Journal of Pathology Informatics 15 (2024) 100344

Contents lists available at ScienceDirect

Journal of Pathology Informatics

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

Learning to predict prostate cancer recurrence from tissue images

Mahtab Farrokh [a,*], Neeraj Kumar [a,b], Peter H. Gann [c], Russell Greiner [a,b]

[a] Department of Computing Science, University of Alberta, Alberta, Canada
[b] Alberta Machine Intelligence Institute, Edmonton, Canada
[c] Department of Pathology, College of Medicine, University of Illinois at Chicago, Chicago, IL, United States

ARTICLE INFO

Keywords:
Machine learning
Computer vision
Contrastive learning
Prostate cancer
Histopathology

ABSTRACT

Roughly 30% of men with prostate cancer who undergo radical prostatectomy will suffer biochemical cancer recurrence (BCR). Accurately predicting which patients will experience BCR could identify who would benefit from increased surveillance or adjuvant therapy. Unfortunately, no current method can effectively predict this. We develop and evaluate PathCLR, a novel semi-supervised method that learns a model that can use hematoxylin and eosin (H&E)-stained tissue microarrays (TMAs) to predict prostate cancer recurrence within 5 years after diagnosis. The learning process involves 2 sequential steps: PathCLR (a) first employs self-supervised learning to generate effective feature representations of the input images, then (b) feeds these learned features into a fully supervised neural network classifier to learn a model for predicting BCR. We conducted training and evaluation using 2 large prostate cancer datasets: (1) the Cooperative Prostate Cancer Tissue Resource (CPCTR) with 374 patients, including 189 who experienced BCR, and (2) the Johns Hopkins University (JHU) prostate cancer dataset of 646 patients, with 451 patients having BCR. PathCLR's (10-fold cross-validation) F1 score was 0.61 for CPCTR and 0.85 for JHU. This was statistically superior (paired t-test with $P < .05$) to the best-learned model that relied solely on clinicopathological features, including PSA level, primary and secondary Gleason Grade, etc. We attribute the improvement of PathCLR over models using only clinicopathological features to its utilization of both learned latent representations of tissue core images and clinicopathological features. This finding suggests that there is essential predictive information in tissue images at the time of surgery that goes beyond the knowledge obtained from reported clinicopathological features, helping predict the patient's 5-year outcome.

Introduction

World-wide each year, around 1.6 million men are diagnosed with prostate cancer, and 366 000 will die of the disease.[22] A majority of these patients undergo radical prostatectomy, either as an initial treatment choice or following a period of active surveillance.[32] Within 5 years of the surgery, about 20%–40% of these men who undergo radical prostatectomy experience biochemical cancer recurrence (BCR), which is detected by elevated prostate-specific antigen (PSA) levels.[29] The accurate prediction of patients prone to experiencing BCR after surgery is crucial for determining the most appropriate post-surgery course of action. This includes identifying patients who might benefit from additional treatment, advanced imaging to detect metastases, genomic testing, or more frequent monitoring.

More specifically, while it is true that most patients with BCR do not die from the disease, it is still a strong risk factor for subsequent metastasis and mortality. Patients with a "Yes" prediction could benefit from more frequent PSA monitoring. In addition, this "Yes" prediction suggests that the patient could benefit from new imaging techniques (e.g., PSMA-targeted PET) (PSMA = prostate-specific membrane antigen; PET = positron emission tomography) to detect occult metastases or local/regional spread soon after surgery. If this produces visualized lesions, the patient could opt for local or systemic therapy or both. In fact, recent clinical trials by Pienta et al.[23] have explicitly demonstrated this. Another trial, CONDOR, found that PSMA-targeted PET was able to detect occult spread in a majority of patients with low levels of BCR (PSA close to 0.2), and 64% of those patients changed management strategy.[18]

Due to the frequency of BCR occurrence, several projects have been developed to create multivariable prediction tools aimed at predicting BCR at the time of diagnosis. These tools include models utilizing clinicopathological features and gene profiling.[2] One such model, the CAPRA-S score, employs clinicopathological features such as PSA level, Gleason sum score, and other pathology-reported variables to calculate a risk score associated with progression-free probability at 3 and 5 years; it has shown favorable results for BCR risk assessment.[6,25] Additionally, some approaches use gene profiling biomarkers, such as the Decipher test, on bulk samples of surgical tissue, to forecast recurrence after prostatectomy.[2,26,27] Despite the advancements facilitated by these models, widespread clinical adoption

* Corresponding author.
E-mail address: farrokh@ualberta.ca (M. Farrokh).

http://dx.doi.org/10.1016/j.jpi.2023.100344
Received 28 August 2023; Received in revised form 21 October 2023; Accepted 1 November 2023
Available online 4 November 2023

has not occurred due to cost barriers or failure to achieve the required level of accuracy.

Therefore, in this study, we introduce a novel deep learning-based method to construct BCR prediction models that leverage both clinicopathological features and information from inexpensive, routinely available images of tumor tissue slides stained with hematoxylin and eosin (H&E), obtained after surgery. Recent research has already demonstrated that deep neural networks can identify recurring tumors.[24] These earlier systems rely on common approach *supervised* learning, as each involves directly learning a model that labels each instance with its outcome. By contrast, our method, called PathCLR (Pathology Contrastive LeaRning), takes a *semi-supervised* approach, which involves a self-supervised learning step before the supervised step. Note that emerging evidence suggests that semi-supervised learning often outperforms traditional supervised learning methods, especially when labeled data is limited, as is often the case in medical research.[1,5,20]

This motivates our PathCLR approach, which first trains a self-supervised SimCLR model[4] on tissue microarray images (TMAs) to generate a set of latent representations for a given tissue core of a new patient. Note this self-supervised step is *unsupervised*, as it does not require a BCR label for each instance. Subsequently, PathCLR utilizes this set of latent representations, along with relevant clinical features and reported pathology variables of the patient, to train a neural network (NN) classifier. The goal is to predict whether a specific patient will experience BCR within 5 years after surgery. This personalized approach sets PathCLR apart from previous methods, as PathCRL: (1) provides a prediction about an individual patient (and not about *groups* of patients) and (2) provides a specific individual "Yes/No" prediction for the patient (and not a hazard ratios relative to a baseline).

Our research findings demonstrate that incorporating a novel semi-supervised machine learning method shows promise for rapid and inexpensive prediction of BCR directly from routinely stained tumor images, especially when combined with clinicopathological variables. Our proposed method is cost-effective because many centers now routinely capture high-resolution H&E scans for all cancer types, with each image costing approximately $4 and a scan duration of around 1 min per slide (or about 3 min at 40× magnification). Consequently, PathCLR, leveraging these readily available images without the need for additional expert annotations, presents a highly cost-effective approach.

We suggest that, indeed, the pathology slides contain valuable information beyond the conventional variables such as Gleason grade, PSA level, surgical margin, etc. Furthermore, we demonstrate that the label-free learning approach on TMA cores, which produces latent representations of their histological patterns, plays a critical role in achieving improved BCR prediction accuracy. Our main contributions are:

1. Our work focuses on challenging datasets where the majority of Gleason grade sum scores are 7 (3 + 4 or 4 + 3). This differs from many previous studies that focus on BCR prediction without matching patients with respect to their clinicopathological characteristics, specifically the Gleason sum score. As we conduct a study on BCR on a carefully selected cohort of patients with matched clinicopathological variables, we focus on more challenging cases for BCR prediction.
2. We introduce PathCLR, an automated pipeline capable of efficiently processing a patient's TMA cores and providing BCR probability predictions in less than 2 s.
3. To demonstrate the effectiveness of our pipeline in recurrence prediction, we compare our semi-supervised SimCLR method with a supervised CNN model, CAPRA-S score, and a previous benchmark[17] for BCR prediction. This comparison shows that our pipeline predicts recurrence more accurately than these alternative approaches.

This paper is organized as follows: Section Material and methods describes the 2 datasets that we used and then proposes the PathCLR pipeline. Section Evaluation metrics provides the evaluation metrics that we used in Section Result. We compare our proposed approach to previous work on BCR prediction in Section Discussion, before concluding in Section Conclusions.

Material and methods

Tissue microarrays and image pre-processing

We utilized images and data from 2 prostate cancer cohorts: TMAs and clinicopathological data from the Cooperative Prostate Cancer Tissue Resource (CPCTR),[21] funded by the National Cancer Institute, USA and the PSA Progression dataset from the Prostate Cancer Biorepository Network (PCBN), funded by the Prostate Cancer Research Program of the US Department of Defense. We refer to the later cohort as JHU as all patients in this dataset were treated at Johns Hopkins University.

Both TMAs in our study employed a matched case–control design to define recurrence based on BCR or clinical progression after surgery. Guidelines describe post-RP BCR defined as a serum PSA of more than 0.2 ng/ml for both CPCTR and JHU datasets. The BCR label in both datasets is based on a single evaluation unless the reported PSA level was very close to the threshold. However, they differed in their approach to selecting control tumors. In the CPCTR dataset, controls were chosen from patients who survived at least 5 years without experiencing BCR. The matching was done on a 1:1 basis, considering factors such as age, race, Gleason grade (primary and secondary), and the treating hospital.

On the other hand, the JHU cohort used an incidence density sampling approach. For each case with BCR, 1 control was selected from the pool of patients who had not experienced BCR up to that specific time point. Case–control matching was done based on age, race, pathologic stage, and Gleason sum. This sampling approach allowed for the possibility that initial controls could later become cases and also that controls could be selected more than once.

Since our primary goal was to compare the risk of developing BCR within 5 years, we limited the selection of control patients to those who survived the entire 5-year period without experiencing BCR. TMA images were pre-processed to remove spots with insufficient tissue, ensuring that those with more than 80% white background were eliminated. The flow chart in Fig. 1 shows how images from recurrent or non-recurrent patients were selected for final analyses. The CPCTR dataset comprised 189 cases and 185 controls, with a total of 1281 TMA cores. In the JHU dataset, there were 451 cases and 195 controls comprising 2983 cores. Both datasets included up to 4 cores per patient.

The main characteristics of the patients and tumors included in our analysis are summarized in Table 1 and Table 2, for CPCTR and JHU datasets, respectively. Note that the PSA feature in Tables 1 and 2 represents the PSA laboratory readings immediately post-prostatectomy. This is distinct from the later rise in PSA levels, which indicates BCR. The TMA image resolution in the CPCTR dataset is 40×, while in the JHU dataset is 20×. We utilized the Python programming language along with OpenCV[1] and the Python Image Library (PIL)[2] for image visualization.

To fine-tune the patch extractor component of PathCLR, we employed the MoNuSAC dataset[31] for learning a model to segment epithelial cells. This dataset is significant as it is extensive, diverse, and hand-annotated, specifically designed for the MoNuSAC2020 challenge, which aimed to identify cancerous epithelial cells and 3 types of immune cells from H&E stained tissue images. Notably, the top methods submitted to this challenge achieved inter-human agreement levels in their performance.

The PathCLR model

Taking inspiration from the success of recent studies in applying machine learning techniques to the medical field, we propose a semi-supervised method for BCR prediction. Semi-supervised learning has

[1] https://opencv.org/.
[2] https://pillow.readthedocs.io/.

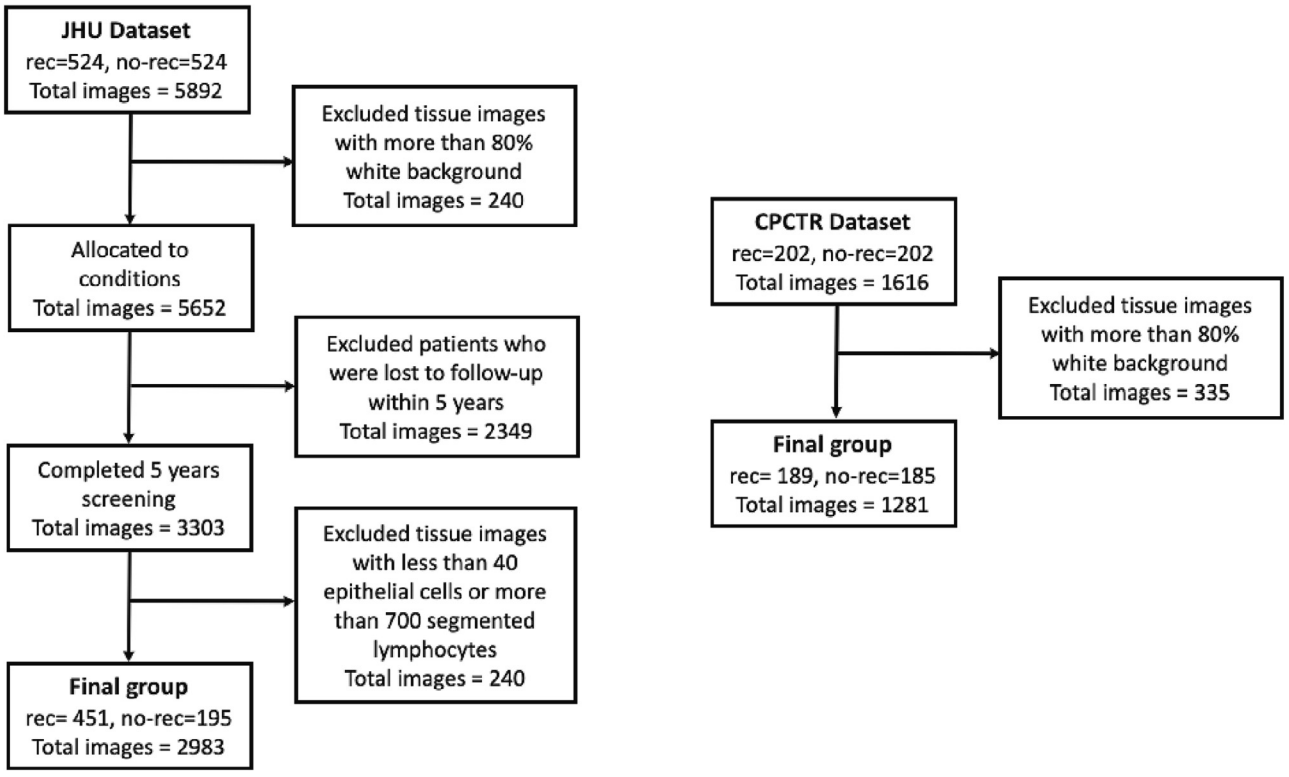

Fig. 1. Pre-processing flow chart for JHU (left) and CPCTR (right) datasets.

proven to achieve notable performance using fewer labeled samples for training when compared to previous fully supervised approaches.[20] In our proposed PathCLR algorithm, the initial step involves learning the latent representations in a self-supervised manner, which means no diagnostic label is required during this training phase. Subsequently, PathCLR utilizes these latent representations to describe each training patient and then employs their outcome labels to train a model capable of predicting BCR in a supervised manner.

To learn the feature representation of tissue images, we adopted the state-of-the-art self-supervised SimCLR algorithm, which is a simple framework for contrastive learning of visual representations. In essence, SimCLR learns the latent representation of an image by maximizing agreement between differently augmented views of the same image through a contrastive loss.[4] To use the training examples effectively, we implemented various data augmentations, such as random cropping, color distortion, and blurring. We employed contrastive learning[15] to acquire highly contrasting features from TMA cores. An advantage of the proposed PathCLR model is its adaptability with varying numbers of TMA cores per patient.

To begin our analysis, we recognize the importance of pre-processing tissue images, considering that these images may vary in size and degree of centering. Extracting informative image patches is a crucial initial step. Additionally, we believe that the most relevant regions for BCR prediction

Table 1
Characteristics of CPCTR dataset after data cleaning. Abbreviations: Gleason Sum (GS), Prostate-Specific Antigen (PSA), Lymph Node Invasion (LN), Seminal Vesicle Invasion (SVI), Surgical Margin (SM), Extra Capsular Extension (ECE).

CPCTR	Total, *n* (%)	No-BCR, *n* (%)	BCR, *n* (%)
#Patients	374	185	189
GS = 3 + 2	2 (0.5)	1 (0.5)	1 (0.5)
GS = 3 + 3	85 (22.7)	41 (22.1)	44 (23.2)
GS = 3 + 4	196 (52.4)	98 (52.9)	98 (51.8)
GS = 3 + 5	7 (1.8)	2 (1.1)	5 (2.6)
GS = 4 + 3	56 (14.9)	27 (14.5)	29 (15.3)
GS = 4+4	18 (4.8)	11 (5.9)	7 (3.7)
GS = 4 + 5	9 (2.4)	5 (2.7)	4 (2.1)
GS = 5 + 3	1 (0.2)	0 (0.0)	1 (0.5)
PSA	10.5 ± 11.5	8.8 ± 6.18	12.2 ± 14.8
LN	17 (4.5)	6 (3.2)	11 (5.8)
SVI	17 (4.5)	8 (4.3)	9 (4.7)
SM = tumor free	226 (60.4)	124 (67)	102 (53.9)
SM = tumor focal at margin	114 (30.4)	49 (26.4)	65 (34.3)
SM = tumor widespread at margin	30 (8.6)	9 (4.8)	21 (11.1)
SM = unknown	4 (1)	3 (1.6)	1 (0.5)
ECE = none	235 (62.8)	111 (60)	124 (65.6)
ECE = multifocal	28 (7.4)	13 (7)	15 (7.9)
ECE = focal	98 (26.2)	56 (30.2)	42 (22.2)
ECE = established	12 (3.2)	4 (2.16)	8 (4.2)
ECE = unknown	1 (0.2)	1 (0.5)	0 (0)

Table 2
Characteristics of JHU dataset after data cleaning. Abbreviations: Gleason Sum (GS), Prostate-Specific Antigen (PSA), Lymph Node Invasion (LN), Seminal Vesicle Invasion (SVI), Surgical Margin (SM), Established Capsular Penetration (ECP), Focal Capsular Penetration (FCP).

JHU	Total, *n* (%)	No-BCR, *n* (%)	BCR, *n* (%)
#Patients	646	195	451
GS = 2 + 3	2 (0.3)	2 (1.0)	0 (0)
GS = 3 + 2	5 (0.7)	3 (1.5)	2 (0.4)
GS = 3 + 3	99 (15.3)	49 (25.1)	50 (11.0)
GS = 3 + 4	263 (40.7)	94 (48.2)	169 (37.4)
GS = 3 + 5	13 (2.0)	3 (1.5)	10 (2.2)
GS = 4 + 3	136 (21.0)	29 (14.8)	107 (23.7)
GS = 4 + 4	60 (9.2)	8 (4.1)	52 (11.5)
GS = 4 + 5	56 (8.66)	6 (3.0)	50 (11.0)
GS = 5 + 3	3 (0.4)	0 (0)	3 (0.6)
GS = 5 + 4	9 (1.3)	1 (0.5)	8 (1.7)
PSA	11.72 ±9.66	10.21 ± 7.75	12.38 ± 10.32
LN	99 (15.3)	11 (5.6)	88 (19.5)
SVI	149 (23.0)	29 (14.8)	120 (26.6)
SM = postive	196 (30.3)	35 (17.9)	161 (35.6)
ECP = positive	413 (63.9)	111 (56.9)	321 (71.1)
FCP = positive	235 (36.3)	102 (52.3)	133 (29.4)

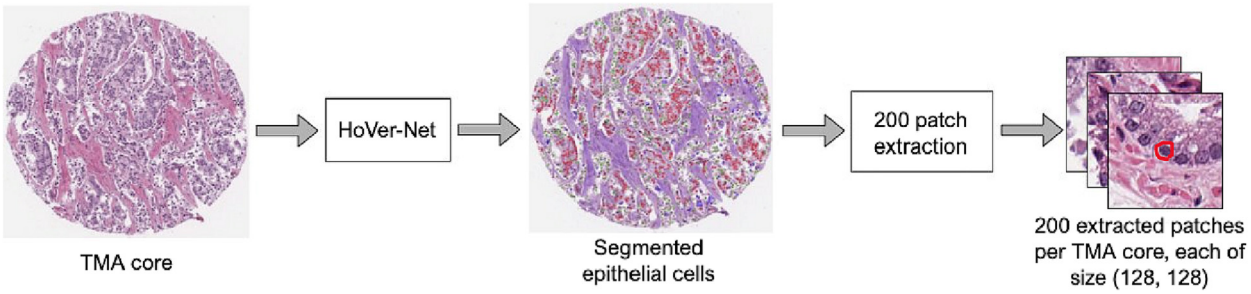

Fig. 2. The first part of the PathCLR process extracts 200 patches in size 128 × 128 pixels in RGB format around segmented epithelial cells. The output of HoVer-Net is a segmented tissue core that shows each epithelial cell with a red circle.

are likely to be around epithelial cells and the surrounding stroma, as demonstrated in a prior study.[16]

To segment epithelial cells, we employed the HoVer-Net model,[10] known for its capabilities in automatic nuclear instance segmentation and classification in histology images. This model simultaneously detects epithelial cells, lymphocytes, macrophages, and neutrophils. The HoVer-Net model is currently a state-of-the-art model for segmenting these nuclei with high accuracy. We fine-tuned the HoVer-Net model on the MoNuSAC2020 dataset, utilizing image augmentation and pretrained weights to achieve even higher segmentation accuracy.[31]

To accommodate SimCLR's requirement of fixed-size inputs, we randomly sampled 200 patches, each measuring 128 × 128 pixels in spatial dimensions, from every TMA image. These patches were centered at the positions of each nucleus detected by the HoVer-Net model. However, the number of detected epithelial cells in each TMA core varied, which could differ from the requirement that each patch include exactly 200 nuclear centers.

To address this, if we identified more than 200 epithelial cells in a TMA core, we employed clustering to group the (x,y) epithelial cell center positions into 200 clusters. From these clusters, we selected the center of each cluster as one of the 200 regions of epithelial cells of interest. This clustering approach aimed to maximize the coverage of tumor-rich areas and minimize the overlapping of extracted patches.

For TMA cores with fewer than 200 epithelial cells, we took the (x,y) spatial position of the k detected epithelial cell centers and added $200-k$ additional spatial positions in the vicinity of the detected cells. This process allowed us to generate 200 patches for each TMA core. Fig. 2 visually illustrates the described patch extraction process from each TMA core image.

As depicted in Fig. 3, the PathCLR pipeline consists of 2 main components. First, the self-supervised SimCLR model learns the latent representation of all extracted tissue patches. In this step, we used ResNet34[11] as the base architecture in SimCLR to produce latent representations of input image patches.

Secondly, the final supervised binary classification neural network incorporates all relevant information specific to each patient. This includes both the imaging features expressed through the SimCLR encoding and 7 observed clinicopathological features, such as PSA level, surgical margin, Gleason grade, and others. The objective of this step is to train a neural network model that can accurately predict whether a patient will experience BCR within 5 years after surgery or not.

Note that the final supervised binary classification neural network can capture nonlinear interactions among various input features, encompassing both latent representations of TMA image patches and clinicopathological variables. This capability is important for accurate BCR prediction.

The described binary classifier predicts the recurrence probability within 5 years for each extracted patch. We then compute the average recurrence probabilities of all patches from a single TMA core. If the average probability from a TMA core is higher than a specified threshold (in this case, 0.6), we predict that the patient will experience BCR. The threshold was empirically determined by internal cross-validation. If a patient has multiple tissue cores available, we calculate the average recurrence prediction probability for each TMA core and return the "disjunction"—indicating a positive prediction if any of the tissue cores is predicted as a recurrence case. We considered various combination rules (e.g., conjunction, majority) but found that the disjunction method performed better (see Section 4).

We also compared the results from the binary classification neural network using: (1) only clinicopathological features, (2) only the SimCLR output, and (3) both input sources combined. The PathCLR pipeline is implemented using the Python programming language and the Keras library[3]. The HoVer-Net and SimCLR parts of the pipeline are implemented using the PyTorch library[4]. Image pre-processing and transformations were done using the transformers in the torchvision[5] library. Table 3 provides more information regarding the main components of the SimCLR pipeline. Note, that the provided configurations for SimCLR in this table are given for the fine-tuning phase because we used pretrained weights of SimCLR.

Evaluation metrics

In the context of the BCR binary classification task, selecting an appropriate evaluation metric is crucial. In our study, we reported 2 main metrics: accuracy and F1-score:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + TP} \tag{1}$$

$$F1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{2}$$

where TP (resp., TN, FP, FN) is the number of true-positive (resp., true-negative, false-positive, false-negative) instances. These metrics are commonly used for binary classification tasks and provide valuable insights into the performance of our predictive model.

To assess the performance of PathCLR, we employed 10-fold cross-validation and reported a 95% confidence interval. This involved training-&-testing 10 times, with each iteration training on 9 parts (representing 90% of the dataset) and testing on 1 left-out part (representing 10% of the dataset). For evaluating our epithelial segmentation, we used the panoptic quality (PQ) metric,

$$PQ = \frac{\Sigma_{(p,g)\in TP} IoU(p,g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \tag{3}$$

which was also used in the MoNuSAC2020 data challenge to evaluate the effectiveness of nuclear segmentation and classification algorithms. Here, p represents the predicted set of pixels in the nucleus, g represents the set

[3] https://keras.io/.
[4] https://pytorch.org/.
[5] https://pytorch.org/vision/.

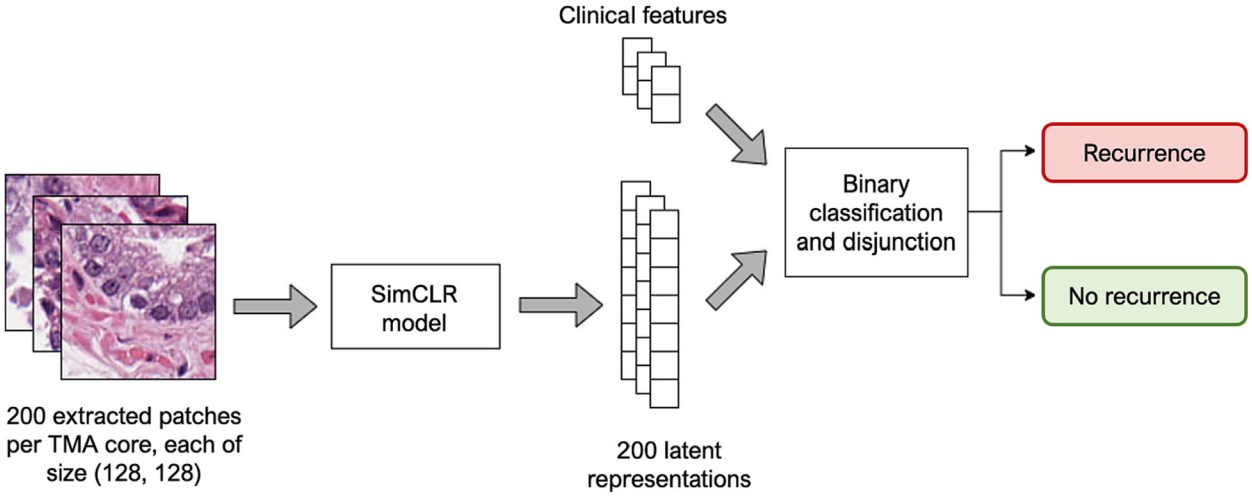

Fig. 3. The second part of the PathCLR pipeline, which predicts prostate cancer recurrence using 200 extracted patches per TMA core. First, SimCLR learns the latent representation of each tissue patch. Next, a learned binary classifier predicts BCR using the learned latent representations along with 7 clinicopathological features, following a disjunction function to decide patient-level recurrence.

Table 3
Configuration and overview of components of PathCLR.

Model	#Parameters	#Layers	Activation function	Optimization	# Epochs	Learning rate	Batch size
SimCLR	21M	34	ReLU	SGD/Adam	100	0.0003	256
Binary classifier	32K	4	ReLU	Adam	200	0.001	32

of pixels in the ground truth, and $IoU(p,g)$ calculates the Intersection over Union (IoU) by dividing the number of elements in the intersection of the sets by the size of their union.

To determine the significance of our results, we utilized a paired 2-sided t-test with a significance level of $P < .05$—i.e., a result is considered significant if the P-value obtained from the t-test is less than .05.

Results

As described in Section 2, the initial step of the PathCLR pipeline involves segmenting epithelial cells and utilizing the segmented epithelial cell centers to extract 200 patches from a TMA core. Our fine-tuned HoVer-Net model demonstrated superior performance compared to the winning results from the MoNuSAC2020 challenge, indicating its promising capabilities for epithelial cell detection. While the best reported PQ for the validation dataset in the MoNuSAC2020 challenge website is 0.611, our model achieved a higher PQ of 0.675, showing its improved performance in epithelial cell segmentation.

To investigate the impact of using H&E-stained tissue slides on recurrence prediction, we trained models in 3 distinct settings:

1. Using only clinicopathological features.
2. Using only H&E-stained TMA cores.
3. Using both clinicopathological features and H&E-stained TMA cores.

In order to conduct a fair evaluation of setting #1, we systematically explored various combinations of clinicopathological features and different learning models. After thorough experimentation, we found that the best outcome was achieved by employing a 2-layer feed-forward neural network with cross-entropy loss. Importantly, the optimal result was obtained by using only the precise set of CAPRA-S score's features.[6] The selected clinicopathological features for the first setting are as follows:

Clinicopathological features = {PSA Level, Surgical Margin, Primary Gleason Grade, Secondary Gleason Grade, Extracapsular Extention, Lymph Node, Seminal Vesicle}.

In setting #3, we maintained the same listed set of clinicopathological features for a fair comparison with the other settings. In both settings #2 and #3, we employed the SimCLR network using ResNet34 with Xavier weight initialization.[9] Note, the other setting #1 does not use ResNet34.

To handle binary classification in all 3 settings, we experimented with different numbers of layers and regularization techniques (both L1 and L2 regularization penalties along with early stopping) to prevent overfitting and optimize the model's performance.

Table 4 and Fig. 4 present the results obtained from 10-fold cross-validation on both the CPCTR and JHU datasets. Notably, using solely clinicopathological features yields results that are almost on par with a trivial "just say BCR" prediction, where BCR is predicted for all cases regardless of their features. This "just say BCR" prediction would achieve an accuracy of 69.81% (451/646) on the JHU dataset and 50.5% (189/374) accuracy on the CPCTR dataset.

We assume the reason clinicopathological variables alone are insufficient for BCR prediction lies in the fact that the case–control pairs are carefully matched for clinicopathological features in both the CPCTR and JHU datasets. Consequently, there are patients with similar clinicopathological features but different outcomes. As a result, classifiers relying solely on clinicopathological features struggle to distinguish between recurrence and non-recurrence cases, which hampers their predictive performance.

In setting #1, we compared the results of our pipeline with a conventional model that relies solely on the CAPRA-S score.[6] However, upon experimentation, we observed that applying a learned neural network

Table 4
10-fold cross-validation result with 95% CI. Setting #1 uses only clinicopathological features, setting #2 uses only H&E-stained TMA cores, and setting #3 uses both clinicopathological features and H&E-stained TMA cores.

	Accuracy		F1-Score	
	CPCTR	JHU	CPCTR	JHU
Setting #1	53.7 ± 2.77	69.81 ± 0.54	0.544	0.82
Setting #2	56.1 ± 3.47	69.82 ± 0.39	0.6	0.82
Setting #3	59.5 ± 3.29	75.25 ± 3.48	0.614	0.85

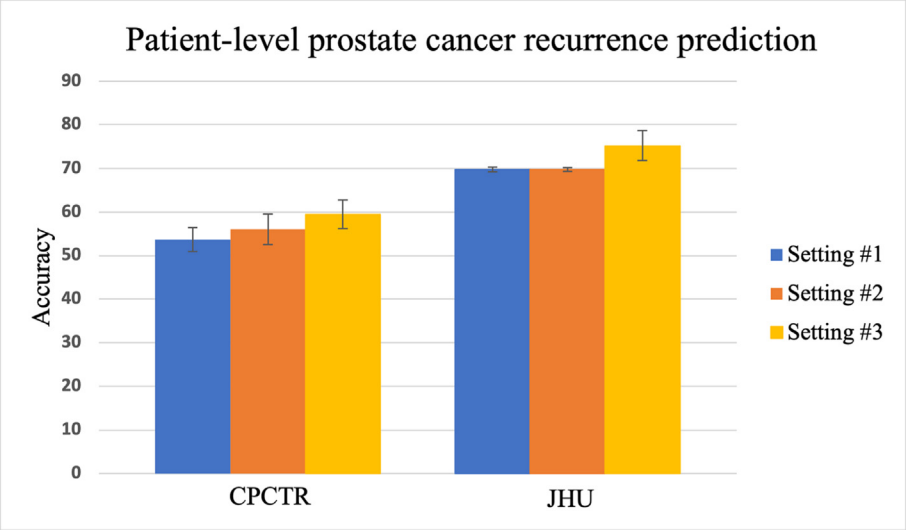

Fig. 4. Accuracy on JHU and CPCTR datasets 10-fold cross-validation. Error bar shows the 95% CI.

to the clinicopathological features yielded slightly better results. We attribute this improvement to the neural network's ability to discover non-linear patterns within the features, enhancing the predictive accuracy.

When using the CAPRA-S score with 10-fold cross-validation, we obtained an accuracy of 53.22 ± 5.93 on the CPCTR dataset, and 69.81 ± 0.0 on the JHU dataset. These results indicate that even the CAPRA-S score fails to accurately predict BCR in cases where patients have matched clinicopathological features, highlighting the challenges in making accurate predictions solely based on clinicopathological variables.

In setting #3, where we combined the clinicopathological features with latent representations of tissue cores, we observed that the accuracy of this model is at least 5% better than using only clinicopathological features. This improvement demonstrates that there is valuable additional information related to cancer recurrence embedded in the tissue slides. To confirm the significance of this improvement, we conducted paired 2-sided *t*-tests on both the JHU and CPCTR datasets, resulting in *P*-values of .005 and .04, respectively. Both *P*-values are below the significance threshold of .05, indicating that the accuracy improvement is statistically significant.

Comparing setting #2 to setting #1, we found that the accuracy and F1-score of setting #2 were higher. However, the paired *t*-tests revealed that the difference between the 2 settings was not statistically significant, with *P*-values of .19 in the JHU dataset and .48 in the CPCTR dataset. Despite not being statistically significant, the improved accuracy in setting #2 suggests the potential benefits of incorporating the learned latent representations of tissue cores along with clinicopathological features in BCR prediction. We also compared setting #2 versus setting #3: for the JHU dataset, setting #3 is significantly better ($P = .007$). However, for the CPCTR dataset, there is no ($P < .05$) significant difference ($P = .17$). Note, we did not apply any false-discovery rate (FDR) correction since the number of involved comparisons is small.

As mentioned in the previous section, we have up to 4 TMA cores per patient, and we predicted recurrence for each TMA core. To achieve this, we used various functions to combine the individual core scores, including disjunction, conjunction, and majority voting. The results of 10-fold cross-validation accuracy for these variants are presented in Table 5 for both the CPCTR and JHU datasets, corresponding to settings #2 and #3.

The results indicate that the accuracy of the disjunction method is higher than that of conjunction and majority voting. We hypothesize that some TMA cores may not provide informative data for cancer recurrence prediction. Therefore, if at least 1 TMA core shows a high probability of BCR, we consider the corresponding patient to have a high probability of experiencing BCR within 5 years. This approach allows us to leverage the most informative TMA cores for each patient, resulting in improved accuracy in predicting recurrence at the patient level.

Table 5
10-fold cross-validation accuracy with 95% CI in different combination modes.

	Setting #2		Setting #3	
	CPCTR	JHU	CPCTR	JHU
Majority	55.08 ± 0.26	69.82 ± 0.39	55.08 ± 0.55	74.44 ± 2.1
Conjunction	55.62 ± 0.26	69.82 ± 0.24	55.88 ± 0.55	74.93 ± 1.7
Disjunction	56.06 ± 3.47	69.82 ± 0.39	59.49 ± 3.29	75.25 ± 3.48

Table 6
Comparing PathCLR, a semi-supervised approach, with supervised and histotyping approaches, using 10-fold cross-validation accuracy and a 95% CI.

	Setting #2		Setting #3	
	CPCTR	JHU	CPCTR	JHU
PathCLR	56.06 ± 3.47	69.82 ± 0.39	59.49 ± 3.29	75.25 ± 3.48
Supervised	54.32 ± 2.2	69.4 ± 0.32	57.4 ± 4.14	70.59 ± 0.81
Histotyping	53 ± 0.01	68.3 ± 0.02	54 ± 0.02	70.00 ± 1.00

In addition to our PathCLR approach, we explored a fully supervised convolutional neural network (CNN) approach using the ResNet34 model. Instead of training on extracted patches of TMA cores, we trained the CNN directly on the whole tissue cores as images. For patients with multiple TMA cores, we still used the disjunction method for both the PathCLR and supervised models.

The results, as shown in Table 6, indicate that the accuracy of the supervised approach was approximately 2% lower than the PathCLR approach on both the CPCTR and JHU datasets. This suggests that our semi-supervised PathCLR approach, which leverages both clinicopathological features and learned latent representations from tissue images, outperforms the fully supervised CNN approach in predicting prostate cancer recurrence.

Drawing inspiration from a previous benchmark by Leo et al.,[17] which predicted BCR using H&E slides and meticulously crafted features of gland morphology, we performed a further comparison of PathCLR. However, rather than focusing on the lumen, we segmented the epithelial cells and derived an identical set of Histotyping features via the HistomicsTK library[6]. Our focus shifted to epithelial cells primarily because the TMA

[6] https://digitalslidearchive.github.io/HistomicsTK/index.html.

cores we examined contained limited lumen regions. Moreover, as previously stated, we hypothesize that epithelial cells provide more valuable information for predicting BCR. The results outlined in Table 6 show that PathCLR performs better than these histotyping features. Hence, the superior performance of PathCLR can be attributed to the learning of each image patch representation through contrastive loss. This embedded information proves to be more effective than the morphology features obtained through histotyping.

In terms of the run-time computational efficiency of the learned system, the processing time for each TMA core is under 2 s. Considering the number of TMA cores (up to 4 in this study), the prediction of BCR for an individual, on average, is completed in less than 7.3 s.

Discussion

Recently, the combination of computational pathology with machine learning (ML) techniques has shown the potential to enhance diagnostic accuracy and optimize patient care.[7] Nonetheless, there are many challenges in applying ML to pathology tasks, including the limited availability of labeled and annotated tissue samples.

Eksi et al.[8] conducted a study exploring projects that applied ML techniques to only clinicopathological parameters, to develop models for predicting BCR. Their findings revealed that all ML models outperformed the traditional statistical regression method. The use of ML methods demonstrated the potential for more accurate risk classification, improved prognosis estimation, earlier intervention, reduced unnecessary treatments, and lower morbidity and mortality. However, it is important to note that their study focused exclusively on clinicopathological features and did not integrate tissue microarray histopathology images into their analysis.

In previous studies, ML techniques have been employed to automate Gleason grading of H&E-stained tissue images.[3,14,19] Another study by Yan et al.[34] utilized contrastive learning on tissue images for cancer diagnosis. However, the application of ML to predict BCR has been relatively scarce.

Kumar et al.[16] attempted BCR prediction using 2 separate convolutional neural networks (CNNs) on a small number of cases from the CPCTR dataset, evaluating their model on only 30 cases. In contrast, our study utilizes 10-fold cross-validation on the entire CPCTR dataset, providing a more comprehensive evaluation. Additionally, Kumar et al.'s approach required manual annotation of tumors to predict BCR for a test patient, and their prediction was based solely on tissue images. In contrast, our approach does not require additional annotations and benefits from the integration of both clinicopathological features and learned latent representations from tissue images.

Manual annotation of histopathology images is a labor-intensive process, often requiring expert pathologists or even a group of pathologists to vote and annotate a tissue image. To overcome these challenges, we proposed a semi-supervised method that first learns latent representations of unlabeled tissue images and subsequently utilizes each patient's outcome as the label to predict BCR.

Yamamoto et al.[33] employed H&E-stained tissue images to predict BCR and generated low-dimensional features using an autoencoder[12] in 2 different resolutions. Then, they used a support vector machine (SVM)[30] and classical regression[13,28] to predict BCR; without incorporating clinicopathological features. In contrast, the PathCLR pipeline first generates latent representations then a neural network learns the non-linear relations between a combination of the latent representation of a TMA image and clinicopathological features. This allows the PathCLR model to leverage all available information, both from tissue images and clinicopathological data, to predict BCR for a specific patient. Moreover, we explored using lower resolutions, or the combination of different resolutions, but observed no enhancement in accuracy.

A recent study by Leo et al.[17] employs deep learning for the segmentation of lumen glands and subsequently extracts handcrafted features. They implement feature selection and pinpoint 6 features for BCR risk stratification. While our approach is distinct in its focus on predicting BCR for individual patients, we took inspiration from their methodology to establish a baseline model. This served as a benchmark against which we contrasted the performance of PathCLR, with the results detailed in Section 4.

The most recent work by Pinckaers et al.[24] utilized deep learning to develop a biomarker for BCR prediction using H&E-stained tissue images. They reported the odds ratio and hazard ratio for the identified group of patients using the developed biomarker. Conversely, our methodology in this investigation steers towards generating individualized BCR predictions, supplying a precise prediction for each individual patient, as opposed to the common task of identifying group biomarkers. A fundamental issue with hazard ratio predictions lies in its nature as a ratio—meaning it can be used to compare one patient to another—e.g., predict that patient A will live longer than patient B—but it does not provide a direct prediction about an individual patient. In contrast, PathCLR predicts a "Yes/No" prediction for the patient. This enables clinicians to devise more targeted treatment and follow-up strategies that are aligned with each patient's unique circumstances.

Lastly, it is worth mentioning that even though we had access to 2 datasets with prostate cancer slides, we decided to not train on one dataset and test on the other, as there are significant differences between these 2 datasets: large variations in the range of clinical feature values and slide resolutions, as well as covariate shift issues, including differences in BCR rates, stage distribution, etc. These differences mean that we do not anticipate that a model trained on one would apply to the other.

Conclusions

We introduced PathCLR, a semi-supervised machine learning approach designed to predict prostate cancer biochemical recurrence (BCR) within 5 years following radical prostatectomy surgery. PathCLR utilizes H&E-stained TMA cores and clinicopathological variables for BCR prediction. Notably, PathCLR stands out as the first approach that predicts BCR without the need for time-consuming manual nuclear or tumor annotations. Instead, it learns contrastive latent representations of H&E-stained TMA cores, contributing to its predictive capabilities.

Our results underscore that the integration of latent representations derived from TMA core images and clinicopathological attributes in PathCLR yields statistically significant improvements of at least 5% compared to using clinicopathological features alone for BCR prediction. This highlights the critical information present in diagnostic TMA images, which complements the data from clinicopathological variables and aids in predicting a patient's BCR within 5 years after surgery. This work represents the initial step towards developing an accurate personalized BCR predictor and sets the stage for future research leading to a deployed system for clinical applications.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Russell Greiner reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Russell Greiner reports financial support was provided by Alberta Machine Intelligence Institute.

Acknowledgments

We gratefully acknowledge financial support from Amii, NSERC, and the University of Alberta. We acknowledge Prostate Cancer Biorepository Network (PCBN), funded by the Prostate Cancer Research Program of the U.S. Department of Defense for providing access to the JHU dataset. Furthermore, We acknowledge Dr. Elizabeth A. Platz and Dr. Angelo M. DeMarzo for providing the JHU dataset and Andre Kajdacsy-Balla for providing access to the CPCTR dataset.

The codebase is available on GitHub.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2023.100344.

References

1. Azizi S, Culp L, Freyberg J, et al. Robust and efficient medical imaging with self-supervision. arXiv preprint arXiv:220509723 2022.
2. Bostrom PJ, Bjartell AS, Catto JW, et al. Genomic predictors of outcome in prostate cancer. Eur Urol 2015;68(6):1033–1044.
3. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol 2020;21(2):233–241.
4. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. International Conference on Machine Learning. PMLR; 2020. p. 1597–1607.
5. Ciga O, Xu T, Martel AL. Self supervised contrastive learning for digital histopathology. Mach Learn Appl 2022;7, 100198.
6. Cooperberg MR, Hilton JF, Carroll PR. The capra-s score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. Cancer 2011;117(22):5039–5046.
7. Cui M, Zhang DY. Artificial intelligence and computational pathology. Lab Investig 2021;101(4):412–422.
8. Eksi M, Evren I, Akkas F, et al. Machine learning algorithms can more efficiently predict biochemical recurrence after robot-assisted radical prostatectomy. Prostate 2021;81(12):913–920.
9. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings; 2010. p. 249–256.
10. Graham S, Vu QD, Raza SEA, et al. Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal 2019;58, 101563.
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770–778.
12. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–507.
13. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 1970;12(1):55–67.
14. Hossain A, Arimura H, Kinoshita F, et al. Automated approach for estimation of grade groups for prostate cancer based on histological image feature analysis. Prostate 2020;80(3):291–302.
15. Jaiswal A, Babu AR, Zadeh MZ, Banerjee D, Makedon F. A survey on contrastive self-supervised learning. Technologies 2020;9(1):2.
16. Kumar N, Verma R, Arora A, et al. Convolutional neural networks for prostate cancer recurrence prediction. Medical Imaging 2017: Digital Pathology. SPIE; 2017. p. 106–117.
17. Leo P, Janowczyk A, Elliott R, et al. Computer extracted gland features from H & E predicts prostate cancer recurrence comparably to a genomic companion diagnostic test: a large multi-site study. NPJ Precis Oncol 2021;5(1):35.
18. Morris MJ, Rowe SP, Gorin MA, et al. Diagnostic performance of 18f-DCFPYL-PET/CT in men with biochemically recurrent prostate cancer: results from the condor phase III, multicenter study. Clin Cancer Res 2021;27(13):3674–3682.
19. Nagpal K, Foote D, Tan F, et al. Development and validation of a deep learning algorithm for gleason grading of prostate cancer from biopsy specimens. JAMA Oncol 2020;6(9):1372–1380.
20. Ouali Y, Hudelot C, Tami M. An overview of deep semi-supervised learning. arXiv preprint arXiv:200605278 2020; abs/2006.05278. https://api.semanticscholar.org/CorpusID:219558952.
21. Patel AA, Kajdacsy-Balla A, Berman JJ, et al. The development of common data elements for a multi-institute prostate cancer tissue bank: the cooperative prostate cancer tissue resource (CPCTR) experience. BMC Cancer 2005;5(1):1-14.
22. Pernar CH, Ebot EM, Wilson KM, Mucci LA. The epidemiology of prostate cancer. Cold Spring Harbo Perspect Med 2018;8(12), a030361.
23. Pienta KJ, Gorin MA, Rowe SP, et al. A phase 2/3 prospective multicenter study of the diagnostic accuracy of prostate specific membrane antigen PET/CT with 18F-DCFPYL in prostate cancer patients (Osprey). J Urol 2021;206(1):52–61.
24. Pinckaers H, van Ipenburg J, Melamed J, et al. Predicting biochemical recurrence of prostate cancer with artificial intelligence. Commun Med 2022;2(1):1–9.
25. Punnen S, Freedland SJ, Presti Jr JC, et al. Multi-institutional validation of the capra-s score to predict disease recurrence and mortality after radical prostatectomy. Eur Urol 2014;65(6):1171–1177.
26. Ross AE, D'amico A, Freedland S. Which, when and why? rational use of tissue-based molecular testing in localized prostate cancer. Prostate Cancer Prostatic Dis 2016;19(1):1–6.
27. Ross AE, Johnson MH, Yousefi K, et al. Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate-and high-risk men. Eur Urol 2016;69(1):157–165.
28. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (Methodol) 1996;58(1):267–288.
29. Tourinho-Barbosa R, Srougi V, Nunes-Silva I, et al. Biochemical recurrence after radical prostatectomy: what does it mean? Int Braz J Urol 2018;44:14–21.
30. Vapnik V. *The Nature of Statistical Learning Theory*. Springer Science & Business Media. 1999.
31. Verma R, Kumar N, Patil A, Kurian NC, Rane S, Sethi A. Multi-organ nuclei segmentation and classification challenge 2020. IEEE Trans Med Imaging 2020;39(1380–1391):8.
32. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005. J Natl Cancer Inst 2009;101(19):1325–1329.
33. Yamamoto Y, Tsuzuki T, Akatsuka J, et al. Automated acquisition of explainable knowledge from unannotated histopathology images. Nat Commun 2019;10(1):1–9.
34. Yan J, Chen H, Li X, Yao J. Deep contrastive learning based tissue clustering for annotation-free histopathology image analysis. Comput Med Imaging Graphics 2022;97, 102053.