359.
33. Farahani N, Parwani AV, Pantanowitz L, et al. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. Pathol Lab Med Int 2015;7(23–33):4321.
34. Huss R, Coupland SE. Software-assisted decision support in digital histopathology. J Pathol 2020;250(5):685–692.
35. Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: a survey. Med Image Anal 2021;67, 101813.
36. Ahmedt-Aristizabal D, Armin MA, Denman S, Fookes C, Petersson L. A survey on graph-based deep learning for computational histopathology. Comput Med Imaging Graph 2021;95, 102027.
37. Dimitriou N, Arandjelovic O, Caie PD. Deep learning for whole slide image analysis: an overview. Front Med 2019:264.
38. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer 2020;1(8):800–810.
39. Van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. Nat Med 2021;27(5):775–784.
40. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw 2008;20(1):61–80.
41. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint. 2016. arXiv:160902907.
42. Wu F, Souza A, Zhang T, Fifty C, Yu T, Weinberger K. Simplifying graph convolutional networks. International Conference on Machine Learning. PMLR; 2019. p. 6861–6871.
43. Prewitt JM, Wu SC. An application of pattern recognition to epithelial tissues. The Second Annual Symposium on Computer Application in Medical Care, 1978. Proceedings. IEEE; Nov 5, 1978. p. 15–25.
44. Aygünés B, Aksoy S, Cinbís RG, Kösemehmetoglu K, Onder S, Uner A. Graph convolutional networks for region of interest classification in breast histopathology. Medical Imaging 2020: Digital Pathology. International Society for Optics and Photonics; 2020. p. 113200K.
45. Adnan M, Kalra S, Tizhoosh HR. Representation learning of histopathology images using graph neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2020. p. 988–989.
46. Pati P, Jaume G, Foncubierta-Rodrıguez A, et al. Hierarchical graph representations in digital pathology. Med Image Anal 2022;75, 102264.
47. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. International Conference on Machine Learning. PMLR; 2018. p. 2127–2136.
48. Li B, Li Y, Eliceiri KW. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021. p. 14318–14328.
49. Keikhosravi A, Li B, Liu Y, Conklin MW, Loeffler AG, Eliceiri KW. Non-disruptive collagen characterization in clinical histopathology using cross-modality image synthesis. Commun Biol 2020;3(1):1-12.
50. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7(1):1–7.
51. Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. Advanced methods of microscope control using µManager software. J Biol Methods 2014;1(2).
52. Pinkard H, Stuurman N, Ivanov IE, et al. Pycro-Manager: open-source software for customized and reproducible microscope control. Nat Methods 2021;18(3):226–228.
53. Keikhosravi A, Li B, Liu Y, Eliceiri KW. Intensity-based registration of bright-field and second-harmonic generation images of histopathology tissue sections. Biomed Opt Express 2020;11(1):160–173.
54. Pielawski N, Wetzer E, Öfverstedt J, et al. CoMIR: contrastive multimodal image representation for registration. Adv Neural Inform Process Syst 2020;33:18433–18444.
55. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9(7):676–682.
56. Macenko M, Niethammer M, Marron JS, et al. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: from Nano to Macro. IEEE; 2009. p. 1107–1110.
57. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770–778.
58. Liu Y, Keikhosravi A, Pehlke CA, et al. Fibrillar collagen quantification with curvelet transform based computational methods. Front Bioeng Biotechnol 2020;8:198.

59. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1026–1034.

60. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint. 2014. arXiv:14126980.

61. Loshchilov I, Hutter F. Sgdr: stochastic gradient descent with warm restarts. arXiv preprint. 2016. arXiv:160803983.

62. Sculley D. Web-scale k-means clustering. Proceedings of the 19th International Conference on World Wide Web; 2010. p. 1177–1178.

63. Huang Z, Li Y. Interpretable and accurate fine-grained recognition via region grouping. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020. p. 8662–8672.

64. Liao PS, Chen TS, Chung PC, et al. A fast algorithm for multilevel thresholding. J Inf Sci Eng 2001;17(5):713–727.

65. Neubert P, Protzel P. Compact watershed and preemptive slic: on improving trade-offs of superpixel segmentation algorithms. 2014 22nd International Conference on Pattern Recognition. IEEE; 2014. p. 996-1001.

66. Lu MY, Williamson DF, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 2021;5(6):555–570.

67. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(11).

68. Laferté JM, Pérez P, Heitz F. Discrete Markov image modeling and inference on the quadtree. IEEE Trans Image Process 2000;9(3):390–404.

69. Sharma H, Alekseychuk A, Leskovsky P, et al. Determining similarity in histological images using graph-theoretic description and matching methods for content-based image retrieval in medical diagnostics. Diagn Pathol 2012;7(1):1-20.

70. Zheng Y, Jiang Z, Shi J, et al. Encoding histopathology whole slide images with location-aware graphs for diagnostically relevant regions retrieval. Med Image Anal 2022;76, 102308.

71. Hein M, Andriushchenko M, Bitterwolf J. Why relu networks yield high-confidence predictions far away from the training data and how to mitigate the problem. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019. p. 41–50.

72. Leibig C, Allken V, Ayhan MS, Berens P, Wahl S. Leveraging uncertainty information from deep neural networks for disease detection. Sci Rep 2017;7(1):1-14.

73. Ren J, Liu PJ, Fertig E, et al. Likelihood ratios for out-of-distribution detection. Adv Neural Inform Process Syst 2019;32.

74. Settles B. Active learning. Synth Lect Artif Intel Mach Learn 2012 Jun 30;6(1):1-14.

75. Hiasa Y, Otake Y, Takao M, et al. Cross-modality image synthesis from unpaired data using CycleGAN. International Workshop on Simulation and Synthesis in Medical Imaging. Cham: Springer; Sep 16, 2018. p. 31–41.

76. Wang H, Rivenson Y, Jin Y, et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nat Methods 2019 Jan;16(1):103–110.