

GrEp: Graph-based epithelial cell classification refinement in histopathology H&E images

Journal Article

Author(s):

Frei, Ana Leni; Garcia-Baroja, Javier; Rau, Tilman; Neppl, Christina; Lugli, Alessandro; Solass, Wiebke; Wartenberg, Martin; Fischer, Andreas; Zlobec, Inti

Publication date:

2026-03

Permanent link:

<https://doi.org/https://doi.org/10.3929/ethz-c-000782735>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Pattern Recognition 171(Part A), <https://doi.org/10.1016/j.patcog.2025.112197>

Funding acknowledgement:

- Trans-omic approach to colorectal cancer: an integrative computational and clinical perspective ()



GrEp: Graph-based epithelial cell classification refinement in histopathology H&E images

Ana Leni Frei^{a,b}, Javier Garcia-Baroja^{a,c}, Tilman Rau^d, Christina Neppel^d, Alessandro Lugli^a,
Wiebke Solass^{id a}, Martin Wartenberg^a, Andreas Fischer^{id e,f}, Inti Zlobec^{id a,*}

^a Institute of Tissue Medicine and Pathology, University of Bern, Bern, Switzerland

^b Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland

^c Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich, Switzerland

^d Institute of Pathology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

^e University of Applied Sciences and Arts Western Switzerland (HES-SO), Fribourg, Switzerland

^f Document, Image and Video Analysis Research Group, University of Fribourg, Fribourg, Switzerland

ARTICLE INFO

Keywords:

Digital pathology
Graph neural networks
Cell classification
Colorectal cancer
Pancreatic ductal adenocarcinoma
Endometrial cancer

ABSTRACT

The automatic cell segmentation and classification from whole slide images plays an important role in digital pathology, unlocking new opportunities for biomarker discovery. Despite extensive research, this task faces persistent challenges such as the differentiation of epithelial cells into normal and malignant. Many existing models lack reporting of epithelial subtyping, and when available, their performance is often suboptimal. This work benchmarks state-of-the-art methods to highlight this limitation and introduces GrEp, a geometric deep learning strategy that considers the broader epithelium tissue architecture to infer cell-level classification rather than relying exclusively on nuclei morphology. The proposed graph-based workflow significantly outperformed state-of-the-art nuclei classification models in colorectal cancer and generalized effectively to two unseen tissue types, endometrium and pancreas, proving the robustness of the geometry-based model. Given its speed and accuracy, we believe GrEp to be a valuable method to refine epithelial cell classification for downstream analyses in clinical and research settings.

1. Introduction

Transformation towards Digital Pathology (DP) opens the field to new opportunities for biomarker discovery and automated clinical applications through the development of Deep Learning (DL) methods. These techniques also offer unprecedented possibilities in analyzing large cohorts of scanned histopathology slides, named whole slide images (WSIs). As personalized medicine advances, the need to characterize tumor composition and understand spatial histological features becomes crucial [1].

In this context, automatic cell segmentation and classification from colorectal cancer (CRC) haematoxylin and eosin (H&E) images has been extensively studied in recent years, leveraging the H&E staining properties that highlights nuclei in purple blue color and other cytoplasmic and extracellular components in pink [2]. Yet, this task is not considered fully resolved [3]. For example, most models either contain a single epithelial class without differentiating normal from malignant epithelial cells, or underperform in this differentiation, often misclassifying normal as malignant. However this distinction is clinically

important as pathology scoring guidelines are based on the reporting of features associated to malignant epithelium such as the extent of tumor invasion, tumor grade, and tumor budding [4]. The incorrect identification of normal epithelial cells as malignant would thus induce an important bias in automated settings. The challenge of accurately differentiating normal and malignant cells arises due to the important intraclass heterogeneity in malignant nuclei regarding size, shape, and morphology as can be seen in Fig. 2. Moreover, the colonic epithelium includes various epithelial cells (goblet cells, enterocytes, paneth cells, tuft cells, enteroendocrine cells, and microfold cells), further complicating the classification [5]. Despite this complexity, on a larger scale, the epithelium tissue covering the colonic tract is very well organized. It is composed of a single-layer columnar epithelium that forms crypts. This simple tissue organization is gradually lost as the stem cells, located at the crypt base, accumulate mutations and begin to over-proliferate and malignant epithelial masses begin to form [6]. The tissue-level changes can be readily identified and are among the key indicators assessed by pathologists to diagnose malignancies. Based on this biological phenomenon, we propose in this work an epithelial nuclei classification

* Corresponding author.

E-mail address: inti.zlobec@unibe.ch (I. Zlobec).

<https://doi.org/10.1016/j.patcog.2025.112197>

Received 23 December 2024; Received in revised form 22 July 2025; Accepted 22 July 2025

Available online 24 July 2025

0031-3203/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

refinement method, using epithelial cell-graphs to capture both local cell features and global tissue structure, emulating pathologists' decision-making. Graph representations of healthy and malignant regions are expected to reflect highly diverse structures that should be easy to classify. Interestingly, throughout cancerization, most tissues feature similar epithelial reorganization processes from non-neoplastic to adenocarcinoma. Under this hypothesis, the developed approach was expected to be easily expanded to other adenocarcinoma types [7]. To evaluate this, the model was applied and finetuned to endometrium and pancreatic tissues.

The main contributions of this work are:

1. We propose a Graph-based Epithelial (GrEp) workflow for epithelial cell sub-classification refinement between normal and malignant for conventional cell classification methods.
2. The proposed GrEp model was compared with other state-of-the-art models and significantly improved the epithelial sub-classification.
3. GrEp demonstrated to be over 4X faster than existing methods for refining epithelial cell classification as post-processing step.
4. The model was extended to two other tissue types (endometrium and pancreas), highlighting the generalizability of the proposed approach to new tissues.
5. Datasets, model weights and code are available on GitHub.¹

2. Related work

2.1. Graph-based approaches in biomedical pattern recognition

Graph-based approaches and graph neural networks (GNN) have gained large popularity in biomedical applications to represent and learn from structured data such as proteins [8], drug design [9], gene interactions [10] but also in histopathology medical imaging to model tissues as networks of cells [11]. The capacity of such methods to capture spatial relationship and cellular interactions between histology elements emulate how pathologists work and allow to extract features that are crucial for diagnosis [12]. In the field of computational pathology, multiple groups have developed graph neural networks for gland classification [13], prediction of cancer subtype [14], cancer grade [15] or mutations [16]. As most of these methods take advantage of cell-based graphs, the initial cell detection and classification of cells must be performed beforehand.

2.2. Nuclei detection and classification

The literature encompasses a diverse array of DL approaches dedicated to nuclei detection and classification.

HoVer-Net [17] was the first method of this kind. It consists of a shared encoder with three decoder branches to simultaneously tackle semantic nuclei instance segmentation and nuclei separation using vertical-and-horizontal distance maps. HoVer-Net remains a state-of-the-art (SOTA) in the field, with some variations proposed over time by introducing modern Vision Transformer encoders [18] or reducing its high computational cost [19].

Chen et al. [20] predict the instance boundary mask for object separation. Ilyas and colleagues [21] build on this idea, adding a bidirectional feature fusion strategy to the backbone to encourage better cross-scale deep representation, which is especially important in DP.

StarDIST [22], another well-established cell segmentation method, predicts centroid probability and distance-to-boundary maps. The model was later extended for classification [23]. StarDIST inspired Chen et al. [24] to leverage contextual information from sampled points within centroid pixels and employ a shape-aware perceptual loss.

Recent AI progress has been fueled by foundation models like the Segment Anything Model (SAM) [25], which have achieved remarkable

performance at zero-shot learning [26]. Multiple groups have adopted it for nuclei segmentation and classification, with or without strategies to employ the promptable properties of SAM [18,27].

Abousamra et al. [28] utilized point annotations to create a dilated nucleus centroid mask. They leverage the cellular spatial configuration for the classification task by predicting Ripley's K function, as a measure of cell clustering.

Other methods use detection transformers, which avoid post-processing to predict the nuclei bounding-boxes [29]. Zhang and colleagues estimate nuclei radii for circular object prediction [30]. Huang et al. infer centroid position instead [31]. They introduce a complex adaptive transformer that employs learnable affine transformations for data augmentation and uses two networks for self-distillation. Huang and colleagues developed another model with a single network trained in two stages [32]. The first fine-tunes the encoder to H&E data. Then, the encoder is frozen, while learnable *grouping prompts* are incorporated into the input to guide the classification by clustering similar cell embeddings.

2.3. Nuclei classification using GNNs

Although most deep learning methods for cell classification rely on convolutional neural networks, some authors have already embarked on integrating topological insights with Graph Neural Networks (GNNs) for nuclei classification with the idea of leveraging spatial relationship to increase classification performance.

Hassan and colleagues used message passing network to identify cell communities [33]. All detected cells in given patches acted as nodes, connected to their nearest neighbors using Euclidean distance. Through message passing layers, edge weights were updated. Inter-community edges were pruned, resulting in a set of community graphs. Each community was finally classified using deep neural network. Their experiments showed improvement over HoVer-Net and MaskR-CNN.

Lou et al. presented SENUCLS, which combines a semantic segmentation block trained alongside a GNN that serves as a feature extractor for vertex and edge embeddings [34]. Node embeddings are enriched by concatenating deep (from a recurrent neural network) and manual shape descriptors of the nuclei mask. Their results are promising without the label forcing of cell communities from the work of Hassan et al. [33], but the feature extraction strategy significantly increases the computational complexity of the model. Recently, the authors have expanded their SENUCLS approach by adding a transformer module that uses nodes and edges as input tokens, enabling the model for global information exchange across the graph differing from immediate neighbors [35].

In contrast, our work proposes a fully automated node feature extraction and classification method for epithelial cell differentiation. Unlike other approaches designed to classify multiple cell types, the proposed GrEp workflow concentrates on the task of epithelial cell subclassification. This strategy allows the model to effectively learn epithelial-specific attributes and to distinguish subtle features useful to accurately differentiate normal from malignant cells. This focused approach reduces the complexity of multiclass classification while specializing for a specific task, resulting in increased performance. The model can be readily applied as post-processing of any nuclei detection/segmentation and classification method and without relying on previous features to infer epithelial differentiation.

3. Methods

3.1. Graph neural networks

Graph neural networks use convolutions on the nodes of a graph to learn new node representations. As graphs can be described by their adjacency matrix A and their node feature matrix H , graph convolutions

¹ <https://github.com/digitalpathologybern/GrEp/>

rely on the multiplication of these two matrices which result in the aggregation of the neighboring node features. In order to weight individual features differently, a weight matrix W is included in the matrix multiplication process. Based on this principle, multiple message passing (MP) aggregation strategies have been proposed. The most common MP layers include GCN [36], GraphSage [37], GIN [38], and GAT [39] and are described below. Although these MP layers rely on the aggregation of features from neighboring nodes to learn new node representations, they all differ slightly in their aggregation strategy.

While GCN aggregates neighbor's nodes' features via mean function and normalizes it according to node degree (using the degree matrix D) as:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (1)$$

GraphSage concatenates a node's own features with the previously averaged features from its neighbors:

$$h_i^{(l+1)} = \sigma \left(W^{(l)} \cdot \text{CONCAT}(h_i^{(l)}, \text{AGGR}(h_u^{(l)}, \forall u \in N(i))) \right) \quad (2)$$

Using this concatenation strategy, GraphSage preserves the features from the node of interest while aggregating the information from neighbors. Following the same idea, GIN proposed a new learnable parameter (ϵ) to adjust the weight of the self node to preserve its identity:

$$h_i^{(l+1)} = \text{MLP} \left((1 + \epsilon^{(l)}) \cdot h_i^{(l)} + \sum_{u \in N(i)} h_u^{(l)} \right) \quad (3)$$

After node feature aggregation a multi-layer perceptron (MLP) is applied to produce distinctive neighborhood features for different communities.

Finally, GAT introduced an attention mechanism to learn relative weights for each node-node connection:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} W h_j^{(l)} \right) \quad (4)$$

Different from GAT, the other three message passing layers weight all neighbors with equal importance. Each MP layer has thus its own specificity and finding GNN architecture can be challenging and depends on the task [40].

After the MP layers, the newly computed node embeddings are usually passed through linear layers for final node classification.

3.2. Proposed graph-based epithelial cell classification method (GrEp)

Multiple models have been developed for cell segmentation/detection and classification in CRC H&E images. These models show great performance for epithelial cell detection and are used here as first step to retrieve epithelial cell centroids. As illustrated in Fig. 1, the proposed Graph-based epithelial cell classification workflow consists of the following three main steps:

(A) Epithelial nuclei feature extraction

Regions of 128×128 px ($64 \times 64 \mu\text{m}$) were cropped around previously detected epithelial centroids. This large crop size has previously been shown to improve nuclei classification by preserving neighborhood information [41]. Then, to direct the model's attention to the central cell, a gamma-focus transformation was applied to enhance the contrast in the tile center (64×64 px), thus highlighting the epithelial nucleus of interest. Conversely, the

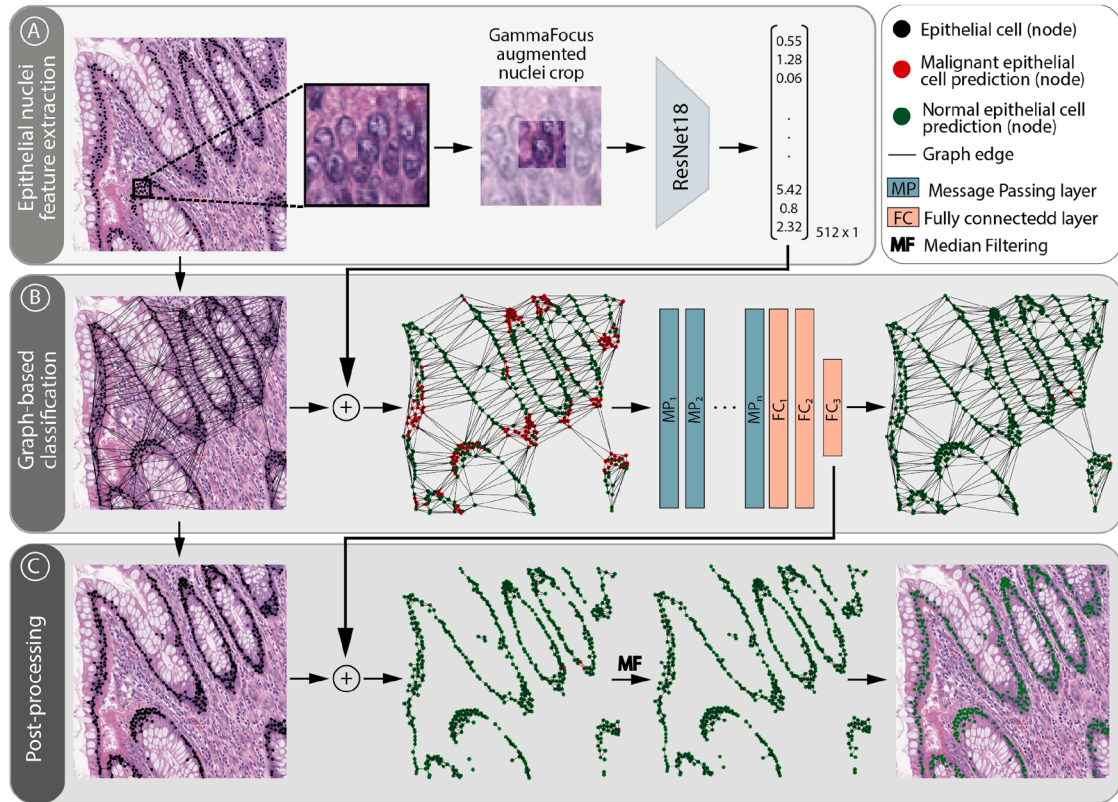


Fig. 1. Overview of the proposed epithelial sub-classification graph-based model pipeline. Starting with a set of detected epithelial cells, the framework consists of: (A) Epithelial nuclei feature extraction using a pretrained ResNet18 from GammaFocus augmented nucleus tiles; (B) Epithelial graph construction using the previously detected epithelial centroids as nodes, the extracted ResNet18 vectors as node embeddings and Delaunay triangulation as edges. The graph is passed through Message Passing (MP) layers for node classification; (C) Epithelial glands are clustered using short Delaunay edges to capture individual glands into single subgraphs on which a Median Filter (MF) is applied to get the final nucleus class. For implementation of the method, a detailed pseudocode is presented in the supplementary material.

contrast of the surrounding region was decreased to minimize the amount of information from neighboring cells [41]. An augmented nucleus tile can be seen in Fig. 1.A. ResNet18 [42] was trained with gamma-augmented nuclei crops for normal versus malignant binary classification. Deep morphology-related nuclear features were extracted using its last hidden layer vector.

(B) Graph-based node classification

For each tile, an undirected epithelial cell-based graph was defined as $G = (V, E)$, where epithelial nucleus in a tile acted as a node, forming the set V . Previously extracted nuclei embeddings from ResNet18 were attached to each node as features. Nearby nodes were then connected using the Delaunay triangulation [43], forming the set of edges E . Delaunay triangulation was selected for its ability to represent cell interactions and reflect how tissue organization changes with malignancy [44]. Edges above a determined length were removed to prevent interactions between distant nuclei that are unlikely to be related. Since the aim of the graph is to capture the overall epithelium structure, the edge thresholds were set above 100px (50 μ m) to ensure that connections span beyond immediate neighbors. As detailed in Section 3.3, different edge thresholds were evaluated during the model optimization process. Graph neural networks were trained and optimized for the specific task of node classification (epithelial cell differentiation) as detailed in Section 3.3. Finally, the optimal GNN was applied to the epithelial cell graph to predict the labels of the nodes (nuclei).

(C) Post-processing

A final post-processing step was applied to smooth the predictions from the graph-based model within individual glands. A very short Delaunay threshold of 40px (20 μ m) was used to cluster epithelial cells into single glands, over which a Median Filter (MF) was applied. The 20 μ m threshold was selected to effectively cluster individual normal glands into single entities, as the average distance between two normal epithelial cells ranged from 9 to 17 μ m in the training dataset. Supplementary Fig. 2 illustrates different clustering values across normal and malignant tissue regions. The edge length must be sufficient to capture all epithelial cells within a gland but short enough to prevent merging adjacent glands. This is particularly relevant when normal and malignant glands are in close proximity to prevent their clustering. This final post-processing step resulted in homogeneous predictions in each epithelial gland or cluster.

3.3. Implementation details

The GrEp workflow was implemented using PyTorch and PyTorch Geometric libraries [45,46]. An AdaptiveAvgPool2d and a Linear layer were added to ImageNet pretrained ResNet18 to adapt for binary classification. ResNet18 was trained on the nuclei crops extracted from the three folds of the Lizard dataset. In order to account for stain variations in histopathology and increase model generalization, nuclei crops were augmented for color variations (ColorJitter) and standard image augmentation (rotations, vertical/horizontal flips) [47]. Finally, the GF augmentation was applied with gamma values of 0.5 and 1.5. ResNet18 was trained for 150 epochs with the AdamW optimizer and a learning rate of $1e^{-6}$ on an NVIDIA GTX1080 GPU. For endometrium and pancreas tissues, the ResNet18 trained on CRC was finetuned for 100 epochs.

When training the GNN models, different graph structures and model architectures were explored to find the best composition. The four MP layers described in Section 3.1 were evaluated and optimized using the open-source python library HyperOpt [48]. For each MP type, the optimization process tested 100 combinations of hyperparameters using the 3-fold cross-validation (CV) from the Lizard dataset (described in Section 4.1). The parameters optimized for each MP type included

the number and depth of MP layers, as well as learning rate, learning rate decay, weight decay, and step size, the definition of the search space can be found in Supplementary Table 1. The optimization procedure was repeated for different graph structures to also determine the optimal edge length (Delaunay triangulation threshold set at: 100, 150, 200, 250 and 300px) and node embedding strategy (ResNet18 embedding or its concatenation with zero-mean normalized node position (x,y) and class prediction). MP layers were followed by 3 fully connected (FC) layers. All graph models were trained on an NVIDIA RTX4090 GPU for 150 epochs with the AdamW optimizer [49], binary cross-entropy loss, LeakyRelu activation function, and dropout ($p = 0.5$) layers in between MP and FC layers. To prevent overfitting, trainings and finetunings were stopped when reaching the smallest validation loss.

3.4. Baselines evaluation of the proposed model

In evaluating our cell classification model, we present three sets of baselines. The first, a Random Forest (RF) model, set a performance threshold that relied on quantitative features extracted from nuclei mask annotations or, when unavailable, the masks produced by the Cerberus model [50]. The feature vector consisted of a total of 83 uncorrelated features that could be categorized as shape, texture, color, spatial arrangement, cell-graph-based, and chromatin distribution [51] descriptors. In this case, the task was simplified to a binary classification problem (normal versus malignant).

The second allows benchmarking against current SOTA cell detection and classification models: HoVer-Net [17], MCSpatNet [28], and PGT [32]. SENUCLS [34] was chosen as the baseline for instance classification refinement using a GNN. The models were retrained using the 3-fold CV from the Lizard dataset to ensure fair comparison. To stabilize training, a minor modification was made to MCSpatNet by incorporating a background channel into the centroid segmentation branch. HoVer-Net, MCSpatNet, and PGT detected and classified cells into three categories: normal, malignant, and "other" (non-epithelial cells). For the subtyping task, SENUCLS distinguished between epithelial subclasses. To ensure better comparability with our method, SENUCLS was trained exclusively on the epithelial predictions of MCSpatNet and PGT, disregarding all other cell types.

The final experiments assess the potential gain of using GrEp as post-processing to existing cell detection and classification models pretrained on separate datasets. To this end, the originally published weights of HoVer-Net, HoVer-Next [19], CellViT [18], MCSpatNet, and Cerberus [52] were downloaded and the models were applied to the test set both without and with GrEp.

3.5. Evaluation metrics

In the context of this study, GNN methods operate as a classification post-processing. Since they cannot influence the detection performance, the assessment of detection and sub-classification tasks is conducted independently.

First, for the detection task and in accordance with the literature [53], a predicted cell instance was considered a True Positive (TP) if it was found within the area defined by a radius of 3 μ m (6px at 20 $\times$) around the closest nuclear center of the Ground Truth (GT). An unmatched prediction was a False Positive (FP_u), while a missed GT cell was considered a False Negative (FN_u). The detection F1-score was calculated globally for all cell types ($\bar{F}_d$), where the score is given by Eq. 5.

$$\bar{F}_d = \frac{TP}{TP + 0.5 \times (FP_u + FN_u)} \quad (5)$$

The classification task only used the matching epithelial predictions with the epithelial Ground Truth (GT) annotations. The classification F1-score (F_c , see Eq. 6) was used to answer whether the models are able

to discern the epithelial subtypes. Here, the FN_m^t and FP_m^t denote wrong classification for matching epithelial cell predictions. The weighted F1-score (F_c^t) is also calculated. All the metrics are reported at the dataset level.

$$F_c^t = \frac{TP^t}{TP^t + 0.5 \times (FP_m^t + FN_m^t)} \quad (6)$$

4. Datasets

The datasets used consist of crops from H&E WSIs of different origin and sizes acquired at a magnification of $20\times$ ($\sim 0.5\mu\text{m}/\text{pixel}$). Example tiles are displayed in Fig. 2 and the specificities of dataset partitions can be seen in Table 1. Fig. 2 also presents nuclei crops from normal and malignant epithelial cells to highlight the intra-class heterogeneity and inter-class resemblances, motivating the use of larger contextual features to classify single nuclei.

4.1. Colorectal cancer (CRC)

The Lizard dataset introduced by Graham et al. [54] was used to train CRC models. The lack of subtyping for epithelial cells led to the selection of a subset of 148 tiles, so each tile clearly only contained either malignant or normal epithelial cells. The PanNuke subset from Lizard was kept aside for testing. The original 3-folds from Lizard were used to train the models and ensure dataset independence between the folds. Additionally, 19 new tiles from TCGA were annotated and added as second independent test set. These tiles were selected to display a variety of epithelial gland morphologies and staining intensities and thus represent a realistic data distribution. Out of these 19 tiles, 5 contained a mixture of normal and malignant epithelium tissue. These regions were specifically selected to observe the graph-based model performance in regions where both epithelial types are found in close proximity.

4.2. Endometrial cancer

Eighty-five tiles ($1000 \times 1000\text{px}$) were extracted from an institutional cohort to contain either normal or malignant epithelium. All epithelial nuclei were annotated at their centroid location as normal or malignant. Tiles were divided into train, validation and test sets according to patient stratification, ensuring that data originating from one patient would be found in a single partition.

Table 1

Dataset Description. For each tissue type, the number of tiles and total number of cells are given. N: Normal, M: Malignant, P:Pancreatitis, Mix: Mixture of normal and malignant cells on the same tile.

Tissue type	Dataset	Epithelial subtype	#tiles	#cells
CRC	Train (Lizard)	N	52	55,665
		M	96	106,823
	Test (PanNuke)	N	7	2724
		M	10	2822
	Test(TCGA)	N	7	6782
		M	7	9502
		Mix	5	8812
Endometrium	Train	N	9	5289
		M	6	5062
	Validation	N	10	4124
		M	10	13,864
	Test	N	33	13,139
		M	17	18,174
PDAC	Train	N	7	7926
		M	14	3784
	Validation	N	6	5415
		M	6	3939
	Test	N	7	6725
		M	10	4484
		P	6	3986

4.3. Pancreatic ductal adenocarcinoma (PDAC)

Fifty-six tiles ($1000 \times 1000\text{px}$) from an internal PDAC WSIs cohort were extracted and their cell nuclei annotated following the same strategy as for CRC and endometrium. The PDAC cohort contained an additional tissue type: pancreatitis. Acute and chronic pancreatitis are both inflammatory diseases of the pancreas that can also be associated with PDAC. Indeed, inflammation and cancer are frequently linked, and pancreatitis can be found in close proximity with PDAC on WSIs. Moreover, despite being different diseases, pancreatitis and PDAC can be difficult to differentiate from a morphology point of view. Examples of normal, malignant, and pancreatitis tiles are shown in Fig. 2. The pancreatitis subtype was not used for finetuning processes. Instead, it was kept in

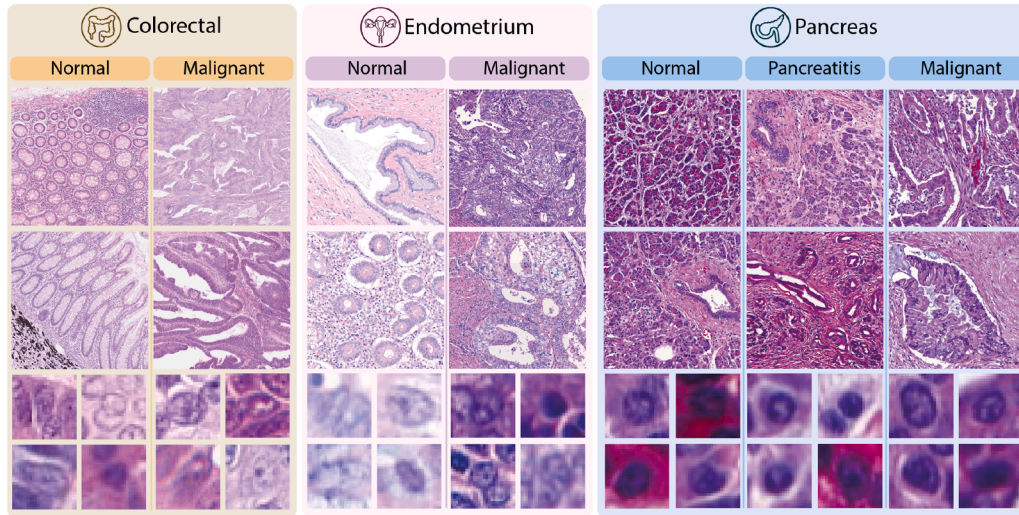


Fig. 2. Dataset visualization: colorectal images in yellow, endometrium in pink and pancreas in blue. The top tiles ($1000 \times 1000\text{px}$) show contextual regions where glands architecture can be seen. For each tissue type, two normal tiles and two malignant tiles are displayed. Normal and malignant epithelial nuclei crops are shown for each tissue type in the two bottom rows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the test set to evaluate how the models behave when presented with an unseen sub-class. In this case, given that pancreatitis originates from non-cancer diseases, these tiles are expected to be assigned to the normal epithelium category. As for endometrium, tiles were divided into training, validation and test sets based on patient-level partitioning.

5. Results

5.1. Model selection

For each MP layer, the model architecture, hyperparameters, and graph structure were selected as the HyperOpt combination that minimized the loss, Supplementary Tables 2–5. Consistent over all four MP types, larger models performed better. The optimization results revealed that the best node embedding consisted of the ResNet18 embedding alone. The addition of node position or ResNet18 prediction was not helpful. In addition, longer edges performed better as the best performances were achieved for edges 250px long. This can be explained by the fact that the aim of the graph-based model is to capture the overall tissue structure at the large scale, which can only be accomplished with longer edges.

Each model was then retrained, using the same folds, with the optimal set of parameters and hyperparameters, Supplementary Table 6. All 4 MP types showed very high classification performance, but the GCN architecture, having the smallest loss and the highest F1-score was selected as the best performing model. The GCN together with the post-processing are referred to as GrEp in the rest of the paper.

Since the model optimization process converged to very large models, we were interested in evaluating the gain of using such large models compared to lighter architecture. The same optimization process was repeated while restricting the number of MP layers to be between 2 and 4. The optimal graph structure was supposed to stay the same, for that reason the edge length and node embedding used were the ones found to be optimal for the larger models. The resulting smaller architecture was found to be 4 layers deep with 512 neurons. The lighter GCN architecture with the post-processing is referred as GrEp-s (GrEp-small). Supplementary Table 7 shows the number of parameters for all models trained in this study.

5.2. GrEp model performance and comparison with existing models

Out of the models trained on the Lizard dataset for cell detection and classification, MCSpatNet showed a superior performance in detection compared to HoVer-Net and PGT and was adopted as the method of choice for epithelial cells detection in the next experiments. Subsequently, graph-based methods were applied to the epithelial cells identified by MCSpatNet. According to Table 2 graph-based models (SENUCLS and GrEp) showed significant better performance than non-graph-based methods on PanNuke ($p < 0.001$, 95%CI [-0.05, -0.03], Cohen-d: 1.88), and TCGA (MCSpatNet vs MCSpatNet + SENUCLS: $p < 0.01$, 95%CI [-0.15, -0.04], Cohen-d: 0.5 and MCSpatNet vs. MCSpatNet + GrEp: $p < 0.001$, 95%CI [0.71, 0.82], Cohen-d: 4.7]). The RF trained

to classify hand-crafted features extracted from nuclei masks performed worse than all DL-based methods, probably due to the higher adaptability to data variability from DL models. This highlights the benefit of using DL techniques for nuclei classification tasks and motivated the use of ResNet18 to extract cell features. Interestingly, our proposed GrEp model achieved the highest average performance on both normal and malignant classes individually and significantly outperformed SENUCLS on TCGA test set ($p < 0.001$, CI95% [0.8, 0.91], Cohen-d: 4.3).

The results can be visualized in Fig. 3. The lower detection score of HoVer-Net and PGT can be seen especially in high-density epithelium regions. Dense normal epithelial crypts were particularly challenging for MCSpatNet, which misclassified most of these cells as malignant. This effect was reverted by using graph-based methods. Notably, GrEp effectively handled regions with adjacent normal and malignant tissues despite its long edges, with most of the remaining misclassified cells after GrEp post-processing located along tile edges. This is probably an effect from the missing part of the neighboring epithelium tissue, degrading the performance.

The benefit of using GrEp as post-processing for existing nuclei classification models was then evaluated. GrEp was applied on top of the published weights from HoVer-Net, HoVer-NetXt, CellViT, MCSpatNet, and Cerberus models, Fig. 4 and Table 3. GrEp significantly improved the epithelial sub-classification for all the models and can add valuable information to models that do not assess normal versus malignant epithelium (MCSpatNet and Cerberus here).

Finally, to verify that GrEp would not be dramatically affected by inaccurate epithelial nuclei detections, we computed the performance of GrEp under perturbation. To do so, some nuclei were randomly removed and some non-epithelial nuclei were randomly added from TCGA ground truth annotation. GrEp performance was computed on true epithelial cells only. Results show that GrEp is not affected by inaccurate epithelial detection and the performance decreases only for large perturbations (more than 50% random cell selection), Supplementary Table 8 and Supplementary Fig. 4.

5.3. Ablation study

The proposed method was evaluated using different variations to show the impact of each of its components on the performance. MCSpatNet was used as epithelial cell detection model and ResNet18 classification of the detected epithelial cells acted as the baseline for comparison. Iteratively, each module of the proposed method was individually incorporated to the pipeline and evaluated, Table 4 and Supplementary Fig. 5. One-way ANOVA revealed a significant difference in accuracy among the full GrEp (RN18 + GCN + MF) and ablated variants ($F = 28.8$, $p = 0.0001$). Post-hoc Tukey tests were performed to pairwise compare variants and indicated that all elements included in the full model significantly increased the performance ($p < 0.05$). It is important to note that the addition of median filtering to smooth predictions inside individual glands on the baseline improves the score without outperforming the graph-based model, showing the benefit of using MP layers before gland clustering. Importantly, MF works because graph-based node

Table 2

Quantitative assessment of existing nuclei classification methods (Cerberus, HoVer-Net, PGT, MCSpatNet). Hand-crafted nuclei features were extracted from Cerberus nuclei masks followed by Random Forest (RF) classification. “-” in the $\bar{F}_d$ columns indicate that the graph-based methods do not change the initial epithelial cell detection and the score remains the same as without graph-based post-processing.

Method	PanNuke				TCGA			
	$\bar{F}_d$	$\bar{F}_c$	F_c^m	F_c^n	$\bar{F}_d$	$\bar{F}_c$	F_c^m	F_c^n
Hand-crafted + RF	-	85.63	84.24	86.99	72.3	71.61	64.29	76.66
HoVer-Net	62.26±2.02	97.31±0.09	97.81±0.10	96.30±0.36	60.50±2.29	81.41±9.46	82.84±11.50	79.13±6.59
PGT	75.06±2.05	99.76±0.22	99.77±0.20	99.74±0.24	65.52 ±3.16	87.94±0.47	89.77±0.35	85.16±1.99
MCSpatNet	79.32±0.38	94.32±0.96	95.41±0.6	93.14±1.40	72.02±0.61	80.84±1.75	86.69±1.28	72.14±3.24
MCSpatNet + SENUCLS	-	100±0.0	100±0.0	100±0.0	-	90.09±1.43	90.59±1.0	89.50±2.04
MCSpatNet + GrEp	-	100±0.0	100±0.0	100±0.0	-	97.2±0.88	97.84±0.65	96.1±1.28

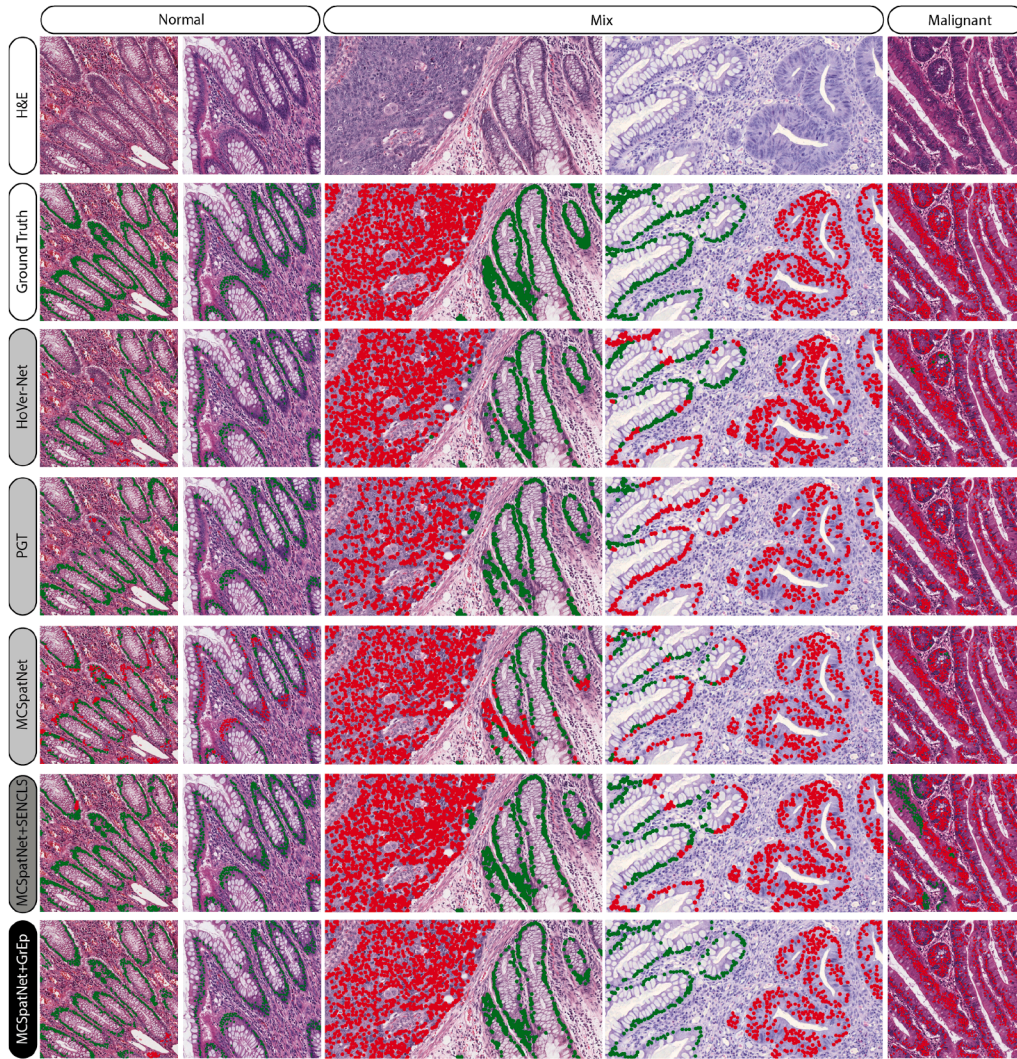


Fig. 3. Visual results on TCGA test tiles using existing cell detection and classification models (HoVer-Net, PGT and MCSpatNet) retrained for normal versus malignant epithelial cell classification in comparison with the ground truth. The two bottom lines present results when applying graph-based epithelial post-processing (SENUCLS and the proposed GrEp). Green dots highlight normal cells and red dots malignant cells. Zoomed insets with predictions are presented in Supplementary Figure 3 with corresponding F1 score. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Quantitative comparison of the epithelial classification performance between existing nuclei classification models without and with the proposed GrEp model. “-” in the $\bar{F}_d$ columns indicate that the graph-based methods do not change the initial epithelial cell detection and the score remains the same as without graph-based post-processing.

Method	$\bar{F}_d$	$\bar{F}_c^e$	F_c^m	F_c^n
HoVer-Net	44.51	56.65	80.29	9.17
HoVer-Net + GrEp	-	96.82	97.7	95.07
HoVer-NEXT	43.62	58.98	78.64	26.51
HoVer-NEXT + GrEp	-	96.89	95.79	97.56
CellViT	56.37	67.27	82.02	41.91
CellViT + GrEp	-	96.48	95.09	97.28
MCSpatNet + GrEp	51.87	97.31	97.95	96.12
Cerberus + GrEp	72.33	98.05	97.46	98.43

classification performance is already high, otherwise this approach might drastically reduce the overall performance. The full version of the proposed approach significantly outperforms all baselines and intermediate steps ($p < 0.001$), demonstrating the necessity of including each element in the final model.

Table 4

Ablation Study on TCGA dataset using GrEp variants. The full GrEp pipeline corresponds to FN18 + GCN + MF. RN18: ResNet18 node classification and embedding extraction, MF: Median Filter.

GrEp elements			TCGA		
RN18	GCN	MF	$\bar{F}_c^e$	F_c^m	F_c^n
✓	-	-	87.58±0.97	90.83±0.84	82.01±1.22
✓	-	✓	90.13±1.14	92.51±1.03	86.04±1.44
✓	✓	-	94.45 ±0.77	95.75 ±0.48	92.21±1.26
✓	✓	✓	97.2 ±0.88	97.84 ±0.65	96.1±1.28

The impact of using GrEp to refine cell embeddings can be seen in Fig. 5. A principal component analysis (PCA) was performed on the node embeddings at the input (ResNet18 embeddings) and output layer of GrEp GNN, as shown in Fig. 5A. The input node features show an important overlap of normal and malignant feature vectors. The same behaviour is observed when applying t-distributed stochastic neighbor embedding (t-SNE), see Supplementary Figure 6. Patches corresponding to these overlapping embeddings were retrieved and displayed either clustered normal epithelium, as can be found at the bottom of

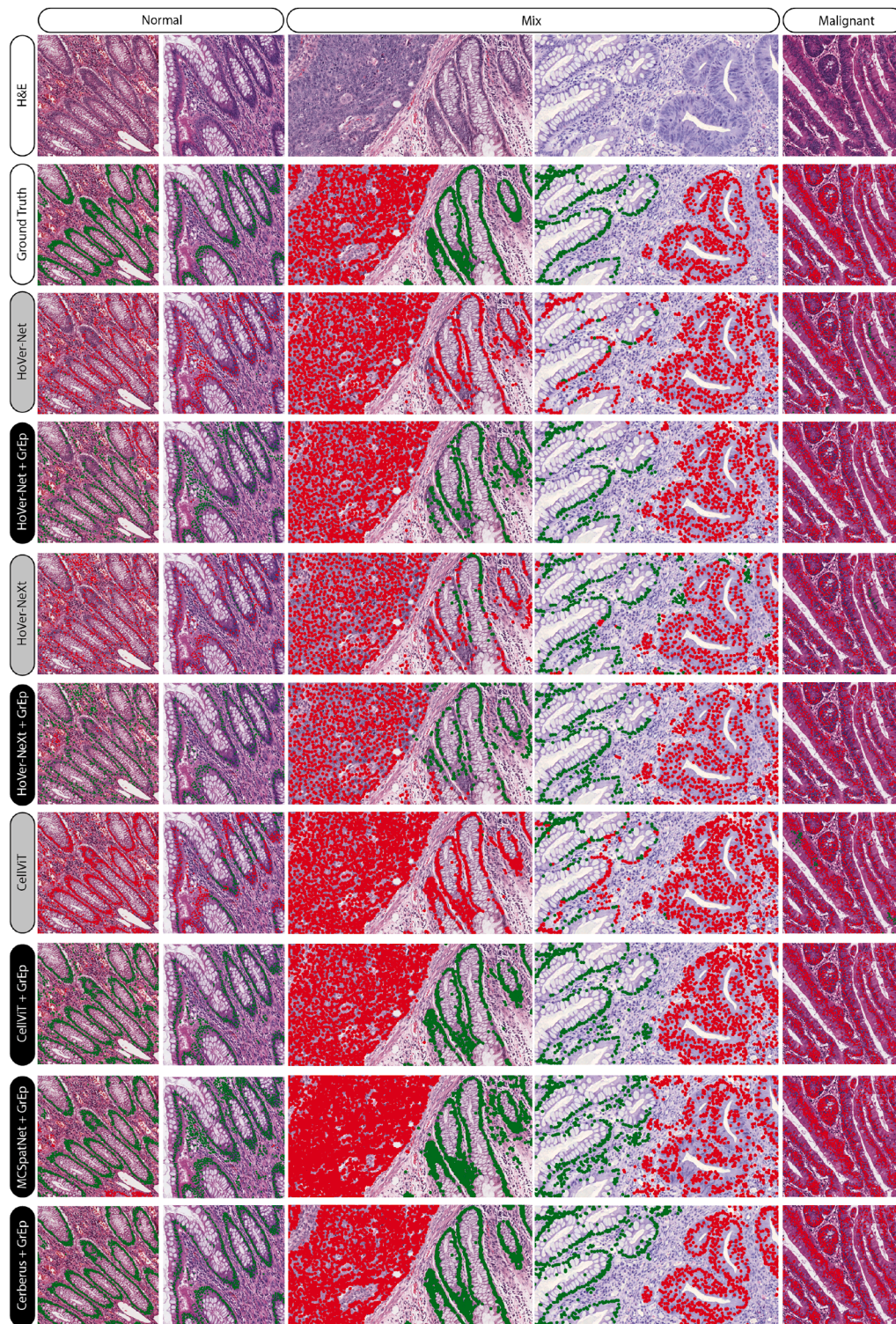


Fig. 4. Visualization of GrEp as post-processing on existing pretrained cell segmentation and classification models (HoVer-Net, HoVer-NeXt, CellViT, MCSpatNet and Cerberus). Red dots indicate malignant cells; green dots indicate normal cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

healthy colonic crypts, or single-layered epithelium from malignant regions. This confusion is expected at the small cell-level context. This uncertainty gets reverted by using GNNs. As a result, the embeddings are more distinctly segregated, highlighting the ability of GrEp to better capture features specific to each epithelial subtype based on contextual information rather than just nuclei features. Fig. 5.B shows GrEp's

capability of correctly classifying normal cells in crowded crypts as well as more organized tumor regions.

To further investigate the node classification performed by GrEp, node importance were computed and visualized using GNNExplainer [55], Supplementary Figure 7. When selecting specific nodes that were initially misclassified by ResNet18 (located at the crowded bottom of

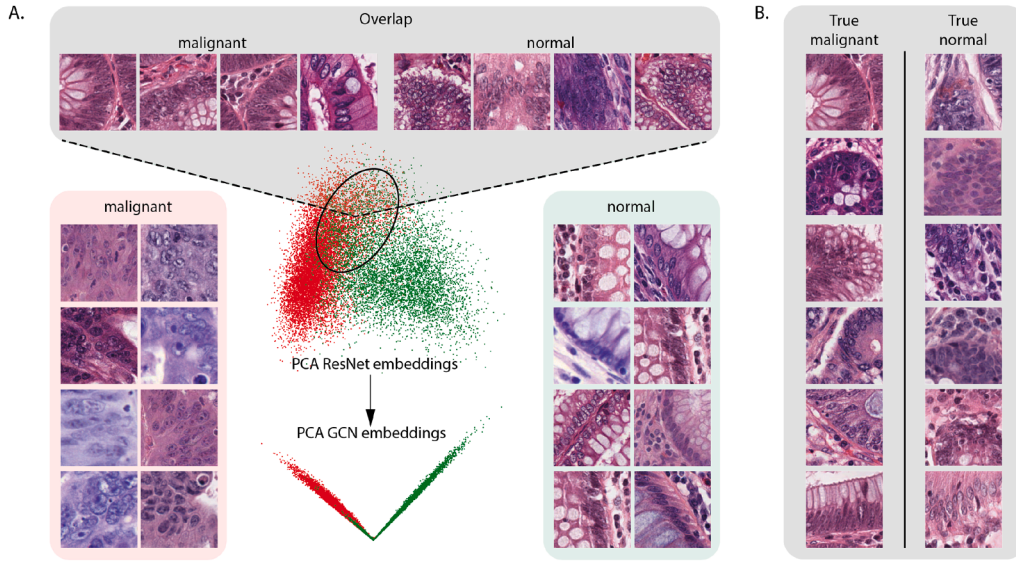


Fig. 5. A. Principal component analysis (PCA) on the input node embeddings (ResNet18 last hidden layer) and GrEp's last hidden layer. Green dots indicate normal cells, and red dots indicate malignant cells. The black circle highlights the region where the normal and malignant embeddings from ResNet18 overlap. For each region in the PCA plot, representative epithelial tiles were retrieved. B. Tiles of malignant and normal epithelial cells that were misclassified by ResNet18 and correctly classified by GrEp. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

healthy crypts) it is interesting to see that GrEp took most advantage of epithelial cells located in less dense crypts regions. Similarly, when looking at malignant nodes, it is interesting to see that the model looks at cells beyond the direct neighbouring region. This emphasize the benefit of using global graph modeling of the tissues to infer the proper epithelial cell classification.

5.4. Inference time

The speed of the GrEp workflow was compared to SENUCLS. Both models were run on an Nvidia A100 GPU (80GB VRAM) and the inference times were assessed on the TCGA test set (total of 25,096 epithelial cells). Our proposed model was more than 4X faster than SENUCLS (57 sec vs. 4min 25 sec), probably due to the more complex feature extraction process of SENUCLS. The feature extraction from ResNet18 was also the most time-consuming step in the proposed GrEp pipeline, accounted for 81 % of the inference time (46sec). The post-processing median filtering, on the other hand, took only 4sec. Interestingly, the lighter GrEp architecture, GrEp-s defined in Section 5.1, was only 0.7sec faster than GrEp. Considering that WSI might contain hundreds of thousands of epithelial cells, the gain in speed might be of a few seconds only when using GrEp-s (28sec for 1M epithelial cells).

5.5. Generalizability to unseen tissue types

The models trained on CRC were first applied to both endometrial and pancreas tissues without retraining the models. Then, finetuning was performed on the extraction of the node embeddings, meaning that only the ResNet18 module from GrEp, or the respective feature encoders of other methods were adjusted. On the other hand, the graph component of the models remained unchanged, meaning that the graph models were only trained with CRC samples.

5.5.1. Endometrium

CRC trained models performed well on endometrium tissue. However, a significant decrease in performance can be observed compared to CRC, especially for normal tissue. The three graph-based methods significantly outperformed the baseline (MCSpatNet vs. MCSpatNet + SENUCLS: $p < 0.001$, 95 %CI [-0.15, -0.07], Cohen-d: 0.3; MCSpatNet vs MCSpatNet + GrEp: $p < 0.001$, 95 %CI [0.14, 0.38], Cohen-d: 0.7), with GrEp significantly outperforming SENUCLS ($p < 0.001$, 95 %CI [0.25, 0.48], Cohen-d: 0.9). Interestingly, GrEp-s performed significantly better than GrEp but with a small effect size ($p < 0.001$, 95 %CI [0.07, 0.16], Cohen-d: 0.29). After finetuning, all models reached almost perfect performance, Table 5 and Supplementary Figure 8. After finetuning,

Table 5

Quantitative results of the generalization performance of CRC-trained models to both endometrium (Endo) and pancreatic tissues. -s stands for the lighter version of GrEp. "-" in the $\bar{F}_d$ columns indicate that the graph-based methods do not change the initial epithelial cell detection and the score remains the same as without graph-based post-processing.

	Method	CRC trained models				Finetuned models			
		$\overline{F}_d$	$\overline{F}_c$	F_c^m	F_c^n	$\overline{F}_d$	$\overline{F}_c$	F_c^m	F_c^n
Endo	MCSpatNet	79.75±0.18	69.28±6.17	84.07±1.90	38.50±15.78	78.86	96.96	97.48	96.12
	MCSpatNet + SENUCLS	-	76.39±5.81	86.29±2.33	58.29±12.27	-	98.43±0.33	98.72±0.27	97.96±0.42
	MCSpatNet + GrEp	-	79.33±3.32	88.09±1.48	61.46±7.07	-	99.77±0.07	99.83±0.05	99.65±0.11
	MCSpatNet + GrEp-s	-	86.22±0.48	91.29±0.34	75.87±1.29	-	99.86±0.01	99.9±0.01	99.79±0.02
Pancreas	MCSpatNet	79.77±0.18	40.33±1.14	67.06±0.52	14.08±1.68	82.28	71.59	50.82	83.26
	MCSpatNet + SENUCLS	-	36.96±0.70	68.77±0.15	1.76±1.38	-	82.77±0.41	75.38±0.86	87.01±0.90
	MCSpatNet + GrEp	-	41.21±4.55	68.78±0.9	7.79±8.99	-	97.26±0.32	96.58±0.38	97.73±0.28
	MCSpatNet + GrEp-s	-	46.07±1.19	67.63±1.0	19.93±1.46	-	96.98±0.25	96.18±0.33	97.53±0.2

GrEp and GrEp-s were not significantly different but both SENUCLS and GrEp outperformed the baseline (MCSpatNet vs. MCSpatNet + SENUCLS: $p < 0.001$, 95 %CI [-0.62, -0.39], Cohen-d: 1.7; MCSpatNet vs MCSpatNet + GrEp: $p < 0.001$, 95 %CI [0.34, 0.59], Cohen-d: 1.5). This highlights the importance of the node features and more importantly the ability of graph-based models to generalize to unseen tissue types since the graph models were not adapted to endometrium.

5.5.2. Pancreas

The inter-tissue differences between PDAC and CRC results in a poor distinction of epithelial cells from other cell types, despite the relatively high $\bar{F}_d$. Said domain shift also affects the quality of the deep cellular representations the GNNs use as nodal features. This, along with the inability to construct meaningful epithelial graphs, renders only little improvement in the subtyping task from the graph subclassification. After finetuning, the confusion of the epithelial cells with the other cells is corrected, but the $\bar{F}_c^e$ is below 80 % for the computer vision model (MCSpatNet), highlighting their difficulty to generalize as the tissue appearance shifts from the original training set. The application of epithelial graphs significantly improved performance (MCSpatNet vs. MCSpatNet + SENUCLS: $p < 0.05$, 95 %CI [-0.35, -0.01], Cohen-d: 0.8; MCSpatNet vs MCSpatNet + GrEp: $p < 0.001$, 95 %CI [0.52, 0.85], Cohen-d: 4.5), with GrEp surpassing SENUCLS ($p < 0.001$, 96 %CI [0.82, 0.91], Chen-d: 7.2) and no significant difference was observed between GrEp and GrEp-s. Visualization of the predictions from the CRC and finetuned models can be seen in Supplementary Figure 9. As for pancreatitis, the results show that they are correctly identified as non-tumor cells in most of the cases (2229 cells correctly classified as normal out of 2967 pancreatitis cells), with the most difficult regions containing glands resembling malignant epithelium, Supplementary Figure 10. The weighted classification F1-score of 85.61 % indicates there is still a good recognition of pancreatitis cells as non-cancerous.

6. Discussion

In this work, we aimed at improving the challenging task of epithelial cell sub-classification into normal and malignant from histology H&E images. The proposed workflow, named GrEp, preserves and exploits the structural changes occurring in tissues when transitioning from normal to malignant to infer cell classification through the use of MP functions on epithelial cell graphs. As shown by the node importance visualizations, nodes located far apart from a node of interest are relevant and play a crucial role in the graph-based classification, highlighting the benefit and importance of including larger tissue regions to infer epithelial sub-classes. These factors help increase model interpretability and explainability, ultimately increasing the trust in the models' predictions. GrEp achieved state-of-the-art performance in distinguishing normal from malignant cells in CRC, reaching an almost perfect classification score.

Since GrEp relies on the extraction of ResNet18 hidden vectors as node features, it can be applied on top of any cell classification model as post-processing without interfering with the initial classification task. However, GrEp could easily be retrained for other node features such as hidden vectors from existing classification models (i.e. Cerberus, HoVer-NeXt, CellViT, MCSpatNet), removing the node feature extraction step from ResNet18 and resulting in a lighter and faster epithelial refinement classification.

GrEp also showed high generalizability performance in both endometrial cancer and PDAC. In both cases, the graph-based models were not changed, only the feature extraction models were finetuned. This highlights the similarity in tissue architecture between different adenocarcinoma types and proves that using topology to infer epithelial-cell type is a powerful tool. However, it is important to mention that while colorectal and endometrium tissues have a clear layered tissue architecture, pancreatic tissue is composed of lobules of epithelial glands, azini and ducts inside fibrous stroma. The absence of layering in pancreatic

tissue makes it more complex to analyze and especially difficult to distinguish chronic pancreatitis from PDAC in some cases. For that reason, the model was also evaluated on pancreatitis and showed a good performance of 85 % weighted F1 score. Eventhough GrEp performance on pancreatitis was lower than on normal versus malignant cells classification, the model was able to classify pancreatitis as normal in the majority of the cases. It is important to note that pancreatitis can be very challenging to distinguish from PDAC even for expert pathologists. To make a decision, they tend to look at tissue slides with very low magnification to have a good overview of the entire processes happening in the surrounding and overall tissue architecture. For that reason, we believe that to overcome this misclassification by the model, increasing the tile size (i.e. the size of the region covered by GrEp) to include more context information might help to improve further the classification of pancreatitis. Based on the high generalizability of GrEp on endometrium and pancreatic tissues, we believe that the model could be further expanded to other adenocarcinoma types.

While GrEp offers many strengths, there are opportunities for further improvement. One limitation is that GrEp only performs binary classification and thus cannot properly handle ambiguous cases such as pancreatitis or dysplastic glands. Knowing that only regular and well-organized glands were included in the normal class, slight changes in the expected normal epithelium architecture might shift the model's predictions towards the malignant class, resulting in dysplastic glands being classified as malignant. However, we believe the GNN part to be easily modifiable to include more classes by retraining and adapting the last linear layer to output more classes.

The results also showed that even under large perturbations (30 % of randomness in the initial epithelial cell detection, Supplementary Table 8) GrEp consistently identified normal from tumor cells. It is however important to note that while GrEp is robust to perturbations, the precise detection of epithelial cells remains crucial as inaccuracies could influence downstream analyses and potentially affect overall conclusions.

Another opportunity for improvement lies in the application of the pipeline to WSIs. As the visualizations of the results revealed that most of the misclassified cells by GrEp were found at tile edges, we believe that extracting overlapping tissue tiles from WSIs will help mitigate tile edge artifacts and potentially improve further the classification performance.

An additional limitation concerns the datasets of the generalization study. While the patient-level split between training, validation and test sets was carefully implemented during data curation, pancreas and endometrium datasets originated from a single institutional cohort. The validation of the results on external datasets would be required to further strengthen the findings. Furthermore, while the pancreatitis subset contained only 6 tiles. However, these images originate from different patients and collectively contained 4000 cells. For these reasons, we believe these 6 images to reflect a significant diversity and a meaningful assessment of the model performance, though further validation on larger cohorts will further solidify our observations.

As shown by the experimental results, the proposed GrEp architecture achieved state-of-the-art performance in the sub-classification of epithelial cells while being faster than other graph-based refinement solutions. We thus believe GrEp to be useful for both clinical and research tasks. For example, the evaluation of the proportion of tumor cells (versus other cells) is routinely performed in pathology labs for both tumor banking and molecular pathology. There, an accurate estimation of the tumor cell fraction is of high importance to ensure reliable mutation calls and select appropriate targeted therapies. Multiple groups have shown the benefit of using DL methods for the estimation of tumor cell fraction [56–58]. Accurate malignant epithelial cell localization on histology images is also crucial to correctly assess the presence of tumor-infiltrating lymphocytes (TILs). TILs have been shown to be prognostic of better survival across malignancies and are also linked to microsatellite instability and mismatch repair deficiency in CRC [59,60]. These two molecular pathways are used to determine treatment strategies. Finally, malignant epithelial cell localization on WSIs might also be of

interest to automatically assess tumor budding (TB) status in CRC where TB can be used to define treatment decisions [61].

7. Conclusion

In this paper, we proposed a graph-based deep learning pipeline, named GrEp, for the classification refinement of epithelial cells into normal and malignant. Leveraging the overall epithelium architecture changes upon transition towards malignancy, GrEp learns structural features to accurately distinguish normal from malignant epithelial cell nuclei. Our proposed GrEp model performed significantly better than state-of-the-art models and showed great generalizability to unseen tissue types (endometrium and pancreas), proving the strength of the tissue architecture-based classification method. Finally, thanks to its simple architecture, GrEp was found to be faster than existing nuclei refinement method, making it a suitable solution for epithelial cell classification post-processing in both research and clinical settings.

Despite the numerous advantages of GrEp further enhancements are possible. First, despite the good generalization performance of GrEp to endometrium and pancreas tissues, additional validations on external datasets should be carried out to strengthen the findings and assess the performance across different institutions. Second, future improvements of the proposed GrEp workflow include its expansion to include intermediate epithelial subtypes such as dysplasia (pre-malignant glands) which might be classified as malignant by the current GrEp which could lead to biased epithelial estimations or false malignant positivity.

CRediT authorship contribution statement

Ana Leni Frei: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Javier Garcia-Baroja:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis; **Tilman Rau:** Writing – review & editing, Validation, Supervision, Data curation; **Christina Neppi:** Writing – review & editing, Validation, Supervision, Data curation; **Alessandro Lugli:** Writing – review & editing, Validation, Supervision, Data curation; **Wiebke Sollas:** Writing – review & editing, Validation, Supervision, Data curation; **Martin Wartenberg:** Writing – review & editing, Validation, Supervision, Data curation; **Andreas Fischer:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization; **Inti Zlobec:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Inti Zlobec reports financial support was provided by Swiss National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the [Swiss National Science Foundation \(CRSII5_193832\)](https://www.id.unibe.ch/hpc). Calculations were performed on UBELIX <https://www.id.unibe.ch/hpc>, the HPC cluster at the University of Bern. Results presented here are partially based on data provided by the TCGA Research Network.

References

- [1] V. Baxi, R. Edwards, M. Montalto, S. Saha, Digital pathology and artificial intelligence in translational medicine and clinical practice, *Mod. Pathol.* 35 (1) (2022) 23–32. <https://doi.org/10.1038/s41379-021-00919-2>
- [2] A.H. Fischer, K.A. Jacobson, J. Rose, R. Zeller, Hematoxylin and eosin staining of tissue and cell sections, *CSH Protoc.* (2008). <https://doi.org/10.1101/pdb.prot4986>
- [3] S. Graham, Q.D. Vu, M. Jahanifar, M. Weigert, U. Schmidt, W. Zhang, J. Zhang, S. Yang, J. Xiang, X. Wang, J.L. Rumberger, E. Baumann, P. Hirsch, L. Liu, C. Hong, A.I. Aviles-Rivero, A. Jain, H. Ahn, Y. Hong, H. Azzuni, M. Xu, M. Yaqub, M.-C. Blache, B. Piégu, B. Vernay, T. Scherr, M. Böhlend, K. Löffler, J. Li, W. Ying, C. Wang, D. Snead, S.E.A. Raza, F. Minhas, N.M. Rajpoot, CoNIC challenge: pushing the frontiers of nuclear detection, segmentation, classification and counting, *Med. Image Anal.* 92 (2024) 103047. <https://doi.org/10.1016/j.media.2023.103047>
- [4] M.B. Loughrey, F. Webster, M.J. Arends, I. Brown, L.J. Burgart, C. Cunningham, J.F. Flejou, S. Kakar, R. Kirsch, M. Kojima, A. Lugli, C. Rosty, K. Sheahan, N.P. West, R.H. Wilson, I.D. Nagtegaal, Dataset for pathology reporting of colorectal cancer: recommendations from the international collaboration on cancer reporting (ICCR), *Ann. Surg.* (2022). <https://doi.org/10.1097/SLA.0000000000005051>
- [5] V. Bonis, C. Rossell, H. Gehart, The intestinal epithelium – Fluid fate and rigid structure from crypt bottom to villus tip, *Front. Cell Dev. Biol.* 9 (2021).
- [6] L. Cernat, C. Blaj, R. Jackstadt, L. Brandl, J. Engel, H. Hermeking, A. Jung, T. Kirchner, D. Horst, Colorectal cancers mimic structural organization of normal colonic crypts, *PLoS ONE* 9 (2014).
- [7] D. Coradini, C. Casarsa, S. Oriana, Epithelial cell polarity and tumorigenesis: new perspectives for cancer detection and treatment, *Acta Pharmacol. Sin.* (2011) 52–564. <https://doi.org/10.1038/aps.2011.20>
- [8] M. Li, Y. Cao, X. Liu, H. Ji, Knowledge-enhanced and structure-enhanced representation learning for protein–ligand binding affinity prediction, *Pattern Recognit.* 166 (2025) 111701. <https://doi.org/10.1016/j.patcog.2025.111701>
- [9] X. Chen, S. Li, R. Liu, B. Shi, J. Liu, J. Wu, K. Xu, Molecular graph contrastive learning with line graph, *Pattern Recognit.* 162 (2025) 111380. <https://doi.org/10.1016/j.patcog.2025.111380>
- [10] D. Hu, Y. Liu, X. Wang, L. Geng, F. Zhang, Z. Xiao, J.C.-W. Lin, Multi-information fusion graph convolutional network for cancer driver gene identification, *Pattern Recognit.* 165 (2025) 111619. <https://doi.org/10.1016/j.patcog.2025.111619>
- [11] D. Ahmedt-Aristizabal, M.A. Armin, S. Denman, C. Fookes, L. Petersson, A survey on graph-based deep learning for computational histopathology, *Computerized Med. Imaging Graph.* 95 (2022) 102027. <https://doi.org/10.1016/j.compmedimag.2021.102027>
- [12] L. Studer, J.M. Bokhorst, I. Nagtegaal, I. Zlobec, H. Dawson, A. Fischer, Tumor budding t-cell graphs : assessing the need for resection in pT1 colorectal cancer patients, in: *Proceedings of Medical Imaging with Deep Learning (MIDL), 2023*.
- [13] L. Studer, S. Toneyan, I. Zlobec, H. Dawson, A. Fischer, Graph-based classification of intestinal glands in colorectal cancer tissue images, in: *MICCAI 2019 Computational Pathology Workshop COMPAY*, 2019.
- [14] P. Pati, G. Jaume, A. Foncubierta-Rodríguez, F. Feroce, A.M. Anniciello, G. Scognamiglio, N. Brancati, M. Fiche, E. Dubruc, D. Riccio, M. Di Bonito, G. De Pietro, G. Botti, J.-P. Thiran, M. Frucci, O. Goksel, M. Gabrani, Hierarchical graph representations in digital pathology, *Med. Image Anal.* 75 (2022) 102264. <https://doi.org/10.1016/j.media.2021.102264>
- [15] Y. Zhou, S. Graham, N. Alemi Koohbanani, M. Shaban, P.-A. Heng, N. Rajpoot, CGC-Net: cell graph convolutional network for grading of colorectal cancer histology images, in: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 388–398. <https://doi.org/10.1109/ICCVW.2019.00050>
- [16] W. Lu, S. Graham, M. Bilal, N. Rajpoot, F. Minhas, Capturing cellular topology in multi-gigapixel pathology images, in: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 1049–1058. <https://doi.org/10.1109/CVPRW50498.2020.00138>
- [17] S. Graham, Q.D. Vu, S.E.A. Raza, A. Azam, Y.W. Tsang, J.T. Kwak, N. Rajpoot, Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images, *Med. Image Anal.* 58 (2019) 101563. <https://doi.org/10.1016/j.media.2019.101563>
- [18] F. Hörst, M. Rempe, L. Heine, C. Seibold, J. Keyl, G. Baldini, S. Ugurel, J. Siveke, B. Grünwald, J. Egger, J. Kleesiek, Cellvit: vision transformers for precise cell segmentation and classification, *Med. Image Anal.* 94 (2024) 103143. <https://doi.org/10.1016/j.media.2024.103143>
- [19] E. Baumann, B. Dislich, J.L. Rumberger, I.D. Nagtegaal, M.R. Martinez, I. Zlobec, HoVer-NeXt: a fast nuclei segmentation and classification pipeline for next generation histopathology, in: *Medical Imaging with Deep Learning*, 2024.
- [20] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, P.-A. Heng, DCAN: Deep contour-aware networks for object instance segmentation from histology images, *Med. Image Anal.* 36 (2017) 135–146. <https://doi.org/10.1016/j.media.2016.11.004>
- [21] T. Ilyas, Z.I. Mannan, A. Khan, S. Azam, H. Kim, F. De Boer, TSFD-Net: tissue specific feature distillation network for nuclei segmentation and classification, *Neural Netw.* 151 (2022) 1–15. <https://doi.org/10.1016/j.neunet.2022.02.020>
- [22] M. Weigert, U. Schmidt, R. Haase, K. Sugawara, G. Myers, Star-convex polyhedra for 3D object detection and segmentation in microscopy, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3666–3673.
- [23] M. Weigert, U. Schmidt, Nuclei Instance Segmentation and Classification in Histopathology Images with Stardist, in: *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)*, IEEE, 2022, pp. 1–4.
- [24] S. Chen, C. Ding, M. Liu, J. Cheng, D. Tao, CPP-net: context-aware polygon proposal network for nucleus segmentation, *IEEE Trans. Image Process.* 32 (2023) 980–994.

- [25] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, P. Dollár, R. Girshick, Segment Anything, 2023, arXiv: 2304.02643
- [26] R. Deng, C. Cui, Q. Liu, T. Yao, L.W. Remedios, S. Bao, B.A. Landman, Y. Tang, L.E. Wheless, L.A. Coburn, K.T. Wilson, Y. Wang, A.B. Fogo, H. Yang, Y. Huo, Segment anything model (SAM) for digital pathology: assess zero-shot segmentation on whole slide imaging, in: Medical Imaging with Deep Learning, Short Paper Track, 2023.
- [27] Z. Shui, Y. Zhang, K. Yao, C. Zhu, S. Zheng, J. Li, H. Li, Y. Sun, R. Guo, L. Yang, Unleashing the power of prompt-driven nucleus instance segmentation, in: A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, G. Varol (Eds.), Computer Vision – ECCV 2024, Springer Nature Switzerland, Cham, 2025, pp. 288–304.
- [28] S. Abousamra, D. Belinsky, J. Van Arnem, F. Allard, E. Yee, R. Gupta, T. Kurc, D. Samaras, J. Saltz, C. Chen, Multi-class cell detection using spatial context representation, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 3985–3994. <https://doi.org/10.1109/ICCV48922.2021.00397>
- [29] O. Pina, E. Dorca, V. Vilaplana, Cell-DETR: efficient cell detection and classification in WSIs with transformers, in: Medical Imaging with Deep Learning, 2024.
- [30] H. Zhang, P. Liang, Z. Sun, B. Song, E. Cheng, CircleFormer: Circular Nuclei Detection in Whole Slide Images with Circle Queries and Attention, Springer Nature Switzerland, 2023, p. 493–502. https://doi.org/10.1007/978-3-031-43993-3_48
- [31] J. Huang, H. Li, X. Wan, G. Li, Affine-consistent transformer for multi-class cell nuclei detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 21384–21393.
- [32] J. Huang, H. Li, W. Sun, X. Wan, G. Li, Prompt-based grouping transformer for nucleus detection and classification, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Springer Nature Switzerland, 2023, p. 569–579. https://doi.org/10.1007/978-3-031-43993-3_55
- [33] T. Hassan, S. Javed, A. Mahmood, T. Qaiser, N. Werghe, N. Rajpoot, Nucleus classification in histology images using message passing network, Med. Image Anal. 79 (2022) 102480. <https://doi.org/10.1016/j.media.2022.102480>
- [34] W. Lou, X. Wan, G. Li, X. Lou, C. Li, F. Gao, H. Li, Structure embedded nucleus classification for histopathology images, IEEE Trans. Med. Imaging (2024) 1. <https://doi.org/10.1109/TMI.2024.3388328>
- [35] W. Lou, G. Li, X. Wan, H. Li, Cell graph transformer for nuclei classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2024.
- [36] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: International Conference on Learning Representations, 2017.
- [37] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Adv. Neural Inf. Process. Syst. 30 (2017).
- [38] K. Xu, W. Hu, J. Leskovec, S. Jegelka, How Powerful are graph neural networks?, in: International Conference on Learning Representations, 2019.
- [39] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: International Conference on Learning Representations, 2018.
- [40] J. You, R. Ying, J. Leskovec, Design space for graph neural networks, 34th Conference on Neural Information Processing Systems (NeurIPS) (2020).
- [41] A.L. Frei, A. Khan, P. Zens, A. Lugli, I. Zlobec, A. Fischer, GammaFocus: an image augmentation method to focus model attention for classification, in: Medical Imaging with Deep Learning, short paper track, 2023.
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [43] B. Delaunay, Sur la sphère vide. a la memoire de georges voronoï, Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na 1934 (6) (1934) 793–800.
- [44] G. Lu, D. Wang, X. Qin, S. Muller, J.V. Little, X. Wang, A.Y. Chen, G. Chen, B. Fei, Histopathology feature mining and association with hyperspectral imaging for the detection of squamous neoplasia, Sci. Rep. 9 (2019). <https://doi.org/10.1038/s41598-019-54139-5>
- [45] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in PyTorch, in: NIPS 2017 Workshop on Autodiff, 2017.
- [46] M. Fey, J.E. Lenssen, Fast Graph Representation Learning with PyTorch Geometric, 2019, 1903.02428
- [47] D. Tellez, G. Litjens, P. Bándi, W. Bulten, J.-M. Bokhorst, F. Ciompi, J. van der Laak, Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology, Med. Image Anal. 58 (2019) 101544. <https://doi.org/10.1016/j.media.2019.101544>
- [48] J. Bergstra, D. Yamins, D. Cox, Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures, in: International Conference on Machine Learning, PMLR, 2013, pp. 115–123.
- [49] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015.
- [50] S. Graham, Q.D. Vu, M. Jahanifar, S.E.A. Raza, F. Minhas, D. Snead, N. Rajpoot, One model is all you need: multi-task learning enables simultaneous histology image segmentation and classification, Med. Image Anal. 83 (2023) 102685. <https://doi.org/10.1016/J.MEDIA.2022.102685>
- [51] I.T. Young, P.W. Verbeek, B.H. Mayall, Characterization of chromatin distribution in cell nuclei, Cytometry 7 (5) (1986) 467–474.
- [52] S. Graham, Q.D. Vu, M. Jahanifar, S.E.A. Raza, F. Minhas, D. Snead, N. Rajpoot, One model is all you need: multi-task learning enables simultaneous histology image segmentation and classification, Med. Image Anal. (2022) 102685.
- [53] K. Sirinukunwattana, S.E.A. Raza, Y.-W. Tsang, D.R.J. Snead, I.A. Cree, N.M. Rajpoot, Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, IEEE Trans. Med. Imaging 35 (5) (2016) 1196–1206. <https://doi.org/10.1109/tmi.2016.2525803>
- [54] S. Graham, M. Jahanifar, A. Azam, M. Nimir, Y.-W. Tsang, K. Dodd, E. Hero, H. Sahota, A. Tank, K. Benes, et al., Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 684–693.
- [55] Z. Ying, D. Bourgeois, J. You, M. Zitnik, J. Leskovec, GNNExplainer: generating explanations for graph neural networks, in: H. Wallach, H. Larochelle, A. Beygelzimer, F.d. Alché-Buc, E. Fox, R. Garnett (Eds.), Adv. Neural Inf. Process. Syst., 32, Curran Associates, Inc., 2019.
- [56] T. Sakamoto, T. Furukawa, H.H.N. Pham, K. Kuroda, K. Tabata, Y. Kashima, E.N. Okoshi, S. Morimoto, A. Bychkov, J. Fukuoka, A collaborative workflow between pathologists and deep learning for the evaluation of tumour cellularity in lung adenocarcinoma, Histopathology 81 (6) (2022) 758–769. <https://doi.org/10.1111/his.14779>
- [57] A.L. Frei, R. Oberson, E. Baumann, A. Perren, R. Grobholz, A. Lugli, H. Dawson, C. Abbet, I. Lertxundi, S. Reinhard, A. Mookhoek, J. Feichtinger, R. Sarro, G. Gadiant, C. Dommann-Scherrer, J. Barizzi, S. Berezowska, K. Glatz, S. Dertinger, Y. Banz, R. Schoenegg, L. Rubbia-Brandt, A. Fleischmann, G. Saile, P. Mainil-Varlet, R. Biral, L. Giudici, A. Soltermann, A.B. Chaubert, S. Stadlmann, J. Diebold, K. Egervari, C. Bénére, F. Saro, A. Janowczyk, I. Zlobec, Pathologist computer-aided diagnostic scoring of tumor cell fraction: a swiss national study, Mod. Pathol. 36 (12) (2023) 100335. <https://doi.org/10.1016/j.modpat.2023.100335>
- [58] V. L'Imperio, G. Cazzaniga, M. Mannino, D. Seminati, F. Mascadri, J. Ceku, G. Casati, F. Bono, C. Eloy, E.G. Rocco, C. Frascarelli, M. Fassan, U. Malapelle, F. Pagni, Digital counting of tissue cells for molecular analysis: the quantum pipeline, Virchows Archiv (2024). <https://doi.org/10.1007/s00428-024-03794-9>
- [59] T.C. Smyrk, P. Watson, K. Kaul, H.T. Lynch, Tumor-infiltrating lymphocytes are a marker for microsatellite instability in colorectal carcinoma, Cancer (2001) 2417–2422.
- [60] K. Brummel, A.L. Eerikens, M. de Bruyn, H.W. Nijman, Tumour-infiltrating lymphocytes: from prognosis to treatment selection, Br. J. Cancer (2023). <https://doi.org/10.1038/s41416-022-02119-4>
- [61] I. Zlobec, M.D. Berger, A. Lugli, Tumour budding and its clinical implications in gastrointestinal cancers, Br. J. Cancer (2020). <https://doi.org/10.1038/s41416-020-0954-z>