Medical Image Analysis 79 (2022) 102480

Contents lists available at ScienceDirect

Medical Image Analysis

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

Nucleus classification in histology images using message passing network

Taimur Hassan[a], Sajid Javed[a,b,*], Arif Mahmood[c], Talha Qaiser[d], Naoufel Werghi[b], Nasir Rajpoot[d,e,f]

[a] Department of Electrical Engineering and Computer Science, Khalifa University of Science and Technology, Abu Dhabi, P.O. Box 127788, UAE
[b] Khalifa University Center for Autonomous Robotic Systems (KUCARS), Abu Dhabi, P.O. Box 127788, UAE
[c] Department of Computer Science, Information Technology University, Lahore, Pakistan
[d] Department of Computer Science, University of Warwick, Coventry CV4 7AL, U.K
[e] Department of Pathology, University Hospitals Coventry and Warwickshire, Walsgrave, Coventry CV2 2DX, U.K
[f] The Alan Turing Institute, London NW1 2DB, U.K

ARTICLE INFO

Article history:
Received 10 January 2022
Revised 7 March 2022
Accepted 10 May 2022
Available online 14 May 2022

MSC:
41A05
41A10
65D05
65D17

Keywords:
Nuclear communities
End-to-end graph learning
Whole slide images
Computational pathology

ABSTRACT

Identification of nuclear components in the histology landscape is an important step towards developing computational pathology tools for the profiling of tumor micro-environment. Most existing methods for the identification of such components are limited in scope due to heterogeneous nature of the nuclei. Graph-based methods offer a natural way to formulate the nucleus classification problem to incorporate both appearance and geometric locations of the nuclei. The main challenge is to define models that can handle such an unstructured domain. Current approaches focus on learning better features and then employ well-known classifiers for identifying distinct nuclear phenotypes. In contrast, we propose a message passing network that is a fully learnable framework build on classical network flow formulation. Based on physical interaction of the nuclei, a nearest neighbor graph is constructed such that the nodes represent the nuclei centroids. For each edge and node, appearance and geometric features are computed which are then used for the construction of messages utilized for diffusing contextual information to the neighboring nodes. Such an algorithm can infer global information over an entire network and predict biologically meaningful nuclear communities. We show that learning such communities improves the performance of nucleus classification task in histology images. The proposed algorithm can be used as a component in existing state-of-the-art methods resulting in improved nucleus classification performance across four different publicly available datasets.

1. Introduction

Recent years have witnesed significant growth and development in the field of computational pathology leveraged by the advancement of slide scanners which create multi-resolution, multi-gigapixel Whole Slide Images (WSIs) (Bui et al., 2019; Kather et al., 2019; Oliveira et al., 2021; Song et al., 2020; van der Laak et al., 2021; Pinckaers et al., 2020). These images are equally effective for the visual examination by pathologists using a microscope and are considered as the gold standard for better cancer diagnosis and prognosis (van der Laak et al., 2021; Song et al., 2020; Srinidhi et al., 2021; Fuchs and Buhmann, 2011; Campanella et al., 2019). Expert pathologists visually analyse the histology landscape using WSIs or tissue slides to assess the degree of malignancy of cancer by analyzing the tumor micro-environment (van der Laak et al., 2021; Kather et al., 2019; Song et al., 2020). The common availability of these WSIs is playing a vital role in the development of a new discipline of artificial intelligence and machine learning techniques within the computational pathology field Tizhoosh and Pantanowitz (2018). An important aim of these developments is to obtain reliable tools to assist pathologists for improving the pathology workflow by yielding reproducible and more objective results (Campanella et al., 2019). Such tools can not only reduce the pathologist burden but can also lead towards improved patient care (Kather et al., 2019; van der Laak et al., 2021; Srinidhi et al., 2021).

Due to gigantic size of WSIs, many computational challenges arise when storage, transfer, or processing is required. For instance, Fig. 1 shows a sample WSI of colorectal cancer (CRC) captured

* Corresponding author at: Department of Computer Science, University of Warwick, Coventry, CV4 7AL, U.K.
E-mail addresses: sajid.javed@ku.ac.ae (S. Javed), n.m.rajpoot@warwick.ac.uk (N. Rajpoot).

https://doi.org/10.1016/j.media.2022.102480
1361-8415/

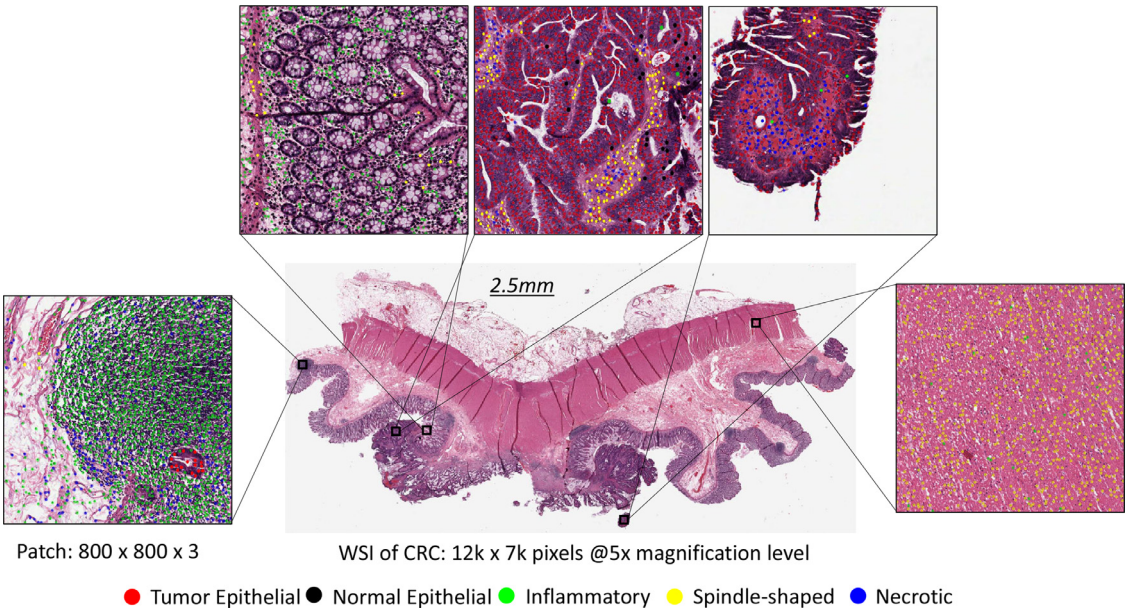


Fig. 1. Exemplar Whole Slide Image (WSI) of colorectal cancer taken from The Cancer Genomic Atlas (TCGA) (Tomczak et al. (2015)) at 5× magnification level. Some high resolution patches are shown at 20× magnification level for nucleus detection and classification. Different colors show distinct nuclei types.

at a maximum resolution of 40× containing 100k × 100k pixels. The existing machine learning-based approaches suffer with increased computational complexity due to large size of WSIs (Srinidhi et al., 2021). Training these algorithms using entire WSIs at a higher magnification is intractable with common hardware resources. Typically, a WSI is divided into small tiles or patches and each tile is then independently stored and processed. The training and testing of machine learning algorithms is also performed using these tiles while the results are integrated to obtain full slide-level inference (Srinidhi et al., 2021; Campanella et al., 2019; Song et al., 2020).

In cancerous WSIs, many different types of densely connected nuclear groups, also known as communities, exist that play an important role in the progression of cancer (Alberts et al., 2015). For instance, a cellular community can be defined based on the intracellular signalling process. Similarly, tissue communities have also been reported by Javed *et al.* based on homogeneity of the tissue and nuclei components (Javed et al., 2018, Javed et al., 2019, 2020a, 2020b). In these networks, representation of a particular tissue patch constitutes a node and pairwise similarity is used as an edge weight. Communities are groups of nodes in a network which are strongly or densely connected within each group and weakly or sparsely connected between groups. Community detection has been explored using spectral clustering, matrix factorization, and statistical-based methods as well as deep learning techniques (Liu et al., 2020; Brasó and Leal-Taixé, 2020; Von Luxburg, 2007; Abbe, 2017; Van Lierde et al., 2019; He et al., 2021; Mahmood et al., 2016; Mahmood and Small, 2015). Detection of nuclear communities within the tumor micro-environment has a significant clinical importance. For example, a community of lymphocytes, debris, or tumor epithelial nuclei may have significant clinical implications for survival analysis. Moreover, such communities can increase the performance of nuclear classification, tissue phenotyping, understanding disease heterogeneity and predicting different clinical outcomes e.g. HER2 prediction from cancerous WSIs.

In the current work, we propose a nucleus classification algorithm using end-to-end message passing network (Battaglia et al., 2004; Battaglia et al., 2018; Gilmer et al., 2017; Kipf et al.. 2018). We pose the problem of nucleus classification as detecting biologically meaningful nuclear communities in histology images. In classical community detection methods, pairwise costs are computed and used to define network structure that is then partitioned to find distinct communities. In contrast, we propose to solve the nuclear community detection problem by global reasoning over an entire histology patch. More specifically, in the proposed algorithm, information flows from each node consisting of nucleus appearance information to all other nodes (e.g., other nuclei) in the patch through edges based on nearest neighbor strategy. It results in a global context which enhances the community structure within that patch. For this purpose, a series of message passing steps are performed over the entire network that propagates the contextual information across directly connected neighbors resulting in strong edges within the nuclear communities and weak edges across nuclear communities. In each message passing step, nodes transfer nuclei appearance information to their incident edges while sharing the appearance information with directly connected nodes. The updated embeddings are learned for all nodes and edges incorporating the network topology and contextual information. Thus, in the proposed algorithm, nuclear communities are directly learnt by passing information from one node to its neighbors while readjusting the edge cost and the node representations. After a few iterations, the weight of weak edges across two different nuclear communities decreases while the weight within the same nuclear community increases. This fact reveals a clear community structure within a complex network of nuclear components.

We explore the significance of communities in the context of histological landscape for nucleus classification. Most existing methods such as HoverNet (Graham et al., 2019), SC-CNN (Sirinukunwattana et al., 2016), TSP-CNN (Tofighi et al., 2019), SCC-DCF (Javed et al., 2021) focus on nuclei detection while nuclei classification has been treated separately. Some recent works such as Graham *et al.* report extensive experimentation on nucleus segmentation while nucleus classification has not been well explored (Graham et al., 2021). Nuclei classification is important for measuring the degree of malignancy of different cancers and plays vital role in cancer diagnosis. In the current work, we improve nuclei classification using the proposed community detection algorithm. We observe that community structure is helpful in improving the nuclei classification. It is because the intra-phenotype edges become stronger compared to the inter-phenotype edges. Therefore,

all nodes in the same community must belong to the same nuclei type. Rigorous experimental evaluations demonstrate the effectiveness of communities in nuclei classification task. Following are the main contributions of this work:

1. A nucleus classification algorithm is proposed using message passing driven end-to-end learnable community detection method in histology images. The improved edge and node representations are obtained by using contextual information in the nearest neighbor graph.
2. The vanilla version of the message passing scheme is extended to multiple nodes based representations which has resulted in improved nuclei classification performance.
3. The proposed algorithm has been used as a component with existing state-of-the-art methods including HoverNet (Graham et al., 2019)) and SC-CNN (Sirinukunwattana et al., 2016) resulting in improve nuclei classification performance.

The rest of this paper is organized as follows: Section 2 presents a literature review on nucleus classification. Section 3 explains the proposed algorithm in detail. Section 4 presents the exhaustive experimental evaluations while Section 5 draws the conclusion and future directions of the current work.

2. Literature review

Most of the existing methods have focused on nuclei detection and segmentation while nuclei classification has not been thoroughly explored (Graham et al., 2019; Tofighi et al., 2019; Javed et al., 2021; Graham et al., 2021; Gamper et al., 2020; Sun et al., 2021; Singh et al., 2017). Most recent methods use a simple fully connected network followed by a softmax layer for nuclei classification (Graham et al., 2019; Sirinukunwattana et al., 2016). We categorize the existing nuclei classification methods into classical machine learning and deep learning methods (Srinidhi et al., 2021; Irshad et al., 2013).

Classical machine learning methods have been employed for nucleus classification in diverse histopathology applications for cancer diagnosis (Xing and Yang, 2016; Demir and Yener, 2004). For instance, nuclear shape, texture, and size based handcrafted features are exploited for nuclear pleomorphism grading for the breast cancer images (Dalle et al., 2009; Cosatto et al., 2008). To discriminate between malignant and normal nuclei in unsupervised manner, Nguyen *et al.* proposed shape features in prostate cancer histology images (Nguyen et al., 2011). Lymphocytes/stromal classification of H&E stained breast cancer WSIs is performed by Yuan *et al.* based on morphological features and nuclei segmentation (Yuan et al., 2012). Adaboost classifier has also been used by Sharma *et al.* for nuclei segmentation and classification using texture, morphological and intensity-based features (Sharma et al., 2015).

The previous approaches have shown promising results for nuclei classification. However, handcrafted features are generally sub-optimal for the final aim of nuclei classification. In contrast, deep features are optimized to increase the classification performance. Therefore, recent deep learning features extraction techniques have reported excellent performance for nuclei classification (Graham et al., 2019; 2021; Gamper et al., 2020; Raza et al., 2019; Abousamra et al., 2021). For instance, Cruz-Roa *et al.* proposed a deep learning-based approach for cell classification (Cruz-Roa et al., 2013). Malon *et al.* trained CNN classifier to classify mitotic and non-mitotic cells (Malon and Cosatto, 2013). Wang *et al.* employed an ensemble of cascaded CNN features for mitotic cell classification (Wang et al., 2014). Sirinukunwattana *et al.* proposed spatially constrained deep network for nuclei detection and classification in routine colon cancer histology images (Sirinukunwattana et al., 2016). This method employed neighboring cell based classification for four classes including epithelial, lymphocytes, spindle-shaped, and miscellaneous. A simple softmax classifier is trained on sparse annotated data which struggled to distinguish between lymphocytes and spindle-shaped nuclei. Gao *et al.* used transfer learning method for breast nuclei classifications using LeNet-5 architecture (Gao et al., 2016). Deep regression models have also been employed for the nuclei classification task. For instance, Hu *et al.* proposed generative adversarial network for tissue as well as cell classification of bone marrow histology images (Hu et al. (2018)). Xing *et al.* used fully connected network to perform regression for combined nuclei detection and classification tasks in pancreatic cancer (Xing et al. (2019)). Qu *et al.* proposed perceptual loss for segmentation and classification of lung histology images (Qu et al., 2019). Recently, Graham *et al.* performed nuclear segmentation and classification using horizontal and vertical distance maps (Graham et al., 2019). The nuclear classification was performed by aggregating the pixel-level nuclear type predictions within each nuclear instance using ResNet architecture. Tripathi *et al.* used fusion of CNN features for nuclei segmentation and classification (Tripathi and Singh (2020)). The fused features were input into a multi-layer perceptron for nuclei classification.

Most of these methods mainly focused on nuclei detection and segmentation while nuclei classification was performed as a secondary task. In contrast to these approaches, we propose an algorithm which leverages community detection for nuclei classification. Our proposed algorithm can be used as a component with any existing state-of-the-art nuclei detection/segmentation and classification method with an effect of increased performance. In our experiments, we observe significant classification performance improvement of a number of existing methods when integrated with the proposed algorithm.

3. Proposed nucleus classification using community detection algorithm

The concept diagram of the proposed Nucleus Classification using Community Detection (NCCD) algorithm is shown in Fig. 2. The NCCD algorithm consists of a pre-processing step for nucleus detection and segmentation. From each detected nucleus, appearance features are extracted using a deep neural network. Also, an initial nuclei-level network is constructed using nearest neighbor strategy which is then refined using message-passing scheme. The communities in the refined network are then detected using an edge classification algorithm. Edges that are classified as intra-community are retained while the inter-community edges are removed. The connected components in the resulting network are considered as communities. For the purpose of nucleus classification task, each community is then classified using deep neural network. The main components of the proposed algorithm are discussed in detail in the following subsections.

3.1. Problem formulation

In nuclear community detection problem, a set of detected nuclei e.g., $N = \{n_1, n_2, \cdots, n_m\}$, is given where m is the total number of nuclei in a given WSI patch. Each detected nucleus is represented as $n_i = \{r_i, c_i\}$, the nucleus centre or a centroid of the segmented region. We model the set of nuclei as an undirected graph $\mathbf{G} = \{\mathbf{V}, \mathbf{E}\}$, where $\mathbf{V} = \{1, \cdots, m\}$ is a set of m nodes each corresponding to a unique nuclei. The set of edges $\mathbf{E}$ is the subset of $\mathbf{V} \times \mathbf{V}$ representing the connections between distinct nuclei. The task is to divide this graph into biologically meaningful non-overlapping nuclear communities $\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$ such that $|\mathbf{x}_j| \geq k_{\min}$, where $k_{\min}$ is the minimum size of a nuclear community. All nuclei

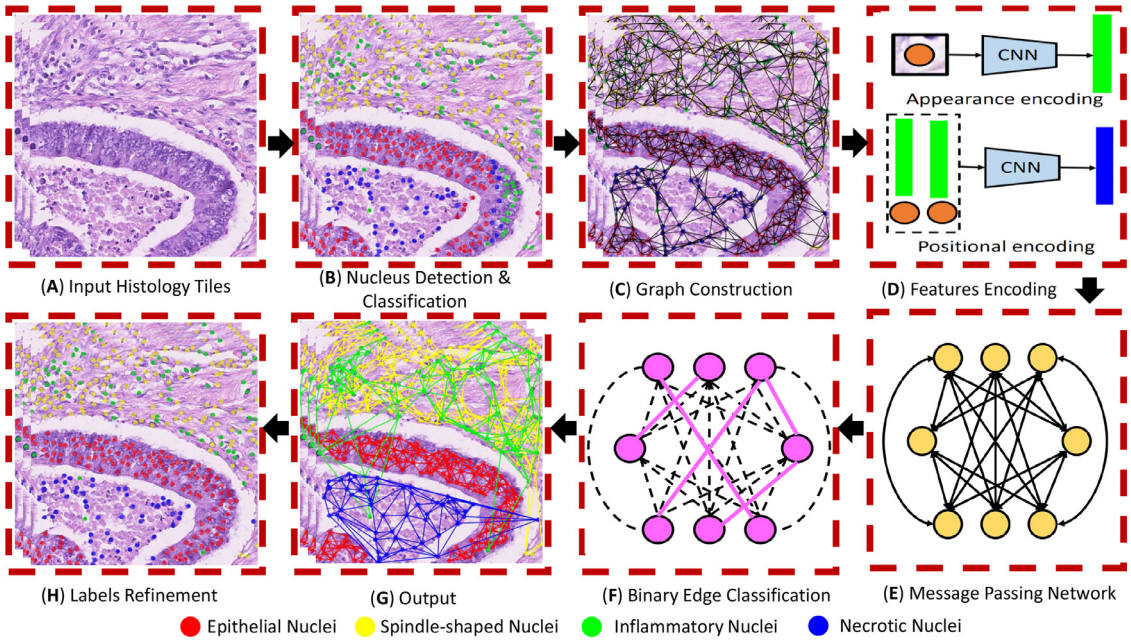


Fig. 2. Schematic illustration of the proposed NCCD algorithm for classifying distinct nuclei. Step (A) shows an input histology image patch. Step (B) shows the results of the nucleus detection and weak classification using an existing model. Step (C) shows the initial graph construction based on nearest neighbor scheme where each nucleus centroid is used as a node and each centroid is connected with its nearest nearest neighbors. Step (D) Shows node appearance encoding and edges positional encoding using pre-trained CNN network. Step (E) shows the message passing network where nodes and edges exchange messages and update accordingly. Step (F) shows binary edge classification for weak edge pruning. Step (G) shows the refined graph based on (F). Step (H) shows the classification label refinement based resulting in improved classification performance..

nodes $\mathbf{n}_i \in \mathbf{x}_j$ share common biological characteristics which are of clinical significance such as a particular nuclear phenotype.

3.2. Nucleus detection, segmentation, and classification

In the proposed NCCD algorithm, nucleus detection, segmentation, and classification are pre-processing steps. For this purpose, any existing state-of-the-art method may be employed. In our experiments, we use different types of nucleus detection methods including Faster R-CNN (Ren et al. (2015)), SCC-DCF (Javed et al. (2021)), TSP-CNN (Tofighi et al. (2019)), and SC-CNN (Sirinukunwattana et al., 2016). For segmentation purpose, we employed HoverNet (Graham et al., 2019). From these methods, we get the nuclei centres $n_i = \{r_i, c_i\}$ and the bounding box coordinates w_i, h_i (Fig. 2 Step (B)). For nucleus classification, we employ Faster R-CNN, SC-CNN, and HoverNet methods as initial labeling. As a baseline, a standalone nucleus classification network is also trained to provide initial label information. We consider the classification labels of the nuclei obtained by these methods as noisy labels and therefore we propose to refine these labels using the proposed community detection algorithm.

3.3. Initial nuclear graph construction

Given the nuclei locations $n_i = \{r_i, c_i\}$, a nearest neighbor-based graph $\mathbf{G}$ is constructed by computing the Euclidean distance between each pair of nuclei $d_{i,j} = ||n_i - n_j||_2$. For each WSI patch, we construct a nuclear graph such that the vertices correspond to the centroids of the nuclei and the edges are assigned using minimum Euclidean distance. Each nucleus is connected to only its k-nearest nuclei keeping in view the inter-cellular signalling (Alberts et al., 2015). The nucleus on the opposite side of tissue constituent white space also known as lumen and endothelium (forming the microvessels) do not communicate to each other and therefore remain disconnected. A visual representation of the nearest-neighbor nuclear graph is shown in Fig. 1 (C). Our nuclear graph has a very specific structure that can be exploited for extraction of clinically significant communities. We aim to encode nuclei-specific inductive bias in the proposed network especially when the nodes are updated. The update steps shown in Fig. 1 (E) depict that information is shared with the neighbors and aggregated over all neighbors of a node. This process results in an updated embedding in the next iteration.

3.4. Appearance modelling

For each segmented nucleus n_i, a bounding box $\mathbf{b}_i$ of size w_i and h_i around the segmented region is extracted which is used for appearance feature extraction and positional encoding. For appearance features extraction, we employ a pre-trained ResNet50 network (He et al., 2016) trained on histology images for nuclei segmentation and classification (Graham et al., 2019).

$$\mathbf{f}_i^0 = \text{ResNet50}(\mathbf{b}_i). \tag{1}$$

The use of deep feature representations assists to encode the varying nuclear shapes and texture appearances. In addition to that, we also perform positional encoding to obtain the representations between each nearest-neighbor pair of nuclei. Let $(r_i, c_i,\ h_i,\ w_i)$ and $(r_j, c_j,\ h_j,\ w_j)$ be the top-left coordinates and height and width of the two bounding boxes. If these two nuclei are connected in the nearest-neighbor graph, the positional encoding for the edge $\mathbf{e}_{i,j}$ is computed as: $\mathbf{p}_{i,j} = (\frac{2(r_i - r_j)}{h_i + h_j}, \frac{2(c_i - c_j)}{h_i + h_j}, \log\frac{h_i}{h_j}, \log\frac{w_i}{w_j})^\top$. The positional encoding is then concatenated using ℓ_2 norm of the differential of the appearance encoding and then input into a Multi-layer Perceptron (MLP) model to get the edge representation as:

$$\mathbf{e}_{i,j}^0 = \text{MLP}_e(||\mathbf{f}_i - \mathbf{f}_j||_2, \mathbf{p}_{i,j}, d_{i,j}), \tag{2}$$

where MLP_e consists of three fully connected layers which is used to get the initial edge embeddings $\mathbf{e}_{i,j}^0$ (Fig. 2 Step (D)). The first term, $||\mathbf{f}_i - \mathbf{f}_j||_2$, in Eq. 2 is a type of perceptual loss function

as defined by Johnson et al. (2016). This loss function measures high-level perceptual and semantic differences between two nuclei patches. We make use of a pre-trained network on histology image classification task as described in Eq. 1. A smaller value of this loss function actually shows that the pixels of the two nuclei have similar feature representations which are actually the activation's of the last layer of the ResNet50 network. In this representation, the nuclear content and overall spatial structure are preserved however, the shape variations are ignored. Therefore, the smaller values of this loss means that the two nuclei are perceptually similar and larger value means these to be perceptually dissimilar (Johnson et al., 2016).

3.5. Message passing network

The aim of message passing network is to learn a function for propagating information encoded in the node appearance features and edge encoding across the graph **G**. The initial node and edge embeddings $\mathbf{f}_i^0$ and $\mathbf{e}_{i,j}^0$ are obtained by Eqs. (1) & (2). The propagation procedure consists of appearance model updates for nodes and embedding updates for edges in each message passing iteration (Battaglia et al., 2004; Battaglia et al., 2018; Gilmer et al., 2017; Kipf et al., 2018). The messages are propagated sequentially using a fixed number of iterations $t = \{1, \cdot, \cdot, \cdot, t_l\}$ across the whole graph **G** (Battaglia et al., 2018). The features initially extracted using Eqs. (1)-(2) are used as the 0-th iteration.

Each iteration is divided into two fundamental steps, the first step consists of a message updating the edge encoding in the current iteration (t) using the encoding from the previous $(t-1)$ iteration:

$$\mathbf{e}_{i,j}^{(t)} = \mathrm{M}_e(\mathbf{f}_i^{(t-1)}, \mathbf{f}_j^{(t-1)}, \mathbf{e}_{i,j}^{(t-1)}). \tag{3}$$

where M_e is a learnable function in the form of MLP consisting of few fully connected layers as described in Section 3.8. The next step in each message passing iteration is node appearance features updating:

$$\mathbf{f}_i^{'(t)} = \frac{1}{|\mathrm{N}_i|} \sum_{\mathbf{e}_{i,j} \in |\mathrm{N}_i|} \mathrm{M}_a(\mathbf{f}_i^{'(t-1)}, \mathbf{e}_{i,j}^{(t)}), \tag{4}$$

where N_i is the set of neighboring nodes of n_i and M_a is a learnable function. The summation here ensures that the contextual information from the neighboring nuclei will flow towards the current nucleus.

Although, the message passing scheme described by the above equations incorporates the contextual information into the node and edge embeddings, the geometric position of a node in a particular neighborhood has not been considered. By using such information, the geometric position can also be used to improve the node and edge embeddings. For this purpose, we propose to partition the neighbors of a particular node n_i based on the angles made by the incident edges due to geometric nuclear location. A fine-grained partitioning can also be made however, some partitions may remain empty if no edges are incident in those angular ranges. We therefore propose to use coarser partitions such that we divide the incident edges into only two partitions. The up-partition consists of those edges which span angles from $0^o - 180^o$ and the down-partition consists of edges in the $181^o - 360^o$. For each partition, a different MLP is trained and the node update $\mathbf{u}_i^{(t)}$ for the up-partition and $\mathbf{d}_i^{(t)}$ for the down-partition is computed as follows:

$$\mathbf{u}_i^{(t)} = \frac{1}{G_i^u} \sum_{j \in |G_i^u|} \mathrm{M}_u(\mathbf{f}_i^{(t-1)}, \mathbf{e}_{i,j}^t, \mathbf{f}_i^0), \tag{5}$$

$$\mathbf{d}_i^{(t)} = \frac{1}{G_i^d} \sum_{j \in |G_i^d|} \mathrm{M}_d(\mathbf{f}_i^{(t-1)}, \mathbf{e}_{i,j}^t, \mathbf{f}_i^0) \tag{6}$$

where M_u and M_d are two learnable functions such as MLPs for up and down-partition of the neighbors of n_i consisting of fully connected layers. G_i^u and G_i^d are sets of nodes in the up and down partitions. The complete node update message is computed as a concatenation of the $\mathbf{u}_i^t$ and $\mathbf{d}_i^t$ as computed by the Eqs. (5) & (6) as follows:

$$\mathbf{f}_i^t = \mathrm{M}_v([\mathbf{u}_i^{(t)}, \mathbf{d}_i^{(t)}]), \tag{7}$$

where M_v is another learnable MLP. In each iteration, all nodes share their information with their immediate neighbors. Therefore, after t_l iterations, information propagates up to t_l hops. Thus, similarity between nodes of the same semantic category increases with each iteration while the similarity across different semantic categories decreases. This effect is reflected in the feature vector of both edges and nodes resulting in a distinct community structure in the nuclear graph. The steps described above are shown in Fig. 2 Step (E).

3.6. Community detection

In this step, weak edges are detected and pruned while strong edges are retained to further enhance the community structure in the nuclear network. We train a multi-layer perceptron model using binary cross-entropy loss function which predicts the strength of each edge as a continuous variable from 0 to 1 $\hat{l}_{(i,j)}^{(t)} = \mathrm{Cdt}(\mathbf{e}_{i,j}^{(t)})$. For this purpose, the initial edge embedding as well as the updated embedding after each iterations are used as follows:

$$\mathcal{L}^{(t)} = \frac{-1}{|N_e|} \sum_{e_{(i,j)} \in E} l_{i,j} \log(\hat{l}_{i,j}^{(t)}) + (1 - l_{(i,j)}) \log(1 - \hat{l}_{(i,j)}^{(t)}), \tag{8}$$

where $l_{(i,j)}$ is the edge-label obtained from the training data using annotations by expert pathologists. If an edge $\mathbf{e}_{i,j}$ in Graph **G** is found to be connecting the nuclei of the same class (e.g., tumor epithelial cells) then $l_{(i,j)} = 1$ otherwise $l_{(i,j)} = 0$. The accumulated loss function over t_l iterations is obtained as:

$$\mathcal{L} = \frac{1}{t_l} \sum_{t=t_0}^{t_l} \mathcal{L}^{(t)}, \tag{9}$$

while training the $\mathrm{Cdt}(\mathbf{e}_{i,j}^{(t)})$ which is a fully connected MLP network the loss $\mathcal{L}$ is minimized over the training data. The trained MLP is then used for the edge pruning task in the test histology networks for the purpose of cellular community detection as shown in Fig. 2 Steps (F)-(G). For the test networks, if $\mathcal{L} \geq 0.50$ then edge is retained otherwise it is pruned.

3.7. Refining nucleus classification

In the previous sub-sections, we have discussed the task of community detection in the nuclei network with the detected communities consisting of homogeneous nuclei. Using this property of detected communities, we remove noise from the classification labels obtained by a simple node classification network. For each detected community, we compute the distribution of different nuclei and we assign the component having maximum probability to all nuclei in the whole community. Let $\mathbf{x}_j$ be a given community computed by the proposed algorithm. The probability of a particular label z over $\mathbf{x}_j$ is obtained as follows:

$$P(z|\mathbf{x}_j) = \frac{n_{z,j}}{\sum_{z_n=1}^{x_j} n_{z,j}}, \tag{10}$$

where $n_{z,j}$ is the number of nuclei having label z in the $\mathbf{x}_j$ community and z_n are the total number of nucleus classes. The aim of the proposed algorithm is each detected community should correspond to a particular nuclear phenotype (e.g., tumor epithelial,

necrotic, fibroblast or miscellaneous). Therefore, we obtain significant noise removal from the current state-of-the-art nuclei classification algorithms resulting in improved classification performance. In Eq. 10, if a tie happens between two classes, we may retained the original noisy labels of each nuclei in such a community as predicted by the baseline method.

3.8. Network architecture

The proposed algorithm consists of seven distinct networks including node appearance modeling network as described by Eq. (1), edge positional encoder described by Eq. (2), the network for updating the edge encoding described in Eq. (3), an MLP updating the node appearance described in upper partition is M_u Eq. (5), and in the down partition M_d Eq. (6). The combined node embedding using up and down node embeddings is computed using M_v in Eq. (7). Finally, the community detection by edge pruning network as described by Eq. (8).

The node appearance modeling network consists of a modified version of ResNet50 (He et al., 2016) initially trained on ImageNet dataset and fine tuned for cell classification datasets used in our experiments. The initial extracted nucleus patch consisting of $31 \times 31 \times 3$ for the case of nucleus detection by keeping the detected nucleus at the centre. For the case of nuclear segmentation, actual size of each nucleus is extracted as a patch. The patch is then resized to $64 \times 64 \times 3$ and input to the network which consists of multiple convolutional layers followed by global average pooling and fully connected layers. The final feature encoding $\mathbf{f}_i^0 \in \mathbb{R}^{32}$ is obtained. A softmax classification layer is used for multi-class nuclei classification task. For the initial positional encoding of edges, an MLP is employed consisting of three fully connected layers resulting in an embedding $\mathbf{e}_{i,j} \in \mathbb{R}^{16}$.

For the message passing updates, messages for updating the edge encoding are obtained by using M_e. For updating node encoding, three different MLPs (M_u, M_d and M_v) are used. All these MLPs consist of two fully connected layers and result in the same size of edge update as $\mathbf{e}_{i,j} \in \mathbb{R}^{16}$ and node update as $\mathbf{f}_i \in \mathbb{R}^{32}$. The edge pruning network for the purpose of community detection also consists of two fully connected layers, first followed by ReLU and second followed by sigmoid activation functions.

The proposed algorithm is trained in an end-to-end fashion using the error gradients produced by the edge pruning network which are propagated to message update networks and edge positional encoding network, while the node appearance encoder is independently trained as discussed above.

4. Experimental evaluations

The proposed Nucleus Classification using Community Detection (NCCD) algorithm is evaluated for nucleus classification task on four different histology images datasets: Colorectal Nuclear Segmentation and Phenotypes (CoNSeP) (Graham et al., 2019), Large-Scale Dataset for Colonic Nuclear Instance Segmentation and Classification (LIZARD) (Graham et al., 2021), CRCHistoPhenotypes (Sirinukunwattana et al., 2016) and multiple cancer types dataset called PanNuke (Gamper et al., 2020). The results are compared with seven existing state-of-the-art nucleus classification methods: Spatially Constrained Convolutional Neural Network (SC-CNN) (Sirinukunwattana et al., 2016), Micro-Net (Raza et al., 2019), Deep regression of the Distance map (DIST) (Naylor et al., 2018), Faster R-CNN (Ren et al., 2015), Mask-RCNN (He et al., 2017), Horizontal and Vertical distances prediction Network (HoVer-Net) (Graham et al., 2019) and DET U-Net (Xie et al., 2018). The proposed NCCD algorithm is used as a component with the detection performed by HoverNet, SC-CNN, and Tunable Shape Prior Convolutional Neural Network (TSP-CNN) (Tofighi et al., 2019), and SCC-DCF (Javed et al., 2021) resulting in improved performance. The implementations of these methods are obtained from the original authors and their recommended parameters are used for fair comparison. For the proposed NCCD algorithm, a maximum of 15 training epochs, $k = 10$ nearest neighbors (ablation study performed in Fig. 6(b)), and $t_l = 14$ message passing iterations (ablation study performed in Fig. 6(a)) are used in Eqs. (3)-(9).

4.1. Baselines of the proposed NCCD algorithm

The performance of the three different baselines for nucleus detection and classification is compared. For each baseline, the proposed NCCD algorithm has improved significant performance. These baselines include SC-CNN (Sirinukunwattana et al., 2016), Faster R-CNN (Ren et al., 2015), and HoverNet (Graham et al., 2019). The corresponding variants are named as SC-CNN-NCCD, FR-CNN-NCCD, and HoverNet-NCCD. Essentially, the proposed NCCD algorithm remains the same while the initial nucleus detection, segmentation, and classification varies with the variations of baselines. In all cases, significant classification performance improvement is observed demonstrating the effectiveness of the proposed NCCD algorithm.

4.2. Nucleus classification evaluation

The nucleus classification performance for the class z is evaluated using F_c measure score as:

$$F_c = 2 \times \frac{Recall_z \times Precision_z}{Recall_z + Precision_z}, \quad \text{where}$$
$$Recall_z = \frac{TP_z}{TP_z + FN_z}, \quad Precision_z = \frac{TP_z}{TP_z + FP_z}, \tag{11}$$

where TP_z denotes the True Positives which are the number of nuclei belonging to class z and also predicted as class z, FN_z is the False Negatives which are the number of nuclei belonging to class z but predicted as some other class, FP are the False positives which are nuclei not belonging to class z but predicted as class z. The aim is to maximize F_c measure so that its value is close to one. The weighted F measure is computed as a weighted average of F_c overall all classes as given below:

$$\widehat{F} = \sum_{z=1}^{n_z} p_z F_c, \tag{12}$$

where $p_z = n_z/n$ is the probability of the z-th class, n_z are the number of samples in that class, and n is the total number of nuclei samples.

4.3. Datasets for nucleus classification

We have evaluated the nucleus classification performance using four different publicly available datasets. In the below sub-sections, we explain each dataset in more detail.

4.3.1. CRCHistoPhenotypes Dataset (Sirinukunwattana et al., 2016)

This dataset was introduced for nucleus detection and classification tasks (Sirinukunwattana et al., 2016). It contains 100 H&E stained CRC histology images of size 500×500 extracted at $20\times$ magnification level (0.55 μm/pixel) and cropped from 10 WSIs of 9 different patients. Experienced pathologists performed manual annotations of 29,756 nuclei centres. It has 22,444 labeled nuclei having four classes including epithelial (7722), inflammatory (6971), miscellaneous (2039), and fibroblast (5712). The epithelial class contained both normal epithelial and tumor epithelial nuclei. The wide variation of normal and malignant tissue has made this dataset very challenging. In our experiments, we

employed two fold cross validation by randomly splitting training and testing using the protocol defined by the original authors (Sirinukunwattana et al., 2016).

4.3.2. Consep dataset (Graham et al., 2019)

This dataset consists of 41 large sized patches having 1000 × 1000 pixels at 40× magnification level which are extracted from 16 WSIs of colorectal adenocarcinoma stained with H&E. All 16 WSIs belong to different patients. Expert pathologists manually annotated 24,319 nuclei boundaries. Every nucleus was labelled in one of the categories: normal and malignant/dysplastic epithelial, inflammatory, miscellaneous, endothelial, muscle, and fibroblast. By following the protocols of HoverNet (Graham et al., 2019), we grouped together normal and malignant/dysplastic epithelial into an epithelial class, and endothelial, muscle, and fibroblast classes are grouped together into a spindle-shaped nuclei class. Using the same official protocols, we have used 27 patches for training and 14 patches for testing. The test split contains nuclei labels as: normal (503), malignant/dysplastic epithelial (2764), inflammatory (1569), miscellaneous (580), endothelial (80), muscle (495), and fibroblast (2312). Sample images of this dataset are shown in Fig. 5. This dataset also introduces the challenges of heterogeneity and nuclei morphological variations as well as clutteredness.

4.3.3. PanNuke Dataset (Gamper et al., 2020)

This dataset has automatically annotated nuclei boundaries which are then validated by veteran pathologists. PanNuke is one of the most diverse nuclei segmentation and classification datasets. It consists of 19 varying cancer types in 481 histology patches at 40*x* resolution level which have over 189,744 annotated nuclei boundaries. The number of labelled nuclei in each of the five classes, in each of the three folds, are neoplastic epithelial (25972, 22418, 28081), non-neoplastic epithelial (8814, 8833, 8818), inflammatory (10460, 10285, 10521), connective (16268, 16636, 17327), and dead (938, 859, 1028). Images of size 256 × 256 are randomly extracted from these patches and split into training, validation, and testing folds. The role of each fold is switched resulting in three fold cross validation experiments. The performance is reported as an average over the three folds by following the experimental protocols defined in Gamper et al. (2020).

4.3.4. Lizard dataset (Graham et al., 2021)

This dataset consists of 291 large histology WSI patches with an average size of 1016 × 917 pixels which are extracted at 20× resolution level. The dataset is collected from six different datasets including TCGA (Tomczak et al., 2015), PanNuke (Gamper et al., 2020), DigestPath, CONSEP (Graham et al., 2019), CRAG, and GLAS. Publicly available annotations in LIZARD consists of 431,913 annotated nuclei distributed over six distinct classes including Epithelial (210372), Lymphocyte (92238), Plasma (24861), Neutrophil (4116), Eosinophil (1979), and Connective (97347).

4.3.5. Experimental protocols

The evaluation of the proposed NCCD algorithm has been performed using two different protocols. In the first protocol named as NCCD-GT, the ground-truth nuclei detections or segmentations are used to compare the classification performance of different methods. In this protocol, the initial classification in the proposed NCCD is performed using two fully connected layers which are trained using the particular training dataset with categorical cross-entropy loss function. The classification is then improved using our proposed message passing network. In the second protocol named as Algo-NCCD, the performance of a particular algorithm is compared with and without NCCD-based label refinement. For instance, the performance of HoverNet is compared with HoverNet-NCCD and performance of SC-CNN is compared with SC-CNN-NCCD.

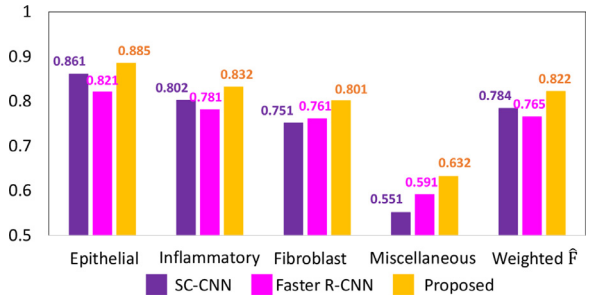


Fig. 3. Classification performance evaluation of different methods using ground-truth nuclei annotations on CRCHistoPhenotypes dataset (Sirinukunwattana et al. (2016)) using individual nucleus class F_1 score and weighted average $\widehat{F}$ score.

4.4. Quantitative results

4.4.1. Quantitative results on CRCHistoPhenotypes dataset

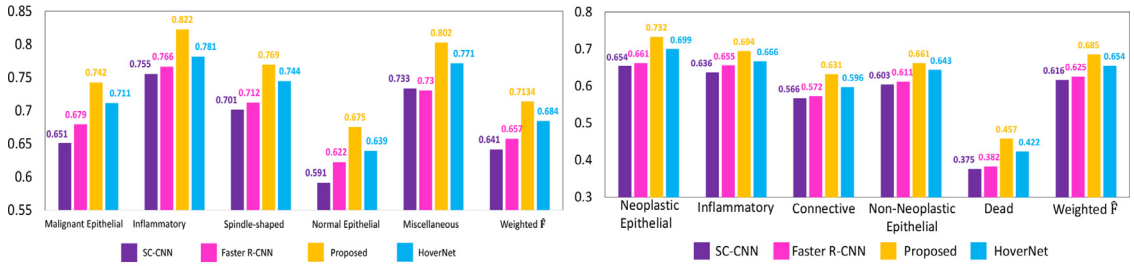
(i): Experiments with Ground-truth Nuclei: In this experiment, the classification performance of the proposed NCCD-GT algorithm is evaluated using ground truth nuclei detection and compared with SC-CNN and Faster R-CNN for nuclei classification using the classification heads of these methods. For fair comparison, the ground-truth nuclei detection are employed with both SC-CNN and Faster R-CNN. On the average, the weighted $\widehat{F}$ score of 82.20% is obtained which is 3.80% better than SC-CNN and 5.70% better than Faster R-CNN as shown in Fig. 3. The F score of individual nuclei classes is also compared and we observed consistently good performance of the proposed NCCD algorithm over the compared methods. Note that some existing state-of-the-art methods such as DIST and HoverNet can not be employed on CRCHistophenotypes dataset due to unavailability of nuclei mask.

(ii): Experiments with NCCD Algorithm: In this experiment, NCCD algorithm is integrated with the existing methods as a classification refinement module. We observed performance improvement for HoverNet and SC-CNN methods in this setting as shown in Table 1. For the case of HoverNet, we observed performance improvement of 2.60% with the addition of NCCD module. For the case of SC-CNN, we also observed the performance improvement of 2.70% using our proposed NCCD algorithm. In these experiments, the initial classification labels are used for the respective methods. Some of the existing methods such as SCC-DCF and TSP-CNN are only proposed for nuclei detection task in histology images. For these methods, initial classification labels are obtained by using a two layer fully connected module for classification which is then improved using NCCD algorithm. The best classification performance of 80.90% weighted $\widehat{F}$ is observed for SCC-DCC + Prop which is 14.20% larger than DIST demonstrating the significance of the proposed NCCD algorithm for nuclei classification task.

4.4.2. quantitative results on CoNSeP dataset

(i): Experiments with Ground-truth Nuclei: The performance of the proposed NCCD-GT algorithm is evaluated using ground truth nuclei segmentation and compared with SC-CNN, Faster R-CNN, and HoverNet for nuclei classification using the classification heads of compared methods. For SC-CNN and Faster R-CNN, we use the centroid of the nuclei segmentation as a ground-truth. In this experiment, the weighted $\widehat{F}$ score of 71.34% which is up to 7.20% better than the existing methods as shown in Fig. 4(a). The proposed NCCD algorithm has also obtained consistently good performance compared to the current state-of-the-art on each nuclei class.

(ii): Experiments with NCCD Algorithm: Table 2 shows the performance improvement of the current state-of-the-art methods when integrated with the proposed NCCD algorithm. The nuclei detec-



(a) CoNSEP Dataset (Graham et al. (2019)) (b) PanNuke Dataset (Gamper et al. (2020))

Fig. 4. Classification performance evaluation of different methods using ground-truth nuclei annotations using individual nucleus class F_1 score and weighted average $\hat{F}$ score.

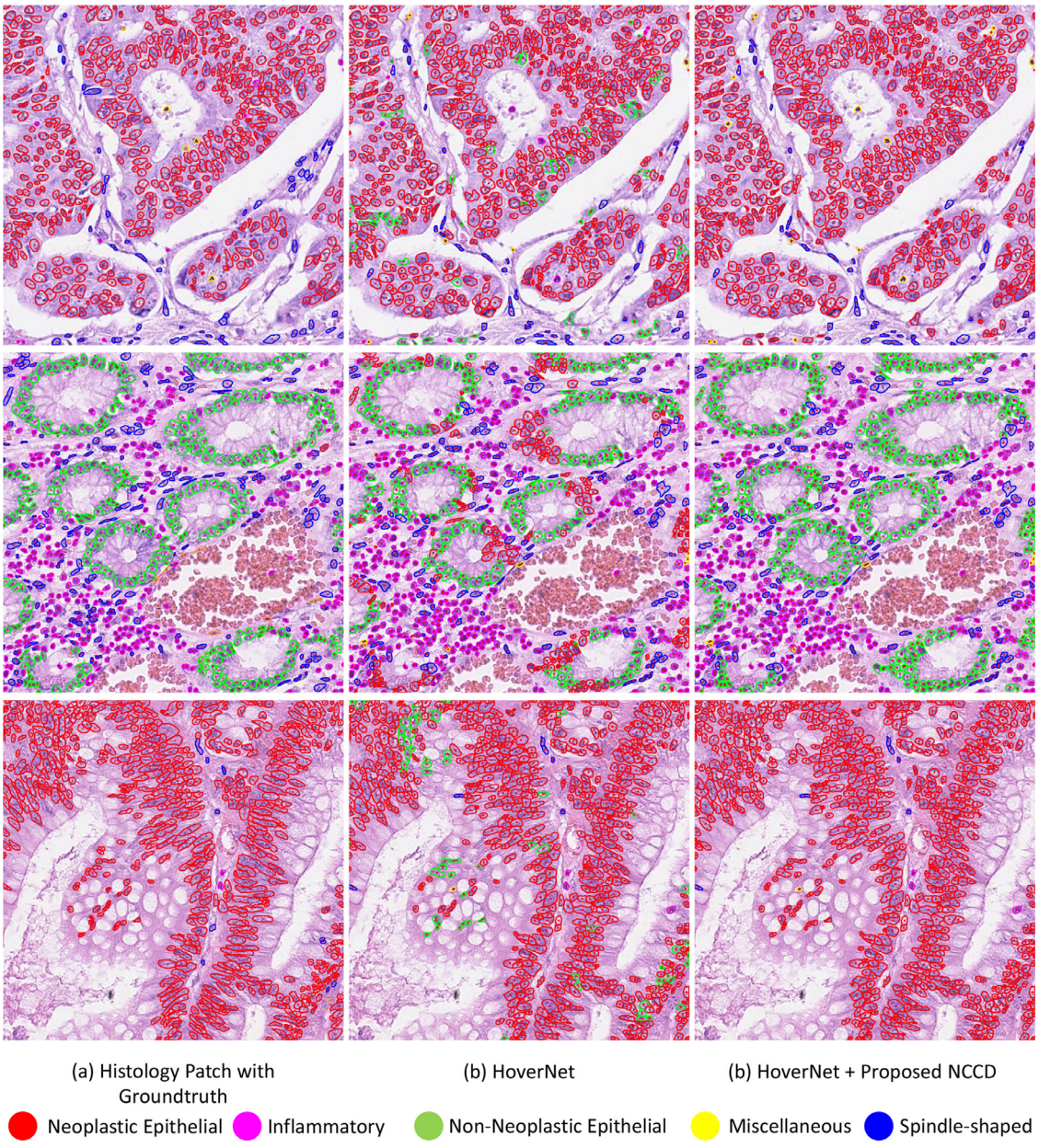


Fig. 5. Qualitative comparison of the proposed NCCD algorithm with HoverNet method on three sample histology images selected from CoNSeP dataset (Graham et al., 2019).

Table 1
Comparative performance of nucleus classification on CRCHistoPhenotypes dataset using NCCD algorithm as a component. F_d denotes the F_1 score of nuclei detection while F_c^e, F_c^i, F_c^f, and F_c^m represent the classification F_1 scores of epithelial, inflammatory, fibroblast, and miscellaneous nuclei classes. $\widehat{F}$ denotes the weighted average classification F score.Imp. denotes the improvement obtained by the proposed NCCD algorithm over the existing baseline method. The first and second best performing methods are shown in the bold and italic fonts, respectively.

Methods	F_d	F_c^e	F_c^i	F_c^f	F_c^m	$\widehat{F}$	Imp.
SC-CNN (Sirinukunwattana et al. (2016))	0.802	0.683	0.786	0.633	0.524	0.687	-
SC-CNN+Prop	0.802	0.714	0.819	0.678	0.545	0.714	2.7%
HoverNet (Graham et al. (2019))	0.841	0.751	0.862	0.711	0.621	0.763	-
HoverNet+Prop	0.841	0.772	0.886	0.742	0.662	0.789	2.6%
FasterR-CNN (Ren et al. (2015))	0.771	0.651	0.775	0.632	0.491	0.670	-
FasterR-CNN+Prop	0.771	0.669	0.791	0.654	0.516	0.692	2.2%
MaskR-CNN (He et al. (2017))	0.788	0.661	0.781	0.644	0.501	0.676	-
MaskR-CNN + Prop	0.788	0.682	0.802	0.661	0.522	0.701	2.5%
DIST (Naylor et al. (2018))	0.742	0.661	0.771	0.621	0.471	0.667	-
TSP-CNN+Prop	0.852	0.778	0.891	0.752	0.668	**0.796**	-
SCC-DCF+Prop	0.914	0.786	0.905	0.771	0.681	*0.809*	-

Table 2
Comparative performance of nucleus classification on CoNSeP dataset NCCD algorithm as a component. F_d denotes the F_1 score of nuclei detection while F_c^e, F_c^i, F_c^s, and F_c^m represent the F_1 classification scores of epithelial, inflammatory, spindle-shaped, and miscellaneous nuclei classes. $\widehat{F}$ denotes the weighted average classification F score. Imp. denotes the improvement obtained by the proposed NCCD algorithm over the existing baseline method. The first and second best performing methods are shown in the bold and italic fonts, respectively.

Methods	PQ	F_d	F_c^e	F_c^i	F_c^s	F_c^m	$\widehat{F}$	Imp.
SC-CNN	-	0.737	0.591	0.501	0.556	0.511	0.556	-
SC-CNN+Prop	-	0.737	0.622	0.546	0.583	0.538	0.582	2.6%
Hover-Net	0.516	0.753	0.635	0.631	0.566	0.626	0.619	-
Hover-Net+Prop	0.516	0.753	0.673	0.663	0.588	0.650	0.651	3.2%
MaskR-CNN	0.450	0.713	0.595	0.590	0.520	0.511	0.573	-
MaskR-CNN +Prop	0.450	0.721	0.602	0.533	0.520	0.519	**0.584**	-
DIST	0.372	0.728	0.617	0.534	0.557	0.522	0.577	-
Micro-Net (Raza et al. (2019))	0.430	0.747	0.615	0.592	0.532	0.481	0.582	-
SCC-DCF+Prop	-	0.848	0.692	0.687	0.595	0.666	*0.670*	-

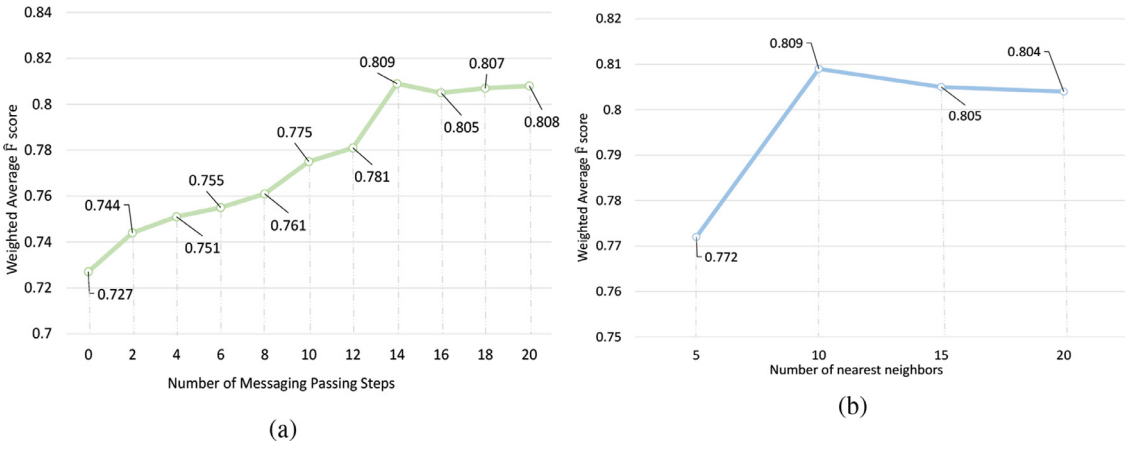


Fig. 6. Nucleus classification performance of the proposed NCCD algorithm on CRCHistophenotypes dataset (Sirinukunwattana et al., 2016) by varying number of message passing steps (t_l) and a number of nearest neighbors (k).

tion performance shown as F_d measure remains same because the proposed NCCD algorithm only improves the classification performance. However, a higher F_d will result in higher classification scores accross different state-of-the-art methods. We observe classification performance improvement of 2.60% and 3.20% in SC-CNN and HoverNet methods. The best performance of 67.0% is obtained by using the SCC-DCF nuclei detection method which is up to 11.40% than existing methods. For this case, the initial classification label is performed by using our two layer fully connected network. The classification performance of this simple network is 60.60% which is 6.40% improved by the use of proposed NCCD algorithm.

4.4.3. quantitative results on PanNuke dataset

(i): Experiments with ground-truth Nuclei: The performance of the proposed NCCD-GT algorithm is evaluated using ground truth nuclei segmentation and compared with SC-CNN, Faster R-CNN, and HoverNet for nuclei classification. In this experiment, the weighted $\widehat{F}$ score of 68.50% obtained by the proposed algorithm which is upto 6.90% better than the existing methods as shown in Fig. 4(b). The proposed NCCD algorithm has also obtained consistently good performance compared to the current state-of-the-art on each nuclei class.

(ii): Experiments with NCCD Algorithm: Table 3 shows the comparison of the proposed NCCD algorithm for nuclei classification

Table 3
Comparative performance of nucleus classification on PanNuke dataset NCCD algorithm as a component. F_d denotes the F_1 score of nuclei detection while F_c^n, F_c^{nn}, F_c^i, F_c^c, and F_c^d represent the classification F_1 scores of neoplastic epithelial, non-neoplastic epithelial, inflammatory, connective, and dead nuclei classes. $\widehat{F}$ denotes the weighted average classification F score accross all nuclei classes. The first and second best performing methods are shown in the bold and italic fonts, respectively.

Methods	F_d	F_c^n	F_c^{nn}	F_c^i	F_c^c	F_c^d	$\widehat{F}$
Det U-Net (Xie et al. (2018))	0.650	0.430	0.290	0.371	0.360	0.00	0.376
DIST	0.730	0.500	0.350	0.420	0.391	0.00	0.436
MaskR-CNN	0.720	0.590	0.520	0.500	0.420	0.220	0.515
MaskR-CNN+Prop	0.720	0.601	0.535	0.512	0.434	0.231	0.526
MicroNet	0.800	0.620	0.580	0.520	0.470	0.190	0.552
HoverNet	0.800	0.620	0.560	0.540	0.490	0.310	**0.560**
HoverNet+Prop	0.800	0.660	0.588	0.571	0.525	0.354	*0.595*

Table 4
Comparative performance of nucleus classification on Lizard dataset NCCD algorithm as a component. F_d denotes the F_1 score of nuclei detection while F_c^{ep}, F_c^i, F_c^p, F_c^n, F_c^e, F_c^c, and F_c^d represent the classification F_1 scores of epithelial, inflammatory, plasma, neutrophil, eosinophil, connective nuclei classes. $\widehat{F}$ denotes the weighted average F score accross all nuclei classes. Imp. denotes the improvement obtained by the proposed NCCD algorithm over the existing baseline method. The first and second best performing methods are shown in the bold and italic fonts, respectively.

Methods	F_d	F_c^{ep}	F_c^i	F_c^p	F_c^n	F_c^e	F_c^c	$\widehat{F}$	Imp.
SC-CNN	0.511	0.322	0.305	0.342	0.256	0.352	0.423	0.341	-
SC-CNN + Prop	0.511	0.344	0.339	0.372	0.285	0.371	0.456	0.369	2.8%
HoverNet	0.633	0.402	0.371	0.436	0.346	0.427	0.503	**0.419**	-
HoverNet + Prop	0.633	0.423	0.404	0.471	0.378	0.461	0.534	*0.446*	2.7%

with existing methods including Det U-NET, DIST, Mask R-CNN, MicroNet, and HoverNet. The proposed NCCD algorithm is used as a component with HoverNet resulting in 59.50% performance which is up to 21.90% better than the existing methods. The other compared methods have obtained lower performance than the HoverNet therefore experiments are not repeated with the other methods.

4.4.4. quantitative results on Lizard dataset

On this dataset, the proposed NCCD algorithm when integrated with SC-CNN and HoverNet methods resulted in 36.90% and 44.60% weighted average $\widehat{F}$ score as shown in Table 4. Compared to SC-CNN, the proposed algorithm outperforms by 2.80% while 2.70% compared to HoverNet. Overall, the proposed algorithm (HoverNet + Prop) obtained up to 10.50 % better performance than existing methods. This experiment demonstrates the effectiveness of our proposed NCCD algorithm for classification label improvement. Lizard dataset being a very recent one and have not been evaluated by many researchers. In our work, we present the classification performance comparison while the results reported in Graham et al. (2021) emphasize on the nuclei segmentation accuracy. The classification performance on this dataset has not been discussed before in detail.

4.5. Visual comparisons

Three different colorectal cancer histology patches are selected from CoNSeP dataset for qualitative comparison as shown in Fig. 5. These patches contain majority of cluterred and clumped nuclei of malignant epithelial and normal epithelial cases. In the top and bottom rows, HoverNet has miss-classified many neoplastic epithelial nuclei which are refined by our proposed NCCD algorithm as neoplastic nuclei. In the middle row, the HoverNet has performed significant mistakes while identifying non-neoplastic epithelial. Most of the wrong predictions made by the HoverNet are corrected by our proposed NCCD algorithm which can be observed in the glandular structure containing clumped of non-neoplastic epithelial nuclei.

4.5.1. Ablation studies

In this section, we evaluate different components of the proposed algorithm as shown in Table 5. First, we remove the message passing component and directly apply edge classification on the nearest neighbor graph (NN Graph). For this case, we obtained 72.70% weighted average $\widehat{F}$ score on CRCHistophenotypes and 59.80% on CoNSeP datasets. In the next step, we replace the message passing network scheme with a vanilla version consisting of only one node-based update as given by Eq. 4. We observed a performance deduction of 3.80% and 3.90% on CRCHistophenotypes and CoNSeP compared to the proposed Message Passing Network(MPN) algorithm. This comparison shows the significance of the proposed message passing scheme because both schemes have resulted in better performance than using no message passing.

In addition to that, the effect of the number of message passing steps on the performance is also analysed by varying the iterations from 0 to 20 as shown in Fig. 6(a). As the number of message passing steps increased from 0 to 14 the performance increases from 72.70% to 80.90%. However, further increase in the number of message passing steps could not produce better results. As the number of message passing steps are increased, the contextual information flows from a node to all of its neighbors. For 14 message passing steps, the contextual information has diffused upto 14th hops.

The performance variation by varying the number of nearest neighbors in the graph using proposed MPN scheme is also analysed as shown in Fig. 6(b). The number of nearest neighbor $k = \{5, 10, 15, 20\}$ are tested. As the number of nearest neighbor increases from 5 to 10 the performance trend increases from 77.20% to 80.90%. However, further increase in the nearest neighbor to $k = 15$ and $k = 20$ have caused a slight decrease in performance on CRCHistophenotypes dataset.

The comparisons have also been performed with the classical community detection method of Louvian (Blondel et al., 2008). These experiments have been performed in two different settings on the PanNuke dataset. In the first setting, nearest neighbor graph is computed using geometric position of each nucleus and Louvian method is directly employed on this graph resulting in $\widehat{F}$ score of 54.1%. In the second setting, proposed ResNet50 + Positional encoding features are employed for the construction of graph and

Table 5
Ablation studies of different variants of the proposed algorithm.

Datasets	NN Graph	Vanilla MPN	Proposed MPN	Edge Classification	$\widehat{F}$
CRCHistoPhenotypes	✓		✓	✓	**0.809**
	✓	✓		✓	*0.771*
	✓			✓	0.727
CoNSeP	✓		✓	✓	0.670
	✓	✓		✓	0.631
	✓			✓	0.598

Table 6
Comparison of NCCD algorithm with classical community detection method Blondel et al. (2008) using PanNuke dataset with HoverNet method as baseline. NN Graph represents the Nearest Neighbor graph.

Datasets	NN Graph	Features (ResNet50+Positional Encoding)	NCCD (HoverNet+Prop)	Louvian Method (Blondel et al. (2008))	$\widehat{F}$
PanNuke Dataset (Gamper et al., 2020)	✓	✓	✓		**0.595**
	✓			✓	0.541
		✓		✓	*0.565*

then Louvian method is used for community detection. This setting resulted in $\widehat{F}$ score of 56.5%. The improved performance shows the importance of the proposed features. Both experiments have resulted lower performance than the proposed NCCD algorithm which achieved 59.5% $\widehat{F}$. These experiments demonstrate the significance of the proposed message passing network-based community detection algorithm.

4.5.2. Discussion on experimental results

Experiments are performed using two different settings, using ground-truth nuclei annotations to compare the classification heads of different methods and using the proposed NCCD algorithm as a component with the existing methods. In the first setting, we observed better classification performance of the proposed algorithm on three datasets while in the second setting the integration of NCCD with existing methods have improved their performance in almost all cases. The improved performance can be attributed towards correction made in incorrect labeling by the existing methods. As we discussed in the visual comparisons section, the communities detected by the NCCD algorithm are biologically meaningful and full community is assigned a single label. Such community detection has not been explored before and it inherits its performance from the proposed underlying message passing network. It is in contrast to the existing methods which independently assign a label to each nuclei. Compared to the existing approaches, our algorithm is more robust because the label assignment is based on the connectivity among distinct nuclear components. Those components which are strongly connected with each other are assigned the same label resulting in the reduction of labeling errors.

Although, the community based labeling offers better labels however, it also has limitations. For instance, if the detected community is isolated consisting of very few nuclei the most frequent label in that community could be erroneous resulting in incorrect label for the full community. Similarly, if the baseline methods used for the initial nuclei labeling are significantly incorrect may result in incorrect label for the full community. It will result in performance degradation of the proposed algorithm. Thus, the proposed algorithm will be able to refine labels as long as the most frequent label in each community is a correct label.

4.6. Computational time

We report the testing time of our proposed NCCD algorithm on the CoNSeP and PanNuke datasets. The testing time for the CoNSeP dataset is almost 5 minutes using two Titan Xp GPUs for the whole test split consisting of 14 histology patches of CRC. For the PanNuke dataset, the total time on testing split is almost 15 minutes on three folds. Compared to other methods, the testing time is not significantly large showing the practical significance of the proposed NCCD algorithm.

5. Conclusion

In this work, a nucleus classification algorithm is proposed by detecting biologically meaningful nuclear communities in histology images. For this purpose, a graph-based message passing algorithm is proposed for finding such communities which are then used for nucleus classification label refinement task. In wide range of experiments, the proposed algorithm is compared with existing state-of-the-art methods. The experimental results using ground-truth annotations demonstrate excellent classification performance of the proposed algorithm. In second set of experiments, the proposed algorithm is used as a component with the existing state-of-the-art methods. By using this integration, significant performance improvement is observed in almost all cases. These results demonstrate the significance of the proposed algorithm for nuclei classification. Interesting future directions including formation of cancer bio-markers, improved tumor segmentation via segmenting tumor epithelial communities from WSIs, improvements using vision transformers, and survival analysis based on tumor-infiltrating lymphocytes indicators can be explored to advance the machine learning applications for computational pathology in clinical practices.

Declaration of Competing Interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgements

This publication acknowledges the support provided by the Khalifa University of Science and Technology under Award No. RC1-2018-KUCARS. The second author of this work, Sajid Javed, is supported by the Khalifa University of Science and Technology under the Faculty Start Up grants FSU-2022-003 Award No. 8474000401. Naoufel Werghi is supported by a research grant from ASPIRE Institute, UAE, Ref: AARE20-279. The last author of this work, Nasir Rajpoot (NR), is supported by the UK Medical Research Council (No. MR/P015476/1). NR is also supported by the PathLAKE digital

pathology consortium, which is funded from the Data to Early Diagnosis and Precision Medicine strand of the governments Industrial Strategy Challenge Fund, managed and delivered by UK Research and Innovation (UKRI).

References

Abbe, E., 2017. Community detection and stochastic block models: recent developments. The Journal of Machine Learning Research 18 (1), 6446–6531.

Abousamra, S., Belinsky, D., Van Arnam, J., Allard, F., Yee, E., Gupta, R., Kurc, T., Samaras, D., Saltz, J., Chen, C., 2021. Multi-class cell detection using spatial context representation. IEEE ICCV.

Alberts, B., Bray, D., Hopkin, K., Johnson, A.D., Lewis, J., Raff, M., Roberts, K., Walter, P., 2015. Essential cell biology. Garland Science.

Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al., 2018. Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.

Battaglia, P. W., Pascanu, R., Lai, M., Rezende, D., Kavukcuoglu, K.,. Interaction networks for learning about objects, relations and physics. In: NIPS.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. 2008 (10), P10008.

Brasó, G., Leal-Taixé, L., 2020. Learning a neural solver for multiple object tracking. IEEE CVPR.

Bui, M.M., Asa, S.L., Pantanowitz, L., Parwani, A., van der Laak, J., Ung, C., Balis, U., Isaacs, M., Glassy, E., Manning, L., 2019. Digital and computational pathology: bring the future into focus. J. of Path. Info. 10.

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Silva, V.W.K., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. 25 (8), 1301–1309.

Cosatto, E., Miller, M., Graf, H.P., Meyer, J.S., 2008. Grading nuclear pleomorphism on histological micrographs. IEEE ICPR.

Cruz-Roa, A.A., Ovalle, J.E.A., Madabhushi, A., Osorio, F.A.G., 2013. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. MICCAI.

Dalle, J.-R., Li, H., Huang, C.-H., Leow, W.K., Racoceanu, D., Putti, T.C., 2009. Nuclear pleomorphism scoring by selective cell nuclei detection. WACV.

Demir, C., Yener, B., 2004. Automated cancer diagnosis based on histopathological images: a systematic survey.

Fuchs, T.J., Buhmann, J.M., 2011. Computational pathology: challenges and promises for tissue analysis. CMIAG 35 (7–8), 515–530.

Gamper, J., Koohbanani, N.A., Graham, S., Jahanifar, M., Khurram, S.A., Azam, A., Hewitt, K., Rajpoot, N., 2020. Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778.

Gao, Z., Wang, L., Zhou, L., Zhang, J., 2016. Hep-2 cell image classification with deep convolutional neural networks. IEEE JBMHI 21 (2), 416–428.

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E., 2017. Neural message passing for quantum chemistry. ICML.

Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.-W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al., 2021. Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification. IEEE ICCV-W.

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

He, C., Fei, X., Cheng, Q., Li, H., Hu, Z., Tang, Y., 2021. A Survey of Community Detection in Complex Networks Using Nonnegative Matrix Factorization. IEEE Trans. Comput. Social Syst..

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. IEEE CVPR.

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

Hu, B., Tang, Y., Eric, I., Chang, C., Fan, Y., Lai, M., Xu, Y., 2018. Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. IEEE JBMHI 23 (3), 1316–1328.

Irshad, H., Veillard, A., Roux, L., Racoceanu, D., 2013. Methods for nuclei detection, segmentation, and classification in digital histopathology: a reviewcurrent status and future potential. IEEE Rev. in Bio. Eng. 7, 97–114.

Javed, S., Fraz, M.M., Epstein, D., Snead, D., Rajpoot, N.M., 2018. Cellular Community Detection for Tissue Phenotyping in Histology Images. In: MICCAI-W.

Javed, S., Mahmood, A., Dias, J., Werghi, N., Rajpoot, N., 2021. Spatially constrained context-aware hierarchical deep correlation filters for nucleus detection in histology images. MIA 72, 102104.

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

Javed, S., Mahmood, A., Werghi, N., Benes, K., Rajpoot, N., 2020. Multiplex cellular communities in multi-gigapixel colorectal cancer histology images for tissue phenotyping. IEEE T-IP 29, 9204–9219.

Javed, S., Mahmood, A., Werghi, N., Rajpoot, N., 2019. Deep multiresolution cellular communities for semantic segmentation of multi-gigapixel histology images. ICCV-W.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. ECCV. Springer.

Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.-A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al., 2019. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med. 16 (1).

Kipf, T., Fetaya, E., Wang, K.-C., Welling, M., Zemel, R., 2018. Neural relational inference for interacting systems. ICML.

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

Liu, F., Xue, S., Wu, J., Zhou, C., Hu, W., Paris, C., Nepal, S., Yang, J., Yu, P.S., 2020. Deep learning for community detection: progress, challenges and opportunities. arXiv preprint arXiv:2005.08225.

Mahmood, A., Small, M., 2015. Subspace based network community detection using sparse linear coding. IEEE Trans Knowl Data Eng 28 (3), 801–812.

Mahmood, A., Small, M., Al-Maadeed, S.A., Rajpoot, N., 2016. Using geodesic space density gradients for network community detection. IEEE Trans Knowl Data Eng 29 (4), 921–935.

Malon, C.D., Cosatto, E., 2013. Classification of mitotic figures with convolutional neural networks and seeded blob features. JPI 4.

Naylor, P., Laé, M., Reyal, F., Walter, T., 2018. Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE T-MI 38 (2), 448–459.

Nguyen, K., Jain, A.K., Sabata, B., 2011. Prostate cancer detection: fusion of cytological and textural features. JPI 2.

Oliveira, S.P., Neto, P.C., Fraga, J., Montezuma, D., Monteiro, A., Monteiro, J., Ribeiro, L., Gonçalves, S., Pinto, I.M., Cardoso, J.S., 2021. Cad systems for colorectal cancer from wsi are still not ready for clinical acceptance. Sci. Reps. 11 (1), 1–15.

Pinckaers, H., van Ginneken, B., Litjens, G., 2020. Streaming convolutional neural networks for end-to-end learning with multi-megapixel images. IEEE T-PAMI.

Qu, H., Riedlinger, G., Wu, P., Huang, Q., Yi, J., De, S., Metaxas, D., 2019. Joint segmentation and fine-grained classification of nuclei in histopathology images. IEEE ISBI.

Raza, S.E.A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., Khan, M., Rajpoot, N.M., 2019. Micro-net: a unified model for segmentation of various objects in microscopy images. MIA 52, 160–173.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. A-NIPS.

Sharma, H., Zerbe, N., Heim, D., Wienert, S., Behrens, H.-M., Hellwich, O., Hufnagl, P., 2015. A multi-resolution approach for combining visual information using nuclei segmentation and classification in histopathological images. In: VISAPP (3), pp. 37–46.

Singh, M., Zeng, Z., Kalaw, E.M., Giron, D.M., Chong, K.-T., Lee, H.K., 2017. A study of nuclei classification methods in histopathological images. IMH.

Sirinukunwattana, K., e Ahmed Raza, S., Tsang, Y.-W., Snead, D.R., Cree, I.A., Rajpoot, N.M., 2016. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE T-MI 35 (5), 1196–1206.

Song, Z., Zou, S., Zhou, W., Huang, Y., Shao, L., Yuan, J., Gou, X., Jin, W., Wang, Z., Chen, X., et al., 2020. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. Nat. Comm. 11 (1), 1–9.

Srinidhi, C.L., Ciga, O., Martel, A.L., 2021. Deep neural network models for computational histopathology: a survey. Med Image Anal 67, 101813.

Sun, Y., Huang, X., Zhou, H., Zhang, Q., 2021. Srpn: similarity-based region proposal networks for nuclei and cells detection in histology images. MIA 102142.

Tizhoosh, H.R., Pantanowitz, L., 2018. Artificial intelligence and digital pathology: challenges and opportunities. JPI 9.

Tofighi, M., Guo, T., Vanamala, J.K., Monga, V., 2019. Prior information guided regularized deep learning for cell nucleus detection. IEEE T-MI 38 (9), 2047–2058.

Tomczak, K., Czerwińska, P., Wiznerowicz, M., 2015. The cancer genome atlas (tcga): an immeasurable source of knowledge. Contemp Oncol 19 (1A), A68.

Tripathi, S., Singh, S.K., 2020. Cell nuclei classification in histopathological images using hybrid olconvnet. ACM-TOMM) 16, 1–22. 1s

Van Lierde, H., Chow, T.W., Chen, G., 2019. Scalable spectral clustering for overlapping community detection in large-scale networks. IEEE Trans Knowl Data Eng 32 (4), 754–767.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Statistics and computing 17 (4), 395–416.

Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., Madabhushi, A., 2014. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. MIDP, Vol. 9041.

Xie, Y., Xing, F., Shi, X., Kong, X., Su, H., Yang, L., 2018. Efficient and robust cell detection: a structured regression approach. MIA 44, 245–254.

Xing, F., Cornish, T.C., Bennett, T., Ghosh, D., Yang, L., 2019. Pixel-to-pixel learning with weak supervision for single-stage nucleus recognition in ki67 images. IEEE TBME 66 (11), 3088–3097.

Xing, F., Yang, L., 2016. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE Rev. in Bio. Eng. 9, 234–263.

Yuan, Y., Failmezger, H., Rueda, O.M., Ali, H.R., Gräf, S., Chin, S.-F., Schwarz, R.F., Curtis, C., Dunning, M.J., Bardwell, H., et al., 2012. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. STM 4 (157). 157ra143–157ra143