

WEAKLY SUPERVISED PROSTATE TMA CLASSIFICATION VIA GRAPH CONVOLUTIONAL NETWORKS

Jingwen Wang¹, Richard J. Chen¹, Ming Y. Lu¹, Alexander Baras², Faisal Mahmood¹

¹Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

²Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD

jwang111@bwh.harvard.edu faisalmahmood@bwh.harvard.edu

ABSTRACT

Histology-based grade classification is clinically important for many cancer types in stratifying patients into distinct treatment groups. In prostate cancer, the Gleason score is a grading system used to measure the aggressiveness of prostate cancer from the spatial organization of cells and the distribution of glands. However, the subjective interpretation of Gleason score often suffers from large interobserver and intraobserver variability. Previous work in deep learning-based objective Gleason grading requires manual pixel-level annotation. In this work, we propose a weakly-supervised approach for grade classification in tissue micro-arrays (TMA) using graph convolutional networks (GCNs), in which we model the spatial organization of cells as a graph to better capture the proliferation and community structure of tumor cells. We learn the morphometry of each cell using a contrastive predictive coding (CPC)-based self-supervised approach. Using five-fold cross-validation we demonstrate that our method can achieve a 0.9637 ± 0.0131 AUC using only TMA-level labels. Our method also demonstrates a 36.36% improvement in AUC over standard GCNs with texture features and a 15.48% improvement over GCNs with VGG19 features. Our proposed pipeline can be used to objectively stratify low and high-risk cases, reducing inter- and intra-observer variability and pathologist workload.

Index Terms— Gleason Score Grading, Graph Convolutional Networks, Deep Learning, Histopathology Classification, Objective Grading, Patient Stratification

1. INTRODUCTION

The subjective interpretation of histology slides is the standard-of-care for prostate cancer detection and prognostication. The commonly used Gleason score, which informs the aggressiveness of prostate cancer, is based on the architectural pattern of tumor tissues and the distribution of glands. Gleason grade ranges from 1 to 5 to suggest how much the prostate tissue looks like normal tissue (lower score) versus abnormal tissue (higher score). If the glands are small, discrete and uniform, the Gleason grade is relatively lower. On the contrary, if

the glands are fused or poorly-formed, the Gleason grade is higher. A primary and a secondary grade which range from 1 to 5 are assigned to prostate tissue and the sum of the two grades determines the final Gleason score. However, its manual assignment is based on visual, microscopy-based evaluation of cellular and morphological patterns, which can be extremely error-prone and time-consuming for pathologists and suffers from interobserver and intraobserver variability [1]. Deep learning has been widely applied to the detection of malignancies in histology images [2, 3, 4]. It has been demonstrated that convolutional neural networks (CNNs) can serve as a tool for automated histology image classification [5, 6, 7], including Gleason score assignment, which alleviates the aforementioned limitations of subjective interpretation. However, CNNs are not sufficiently context-aware and do not capture the dense spatial relationships between cells that are predictive of cancer grade and proliferation. In addition, CNN-based methods require detailed pixel or patch level annotations which are tedious and time consuming to curate [8].

On the contrary, graphs are a reasonable and natural option to model the distribution of the cells in histopathology images. With the cells as nodes and the edges generated by a proper algorithm, graphs can accurately capture the distribution of cells and the spatial relations between the cells. GCN is a deep learning approach for performing feature extraction and classification on graphs[9]. In this paper, we propose a GCN based-approach for automatic patient stratification trained using TMA-level labels and node features learned via a self-supervised technique known as contrastive predictive coding (CPC) [10, 11]. GCNs can learn from the global distribution of cell nuclei, cell morphometry and spatial features without requiring exhaustive pixel-level annotation.

2. RELATED WORK

Objective Gleason score grading: Previous work on computer-aided Gleason score grading has used machine learning with pixel-level annotations. Khurd *et al.* [12] developed a texture-based classification system for automatic and reproducible Gleason grading. They use random forests to cluster extracted filter responses at each pixel into ba-

sic texture elements (textons) and characterize the texture in images belonging to a certain tumor grade. Del Toro *et al.* [13] stratify the patients into high-risk and low-risk with the decision boundary of Gleason score 7-8. They train a CNN that requires pixel-level labels on regions of interests (ROI) extracted from WSIs and are able to detect prostatectomy WSIs with high-grade Gleason score. Arvaniti *et al.* [8] present a deep learning approach for automated Gleason grading of prostate cancer TMAs and are able to assign both the primary score and secondary score to a single pixel in the patients' TMAs. However, all current methods require pixel or patch level labels and do not incorporate the spatial organization of cells to fully capture the architecture of the tumor micro-environment.

Classification using GCNs: GCNs proposed by Kipf *et al.* [9] has been utilized in many classification tasks. Recently, Zhou *et al.* [14] propose using GCN for colorectal cancer classification, where each node is represented by a nucleus within the original image and cellular interactions are denoted as edges between these nodes according to node similarity. Then they extract spatial features like the centroid coordinates and texture features like angular second moment (ASM) obtained from grey level co-occurrence matrix (GLCM) [15, 16] for each node. GCNs are utilized to classify each image into normal, low-grade and high-grade based on the degree of gland differentiation. Anand *et al.* [17] also propose to classify cancers using GCNs by modeling a tissue section as a multi-attributed spatial graph of its constituent cells. They not only extract the spatial and texture features but also use a pre-trained VGG19 model to generate features from a window of size 71×71 around the nuclei centroids. Then, a GCN is trained on the breast cancer BACH dataset [18] to classify patients into cancerous or non-cancerous groups. However, these methods rely on gray-scale on ImageNet features and do not use unsupervised features.

3. METHODS

Our method is inspired by the process through which pathologists examine prostate tissue and assign Gleason scores. Pathologists inspect the spatial distributions of the glands in prostate histology slides. In order to model the distribution of the nuclei, we segment out the nuclei, and then construct graphs for TMAs with nuclei as the nodes and the distance between neighbor nuclei as the edges.

3.1. TMA Graph Construction

In order to capture the information stored in prostate TMAs, the images first need to be stain-normalized to remove the color variance. Then, each image is converted to a cell graph where the nuclei are the nodes and the possible interactions are the edges. We constructed a graph for each TMA using the following steps: i) segmenting the nuclei, which form the set of nodes. ii) modeling interactions between nuclei using edges. iii) generating features for each node.

Nuclei Segmentation: To detect atypical and packed nuclei and other features that suggest cancerous tissues, precise segmentation results are fundamental. Fully convolutional neural networks (FCNN) have been used for nuclei segmentation in the past, which aims to minimize pixel-wise loss [19]. However, this can sometimes lead to the fusion of nuclei boundaries and these blurry outlines of nuclei can cause feature extraction and graph structure construction to be inaccurate. In order to detect the nuclei robustly and clearly, we utilize conditional GANs (cGANs) [20]. Such methods, which use an proper loss function for semantic segmentation, have been proved to avoid the aforementioned issues. We train our model for segmentation on the Multi-Organ Nuclei Segmentation dataset [21]. This dataset, which consists of 21,623 nuclei in 30 images, comes from 18 different hospitals and include diverse nuclear appearances from a variety of organs like liver, prostate, bladder, etc.

Nuclei Connection: We assume that nuclei that are close to each other are more likely to have interactions and we intend to capture the architectural structure between neighboring nuclei. Based on this idea, we adopt the K -nearest neighbors (KNN) algorithm [22], in which each nucleus is connected to its top 5 nearest neighboring nucleus if they are within a certain Euclidean distance (100 pixels in our method).

Nuclei Feature Extraction: Based on the nuclei masks obtained using a cGAN, we then extract both morphological and texture features for each nucleus, along with features extracted from CPC-based self-supervised learning.

Morphological Features: For morphological features, we compute area, roundness, eccentricity, convexity, orientation, etc for each of the nucleus.

Texture Features: To analyze the nuclei texture in the TMAs, we calculate GLCM for each nucleus and then compute second order features like dissimilarity, homogeneity, energy and ASM based on the obtained GLCM.

Contrastive Predictive Coding Features: Contrastive predictive coding is an unsupervised learning approach to extract useful representations from high-dimensional data. By predicting the future in the latent space, CPC [11] is able to learn such representations through autoregressive models. With a certain data sequence $\{x_t\}$, a feature network g_{enc} first calculates a low-dimensional embedding z_t for each observation. Then, an auto-regressive context network g_c can accumulate observations prior to t , namely, $c_t = g_c(\{z_i\})$, for $i \leq t$. CPC utilizes a contrastive loss, through which the mutual information shared between the context c_t , the present, and future observations z_{t+k} , $k > 0$ can be maximized. The network aims to correctly distinguish the positive target z_{t+k} from sampled negative candidates. Predictions for z_{t+k} can be achieved linearly using weights W_k : $\hat{z}_{t+k} = W_k c_t$. If the probability score for each candidate z_i is assigned through a log-bilinear model, the CPC objective is just the standard cross-entropy loss for the positive target. When applying CPC to images, small, overlapping patches need to be extracted from each

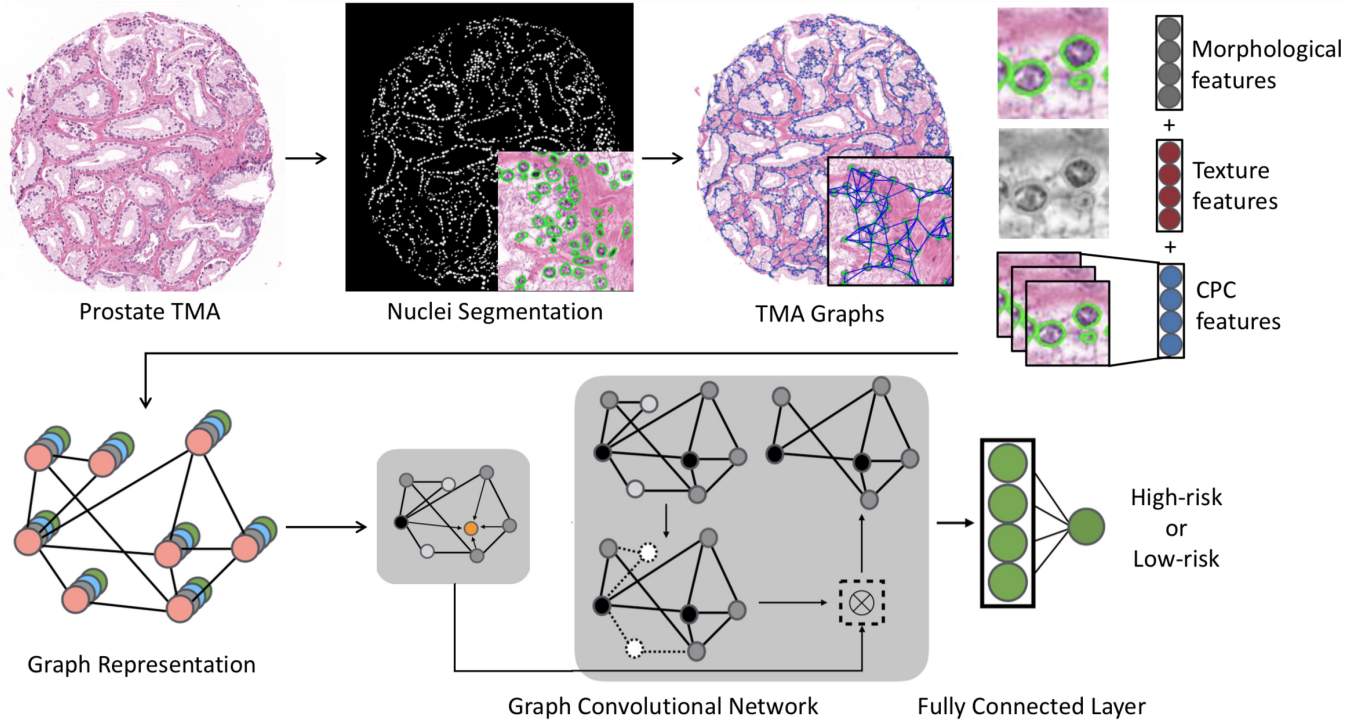


Fig. 1. For cell identification nuclei are segmented in each TMA and a graph is built on the centroid of each nuclei using the K-nearest neighbours (KNN) algorithm. Morphological, texture and contrastive predictive coding (CPC) features are then extracted for each nucleus and GCNs are applied on the graph representation. Finally, a fully connected layer is used at the end for classification.

image. We train a CPC encoder on extracted TMA patches of size 256×256 and a stride of 128. We then use the trained model to extract features from a window of size 64×64 around the centroid of each detected nucleus.

When applying CPC to prostate TMAs at the scale of 256×256 patches, we expect that the context (summarized from rows of features that are visible to the network) and unknown future observations (rows of features hidden from the network) are conditionally dependent on shared high-level information specific to the underlying pathology and the tissue site at this resolution. Examples of such high-level information might include the morphology and distinct arrangement of different cell types, the tissue texture and micro-environment, etc. By minimizing the CPC objective in the latent space, we implicitly encourage learning of such shared high-level abstractions specific to prostate tissue, which one is unlikely to effectively capture by using features extracted through naive transfer learning from ImageNet.

For each nucleus in a prostate TMA, we generate 8 morphological features, 4 texture features from GLCMs and 1024 features from CPC. Then, we concatenate them together to form a feature matrix $V \in R^{N_i \times F}$ where N_i is the number of nuclei in the graph and F is the number of features (1036 in our method).

3.2. Graph Convolution Networks

While CNNs remain a powerful deep learning tool for medical image analysis, they can only perform on grid-like Euclidean data like 2-dimensional images. Studies have been exploring how to generalize convolutional networks onto non-Euclidean domains. Convolution operators have been refined by Kipf *et al.* [9] to perform on non-grid-like data as graphs like protein structure and nuclei distribution. Through message passing, each node is able to iteratively accumulate feature vectors from its neighboring nodes and generate a new feature vector at the next hidden layer of the network, thus GCNs can learn to represent for each feature in a node. This process is very similar to feed-forward networks or CNNs. Then, by pooling over all the nodes, we are able to achieve the presentation for the entire graph. Such graph presentation can then serve as inputs for tasks as graph-level classification.

In our method, we used the GraphSAGE convolution [23]. The convolution and pooling operations can be defined as follows:

$$a_v^{(k)} = \text{MAX} \left(\left\{ \text{ReLU} \left(W \cdot h_u^{(k-1)} \right), \forall u \in \mathcal{N}(v) \right\} \right)$$

$$h_v^{(k)} = W \cdot \left[h_v^{(k-1)}, a_v^{(k)} \right]$$

At the $(k-1)^{th}$ iteration of the neighborhood accumulation, the feature vector for node v is denoted by $h_v^{(k)}$ while the feature vector of node v at the next iteration is represented by $a_v^{(k)}$.

Table 1. A comparative analysis of prostate TMA classification using various models and features.

Model	Accuracy \uparrow	AUC \uparrow
GCN + GLCM features	0.6934 \pm 0.0640	0.7067 \pm 0.0591
GCN + GLCM + Transfer Learning (VGG19) [17]	0.7373 \pm 0.042	0.8345 \pm 0.0319
GCN + GLCM + CPC features (Proposed)	0.8861 \pm 0.056	0.9637 \pm 0.0131

For the pooling part, we utilize the self-attention graph pooling method [24]. Self-attention can reduce computational complexity and take the topology into account. By performing self-attention graph pooling, our method is able to calculate attention scores, focus on the more meaningful node features and aggregate information of nuclei graph topology on different levels. The attention score $Z \in \mathcal{R}^{N \times 1}$ for nodes in G can be calculated as follows:

$$Z = \sigma(\text{SAGEConv}(X, A + A^2))$$

The node features are represented by X , the adjacency matrix is suggested by A , while SAGEConv is the convolution operator from GraphSAGE.

4. EXPERIMENTS AND RESULTS

We train both the CPC and the GCN on a high-resolution H&E stained image dataset from 5 prostate TMAs, each containing 200-300 spots [25]. There are 886 images in total. For each image, there is a pixel-level annotation mask suggesting the Gleason scores. In order to obtain the TMA-level labels, we calculate the primary grade and secondary grade based on the area in the mask and sum up the two grade together as the final Gleason score. For data augmentation, we extracted 1550×1550 crops from the four corners and the center of each image, and computed their corresponding binary labels. The crops that do not contain tissue are discarded. We classify the TMAs with Gleason score higher than (including) 6 as high-risk while TMAs with Gleason score lower than 6 as low-risk. Gleason scores of $3 + 3$ and $3 + 4$ are known to have the most interobserver and intraobserver variability [26] and our proposed method can stratify between the two highly variable scores. We validate our approach on 5 class-balanced splits where each splits contain 3498 crops for training, 388 crops for validation and 432 crops for testing. Table 1. shows a comparative analysis of our proposed method with other methods in a five-fold cross validation.

For comparison, we trained GCNs with the features generated from GLCMs and VGG19 in contrast to the features generated from CPC. Results show GCNs with a combination of GLCM and CPC features achieve state-of-the-art results with an AUC of 0.9659.

5. CONCLUSIONS

In the paper, we propose an approach for using GCNs as a tool to stratify patients on prostate cancer. Our approach is able to accurately identify high-risk patients using the prostate TMAs with only the image-level labels instead of the pixel-level labels. Also, we demonstrate that using CPC features in

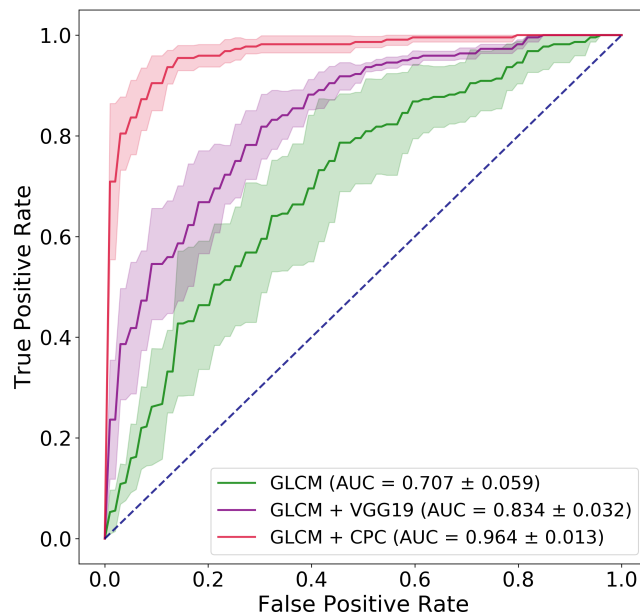


Fig. 2. Comparative analysis of our approach (GCN with GLCM+CPC features) as compared to other approaches demonstrates that our method achieves a 36.36% higher AUC over GCNs with just GLCM features and a 15.48% improvement over GCNs with VGG19 features. The shade suggests the confidence interval from the 5 random cross validation.

graph convolution can outperform features generated through simple transfer learning from ImageNet.

Our deep learning-based TMA stratification pipeline can be used as an assistive tool for pathologists to automatically stratify out high-risk cases before review. This work has the potential to alleviate the overall pathologist burden, accelerate clinical workflows and reduce costs. Future work will involve scaling this to a six class classification problem, identifying nodes being activated for interpretability and fusing information from patient and familial histories and incorporating multi-omics information for improved diagnostic and prognostic determinations. We will also explore methods to scale this to whole slide images.

6. REFERENCES

- [1] Thomas J Fuchs and Joachim M Buhmann, "Computational pathology: Challenges and promises for tissue analysis," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7-8, pp. 515-530, 2011.
- [2] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermesen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-

- Van De Kaa, Peter Bult, Bram Van Ginneken, and Jeroen Van Der Laak, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Scientific reports*, vol. 6, pp. 26286, 2016.
- [3] Rasool Fakoor, Faisal Ladhak, Azade Nazi, and Manfred Huber, "Using deep learning to enhance cancer diagnosis and classification," in *Proceedings of the international conference on machine learning*. ACM New York, USA, 2013, vol. 28.
 - [4] Anant Madabhushi and George Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," 2016.
 - [5] Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, and Michael Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," *Annual Review of Pathology: Mechanisms of Disease*, vol. 8, pp. 331–359, 2013.
 - [6] Yaniv Bar, Idit Diamant, Lior Wolf, Sivan Lieberman, Eli Koenen, and Hayit Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 294–297.
 - [7] Mehmet Günhan Ertosun and Daniel L Rubin, "Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks," in *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, 2015, vol. 2015, p. 1899.
 - [8] Eirini Arvaniti, Kim S Fricker, Michael Moret, Niels Rupp, Thomas Hermanns, Christian Fankhauser, Norbert Wey, Peter J Wild, Jan H Rueschoff, and Manfred Claassen, "Automated gleason grading of prostate cancer tissue microarrays via deep learning," *Scientific reports*, vol. 8, 2018.
 - [9] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
 - [10] Aäron van den Oord, Yazhe Li, and Oriol Vinyals, "Representation learning with contrastive predictive coding," *CoRR*, vol. abs/1807.03748, 2018.
 - [11] Ming Y Lu, Richard J Chen, Jingwen Wang, Debora Dillon, and Faisal Mahmood, "Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding," *arXiv preprint arXiv:1910.10825*, 2019.
 - [12] Parmeshwar Khurd, Claus Bahlmann, Peter Maday, Ali Kamen, Summer Gibbs-Strauss, Elizabeth M Genega, and John V Frangioni, "Computer-aided gleason grading of prostate cancer histopathological images using texton forests," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010, pp. 636–639.
 - [13] Oscar Jiménez del Toro, Manfredo Atzori, Sebastian Otálora, Mats Andersson, Kristian Eurén, Martin Hedlund, Peter Rönquist, and Henning Müller, "Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade gleason score," in *Medical Imaging 2017: Digital Pathology*. International Society for Optics and Photonics, 2017, vol. 10140, p. 101400O.
 - [14] Yanning Zhou, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir Rajpoot, "Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images," *arXiv preprint arXiv:1909.01068*, 2019.
 - [15] Fritz Albrechtsen et al., "Statistical texture measures computed from gray level cooccurrence matrices," *Image processing laboratory, department of informatics, university of oslo*, vol. 5, 2008.
 - [16] P Mohanaiah, P Sathyanarayana, and L GuruKumar, "Image texture feature extraction using glcm approach," *International journal of scientific and research publications*, vol. 3, no. 5, pp. 1, 2013.
 - [17] Shrey Gadiya, Deepak Anand, and Amit Sethi, "Histograms: Graphs in histopathology," *arXiv preprint arXiv:1908.05020*, 2019.
 - [18] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al., "Bach: Grand challenge on breast cancer histology images," *Medical image analysis*, 2019.
 - [19] Peter Naylor, Marick Laé, Fabien Rey, and Thomas Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 933–936.
 - [20] Faisal Mahmood, Daniel Borders, Richard Chen, Gregory N McKay, Kevan J Salimian, Alexander Baras, and Nicholas J Durr, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE transactions on medical imaging*, 2019.
 - [21] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550–1560, July 2017.
 - [22] Sahibsingh A Dudani, "The distance-weighted k-nearest-neighbor rule," *IEEE Transactions on Systems, Man, and Cybernetics*, , no. 4, pp. 325–327, 1976.
 - [23] Will Hamilton, Zhitao Ying, and Jure Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, 2017, pp. 1024–1034.
 - [24] Junhyun Lee, Inyeop Lee, and Jaewoo Kang, "Self-attention graph pooling," *arXiv preprint arXiv:1904.08082*, 2019.
 - [25] Qing Zhong, Tiannan Guo, Markus Rechsteiner, Jan H Rüschoff, Niels Rupp, Christian Fankhauser, Karim Saba, Ashkan Mortezaei, Cédric Poyet, Thomas Hermanns, et al., "A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients," *Scientific data*, vol. 4, pp. 170014, 2017.
 - [26] Tayyar A Ozkan, Ahmet T Erucar, Oguz O Cebeci, Omur Memik, Levent Ozcan, and Ibrahim Kusonmaz, "Interobserver variability in gleason histological grading of prostate cancer," *Scandinavian journal of urology*, vol. 50, no. 6, pp. 420–424, 2016.