

RESEARCH

Open Access



Development of a deep learning system for predicting biochemical recurrence in prostate cancer

Lu Cao¹, Ruimin He¹, Ao Zhang², Lingmei Li¹, Wenfeng Cao^{1*}, Ning Liu^{3*} and Peisen Zhang^{2*}

Abstract

Background Biochemical recurrence (BCR) occurs in 20%–40% of men with prostate cancer (PCa) who undergo radical prostatectomy. Predicting which patients will experience BCR in advance helps in formulating more targeted prostatectomy procedures. However, current preoperative recurrence prediction mainly relies on the use of the Gleason grading system, which omits within-grade morphological patterns and subtle histopathological features, leaving a significant amount of prognostic potential unexplored.

Methods We collected and selected a total of 1585 prostate biopsy images with tumor regions from 317 patients (5 Whole Slide Images per patient) to develop a deep learning system for predicting BCR of PCa before prostatectomy. The Inception_v3 neural network was employed to train and test models developed from patch-level images. The multiple instance learning method was used to extract whole slide image-level features. Finally, patient-level artificial intelligence models were developed by integrating deep learning -generated pathology features with several machine learning algorithms.

Results The BCR prediction system demonstrated great performance in the testing cohort (AUC = 0.911, 95% Confidence Interval: 0.840–0.982) and showed the potential to produce favorable clinical benefits according to Decision Curve Analyses. Increasing the number of WSIs for each patient improves the performance of the prediction system. Additionally, the study explores the correlation between deep learning -generated features and pathological findings, emphasizing the interpretative potential of artificial intelligence models in pathology.

Conclusions Deep learning system can use biopsy samples to predict the risk of BCR in PCa, thereby formulating targeted treatment strategies.

Keywords Prostate cancer biopsy, Biochemical recurrence, Deep learning, Multi instance learning

*Correspondence:

Wenfeng Cao
caowenfeng2017@163.com

Ning Liu
cningning@sina.com

Peisen Zhang
zpsbit@foxmail.com

¹ Department of Pathology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Breast Cancer Prevention and Therapy, Tianjin Medical University, Ministry of Education, Tianjin 300060, China

² Tianjin University of Science and Technology, Tianjin 300222, China

³ Department of Pathology, Tianjin Baodi Hospital, Tianjin 301899, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

According to the Global Cancer Statistics 2020, prostate cancer (PCa) is a common malignancy among men [1]. A majority of these patients undergo radical prostatectomy, either as their initial treatment choice or after a period of active surveillance. Prostate-specific antigen (PSA) is a protein produced by both cancerous and noncancerous tissue in the prostate. Its concentration in the blood is used to judge the presence of PCa. In a successful prostatectomy surgery, PSA concentration will mostly be undetectable (<0.1 ng/mL) after 2–6 weeks.

However, in 20%–40% of these patients, PSA levels will rise again after surgery, indicating biochemical recurrence (BCR) and suggesting the regrowth of PCa cells [2–4]. BCR is a strong risk factor for subsequent metastases and mortality. The accurate prediction of patients prone to experiencing BCR before prostatectomy is crucial for determining the pre-surgery course of action. For example, more aggressive treatment options, such as additional chemotherapy, radiotherapy, hormone therapy, immunotherapy and prophylactic extended pelvic lymphadenectomy, should be considered for patients with a high risk of BCR.

Prostate multi-core needle biopsy is the most reliable diagnostic method for patients suspected of having PCa, and is also one of the standard procedures before prostatectomy [5]. A systematic review showed that the cancer detection rate is associated with the number of cores [6, 7]. Histopathological grading of prostate biopsy, along with digital rectal examination and PSA level, forms the basis of most preoperative prediction systems used in clinics [8]. These systems can effectively estimate PCa progression, but have repeatedly been shown to have suboptimal prognostic and discriminatory performance, which is partly due to the subjective and non-specific nature of the core variables [9, 10]. For instance, Gleason grading was developed in the 1960s and has suboptimal interobserver reproducibility even among expert urologic pathologists [11, 12].

In recent years, whole slide image (WSI), generated by digitizing glass slide with histology samples, has seen a rise in popularity. The development of cancer always involves changes in cellular morphology and the micro-environment [13], therefore, there is a consensus among pathologists that WSI contains abundant cancer-related information [14, 15]. Additionally, deep learning (DL) based on convolutional neural network (CNN) has demonstrated excellent capabilities to obtain cancer-related information from WSI [16]. DL has been widely applied in prostate pathology, including cancer detection [17, 18], Gleason grading [19], genomic signatures prediction [20, 21], and post-surgery BCR prediction [22–24].

In this study, we developed a pre-surgery BCR prediction system based on DL and multiple instance learning (MIL) framework, using prostate multi-core needle biopsy. The system demonstrated excellent performance on the testing dataset, providing evidence for the potential contribution of AI in medical diagnosis.

Methods

Data preparation

This study incorporated two independent cohorts of patients who underwent multi-core needle biopsy diagnosis prior to prostatectomy for clinically localized PCa between January 1, 2018 and December 31, 2020. All the resources were comprehensively characterized, including patients' clinical and pathological data. None of the patients had received any preoperative treatment. All patients were investigated and followed up for a period ranging from 47 to 83 months until November 30, 2024. A total of 3092 hematoxylin and eosin (H&E) stained slides from 342 PCa patients were scanned by KFBIO PRO400 (Ningbo Konfoong Bioinformation Company, Zhejiang, China, $0.25\mu\text{m}/\text{pixel}$ at $40\times$ magnification and $0.50\mu\text{m}/\text{pixel}$ at $20\times$ magnification) at a resolution of $20\times$, sourced from Tianjin Medical University Cancer Institute and Hospital (TMUCH) and Tianjin Baodi Hospital (TBH). All patients underwent 10–16 cores needle biopsy according to the hospital guidelines. Pathologists, blinded to the diagnosis, reviewed these slide images and selected 5 WSIs for each patient. The selection criteria required that all 5 WSIs for each patient contain PCa tissue, and the WSIs with higher Gleason scores were preferentially selected. Thus, 1585 WSIs from 317 patients were selected for model training and testing. The digital pathology slides and clinical information of 254 patients treated at TMUCH were used as the training cohort, and 63 patients treated at TBH were used as the testing cohort in this study. This study was approved by the Ethics Committee of TMUCH (No. Ek2020074) and was conducted in accordance with the 1964 Helsinki Declaration and its subsequent amendments, or comparable ethical standards. Our Ethics Committees granted a waiver of informed consent.

The clinical information of patients, including age, PSA values, primary and secondary Gleason scores, was collected from electronic surgical pathology reports. BCR was defined as the detection of two consecutive postoperative PSA values equal to or higher than 0.2 ng/mL. Meanwhile, other patients were confirmed to have non-BCR through telephone follow-up. The baseline data of patients before the radical prostatectomy period are shown in Table 1.

Table 1 Baseline data of the training cohort and testing cohort

	Training cohort (TMUCH)		Testing cohort (TBH)		Total
	Recurrence cases	No events cases	Recurrence cases	No events cases	
Patients, n	89	165	22	41	317
WSIs, n	445	825	110	205	1585
Age, mean (SD)	67.30(± 7.18)	68.08(± 7.41)	67.91(± 5.80)	67.44(± 7.43)	67.77(± 7.25)
PSA (ng/ml), mean (SD)	7.16(± 3.40)	3.34(± 3.41)	6.79(± 3.56)	3.75(± 3.54)	4.70(± 3.88)
GS = 3 + 3, n	4	46	1	11	62
GS = 3 + 4, n	2	13	2	4	21
GS = 4 + 3, n	2	4	1	1	8
GS = 4 + 4, n	42	62	13	17	134
GS = 4 + 5, n	22	15	2	4	43
GS = 5 + 4, n	2	5	0	0	7
GS = 5 + 5, n	15	20	3	4	42
pT2, n	31	119	12	28	190
pT3a, n	33	34	3	11	81
pT3b, n	25	12	7	2	46

SD Standard deviation, GS Gleason scores, pT Pathology tumor stage

Overview of the BCR prediction system

The workflow of the BCR prediction system is shown in Fig. 1. The system is implemented across three stages, with each stage mapping to BCR prediction at the patch-level, WSI-level, and patient-level. In stage 1, cropped patches from WSIs are input into a pre-trained CNN model to predict the recurrence probability. Thus, this stage outputs the patch-level prediction. In stage 2, a Patch Likelihood Histogram (PALHI) pipeline and a Bag of Words (BoW) pipeline are used to extract the WSI-level features of all patches' prediction probabilities in each WSI [25]. Meanwhile, the model based on PALHI and BoW methods can predict the recurrence probability of each WSI, thus this stage can output the WSI-level prediction. In stage 3, due to each patient involves 5 WSIs, patient-level features are generated by aggregating WSI-level features of each patient using pooling operation. These features are then combined with clinical characteristics and inputted into machine learning (ML) classifiers to predict patient-level recurrence risk.

Data processing

To address the challenge of handling large-scale digital images, we implemented a systematic pre-processing strategy. First, the WSIs were divided into 512×512 patches using a non-overlapping partitioning approach, strictly maintaining a resolution of 0.5 μm/pixel. Then, to ensure high-quality data, we removed the background images according to the pixel threshold and bright threshold. In addition, we used the Vahadane method to normalize the color of these patches. During the training

process, we applied data augmentations, including random horizontal and vertical flipping, and for the testing process, we only used color normalization.

Deep learning training

Proposed DL process comprises three levels of predictions: patch-level, WSI-level and patient-level predictions, using CNN and MIL.

For patch-level predictions, we implemented transfer learning to enhance the model's generalization across heterogeneous cohorts. This involved initializing the model's parameters with pretrained weights from the ImageNet dataset. Afterward, we fine-tuned the entire model using the training cohort dataset (254 samples) which has been annotated at the patient level, since the WSI area related to BCR in PCa could not be identified. Utilizing transfer learning, we evaluated the efficacy of various DL models, including Inception_v3, ResNet50, VGG19 and ResNet18, as shown in Table 2. The Inception_v3 demonstrated superior performance in terms of the area under the receiver operating characteristic curve (AUC) and accuracy, thus becoming the model selected in this study.

To enhance generalization, in this study, we employed the cosine decay learning rate algorithm to set the learning rate. The learning rate is presented as follows:

$$\eta_t^{\text{task-spec}} = \eta_{\min}^i + \frac{1}{2} \left(\eta_{\max}^i - \eta_{\min}^i \right) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_i} \pi \right) \right)$$

$\eta_{\max}^i = 0.01$, $\eta_{\min}^i = 0$, $T_i = 50$ represent the maximum learning rate, the minimum learning rate, and the number of iteration epochs, respectively. For transfer learning

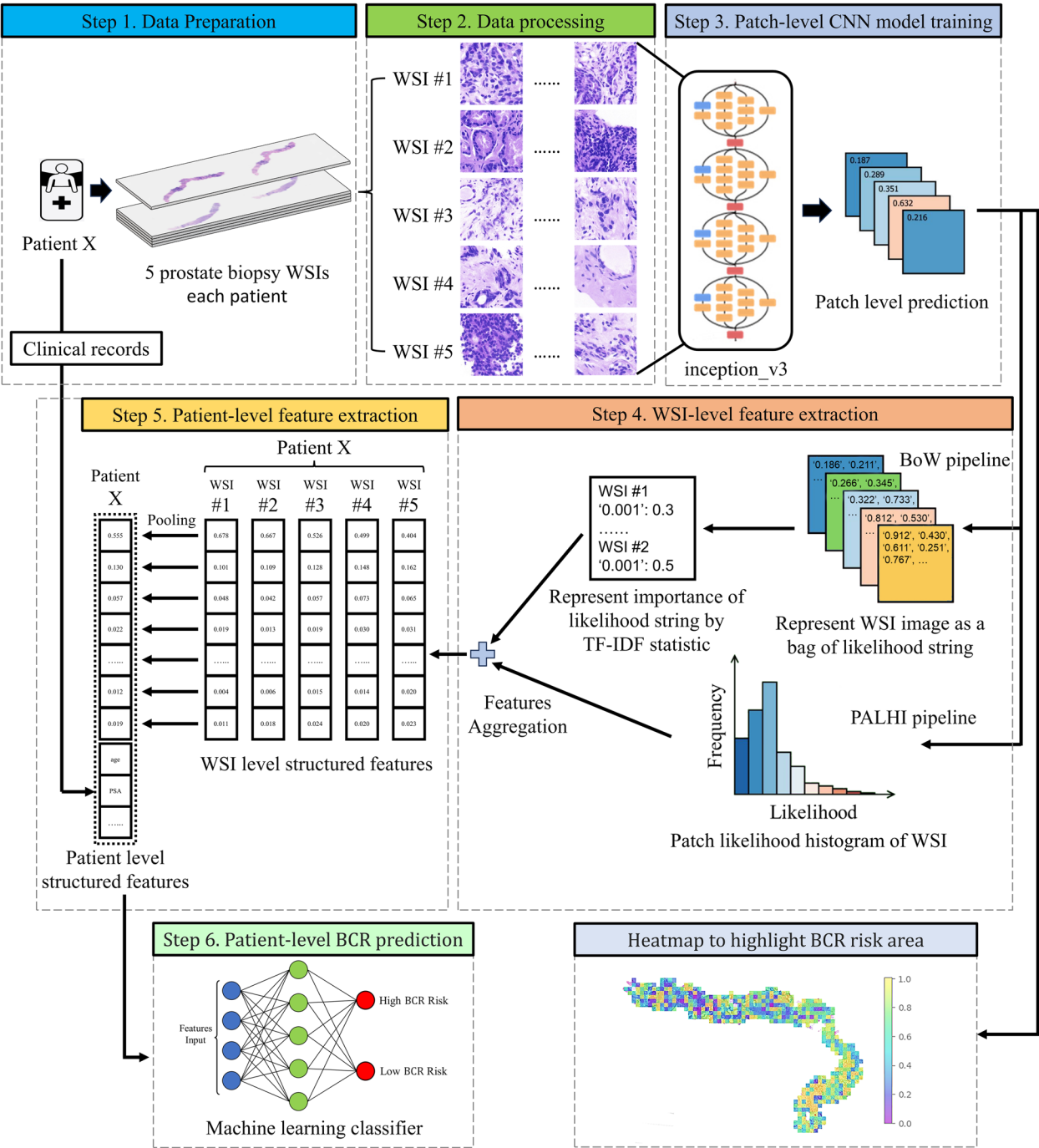


Fig. 1 Illustration of the development and usage of the BCR prediction system. Steps 1 and 2, WSIs of the patients were cropped as tile-like patches. Step 3, a CNN model was trained to output the recurrence probability of each patch. Its parameters, pretrained from the ImageNet dataset, were fine-tuned using the training cohort dataset (1270 WSIs). Step 4, patches' prediction probabilities were aggregated by PALHI and BoW pipeline to output structured data of WSI-level features. Step 5, when the patient had multiple WSIs, the patient-level features were extracted by pooling corresponding WSI-level features. Step 6, clinical features and patient-level features were input into ML classifiers to predict the recurrence risk of the patient. Steps 4–6 form the MIL model of this study

Table 2 Patch-level performances of the BCR prediction system

Model	AUC	95% CI	Accuracy	Sensitivity	Specificity
inception_v3	0.803	0.8014–0.8053	0.732	0.663	0.78
resnet50	0.802	0.8002–0.8041	0.727	0.685	0.757
vgg19	0.789	0.7869–0.7910	0.725	0.645	0.779
resnet18	0.78	0.7775–0.7816	0.708	0.653	0.745

CI confidence interval

algorithms, fine-tuning is essential as the backbone component already contains pre-trained parameters. Therefore, we fine-tuned the backbone component parameters when $T_{cur} = \frac{1}{2}T_i$. Furthermore, the learning rate of the backbone component is defined as follows:

$$\eta_t^{backbone} = \begin{cases} 0 & \text{if } T_{cur} \leq \frac{1}{2}T_i \\ \eta_{min}^i + \frac{1}{2}(\eta_{max}^i - \eta_{min}^i) \left(1 + \cos\left(\frac{T_{cur}}{T_i}\pi\right)\right) & \text{if } T_{cur} > \frac{1}{2}T_i \end{cases}$$

Other hyperparameter configurations are as follows: Optimizer—SGD, Loss function—SoftMax cross-entropy, with a batch size of 32.

Multi-instance learning process

After DL training, we used the trained Inception_v3 model to carry out label predictions for all patches and obtained the corresponding probabilities for each patch. Afterward, we used the MIL method to aggregate the probability of all patches for each WSI as the feature vector of corresponding WSI. The MIL method used in this study was refined from the commonly used methods in pathological image analysis [25]. It was composed of the PALHI pipeline and the BoW pipeline. The PALHI pipeline used histogram to represent the distribution of patch probabilities in the WSI. The distribution of all patch probabilities in different intervals of the histogram constitutes the feature vector of the corresponding WSI. The BoW pipeline used TF-IDF variable mapping for each patch, and these variables constituted a TF-IDF feature vector to represent corresponding WSI [26]. Term frequency-inverse document frequency (TF-IDF) is a statistical method commonly used in information retrieval, combining the term frequency (TF) and the inverse document frequency (IDF). The TF measures the frequency of each patch's feature in a single WSI, and the IDF assigns a weight to the feature based on the rarity of each feature across the entire WSIs. By multiplying the TF and IDF values, TF-IDF assigns a higher weight to patches that are both frequent within a particular WSI (high TF) and relatively rare across the whole set of WSIs (high IDF), thus effectively identifying the significance

of features by quantifying their importance based on their frequency within and across WSIs. Feature vectors obtained by PALHI and the BoW pipelines were then spliced together and used for training classifiers to predict the label in each WSI.

Through the deployment of PALHI and BoW pipelines, we integrated the initially scattered patch-level predictions and subsequently obtained WSI-level features, which were used for subsequent analysis operations, including t-distributed stochastic neighbor embedding (t-SNE) features projection and training ML classifiers.

Signature building

After constructing WSI-level features using patch-level predictions, probability histograms, and TF-IDF features in combination, we aggregated them into patient-level features through pooling operation. Consequently, final patient representations were formed by integrating patient-level pathology features with clinical features.

For feature selection, initially, all feature lines were standardized with the z-score standardization method. The z-score standardization method is a technique that transforms data points in a dataset to have a mean of 0 and a standard deviation of 1, and its formula is $z = \frac{x-\mu}{\sigma}$ where x is the original data point, μ is the mean of the dataset, and σ is the standard deviation of the dataset.

Next, Pearson's rank correlation coefficient was employed to compute the correlation among features. When there is a monotonically increasing or monotonically decreasing relationship between two features, this coefficient can quantify the closeness of such a relationship. Among the features, if the correlation coefficient between any two features was greater than 0.9, only one of these features was retained. Finally, the least absolute shrinkage and selection operator (LASSO) regression was used for feature selection, since it can shrink the coefficients of some unimportant features to zero, thereby achieving the purpose of variable selection and reducing the risk of model overfitting.

Model evaluation

After signature building, we utilized machine learning algorithms, including multilayer perceptron (MLP), logistic regression (LR), support vector machine (SVM) and Random Forest, to develop classifiers. The MLP model was a fully connected 3-layer perceptron, comprising 128, 64, and 32 hidden nodes, respectively. The SVM model used the radial basis function kernel function, while the other parameters were kept as default. The Random Forest model set the value of $n_{estimators}$ to 10. All of these models employed the implementation of scikit-learn. Scikit-learn is a popular Python library

and integrates various ML algorithms, allowing users to directly call these algorithms for different ML tasks [27].

The receiver operating characteristic (ROC) curve was employed to validate the performance of the Inception_v3 model in region identification at the patch-level. Probability heatmaps were used for WSI-level visual evaluation after the MIL process. For the BCR prediction model, we employed the AUC value as the performance metric, along with accuracy, sensitivity and specificity calculations. Its clinical practicability was evaluated using decision curve analysis (DCA).

Results

Performance evaluation and visualization

The performance of the BCR prediction system was evaluated at the patch-level, WSI-level, and patient-level using ROC curves. In the patch-level prediction, the AUC for the Inception_v3 architecture was 0.968 for the training dataset and 0.803 for the testing dataset, as shown in Fig. 2A. In the WSI-level prediction, all ML classifiers demonstrated improved performance on the testing dataset compared to the patch-level prediction, as shown in the ROC curves, and the RandomForest classifier achieved a higher AUC value of 0.848, as shown in Fig. 2B. In the patient-level prediction, these classifiers' performances were further improved after both average pooling and maximum pooling feature aggregation, and the average pooling operation displayed better performance, yielding AUC values of 0.908 with MLP and LR classifiers, as shown in Fig. 2C and D. These results demonstrated the considerable effectiveness of our feature aggregation approach.

Probability maps were utilized to assess the patch-level prediction outcomes, as shown in Fig. 3A. It could be observed that, compared to non-BCR patients, the WSIs of BCR cases exhibited a greater number of patches with higher probability values approaching 1. Thus, the MIL model combining BoW with PALHI method was utilized in this study. Another reason for choosing the MIL method is that it does not require manual pixel-level annotation, which is crucial for building a BCR prediction model, because even experienced pathologists cannot accurately match the pathological morphology of H&E slices with BCR. Gradient-weighted Class Activation Mapping (Grad-CAM) is a technique that creates maps to display the localization of different classes by visualizing the gradients entering the final convolutional layer of a neural network. Figure 3B shows the utilization of Grad-CAM in displaying the activation of the last convolutional layer for BCR prediction evaluation. This visualization highlighted the regions of the input image that significantly contributed to the prediction, offering valuable insights into the model's decision-making process.

To comprehend the enhancement in performance at the patient-level prediction, we employed the t-SNE algorithm. WSI-level features were extracted from patches' prediction probabilities by PALHI and BoW pipelines, and their t-SNE projections were shown in Fig. 4A. The inter-class distance and intra-class distance were calculated to quantitatively describe the changes in features from WSI-level to patient-level, as shown in Fig. 4B. The t-SNE projection of patient-level features which were aggregated by maximum pooling was shown in Fig. 4C, and the t-SNE projection of patient-level features which were aggregated by average pooling was shown in Fig. 4D. Clear differentiation between BCR and non-BCR cases was observed in both WSI and patient level t-SNE projections. It indicated a significant decrease in the intra-class distance of both BCR and non-BCR cases after aggregating features with pooling operations (decrease from 25.94 & 32.84 to 9.47 & 13.26), which correlated with higher AUC values at the patient level. In addition, the inter-class distance of the average pooling projection was larger than that of the maximum pooling (35.64 vs. 34.23), which correlated with the higher AUC value for average pooling. Therefore, the average pooling operation was employed in our BCR prediction system.

The impact of WSIs quantity on prediction performance

The impact of the number of WSI per patient on the efficacy of models was evaluated and shown in Fig. 4E. For models utilizing MLP, LR, and SVM classifiers, their AUC values increased as the number of WSIs per patient increased, and for all classifiers, the maximum AUC value was achieved when all WSIs of each patient were involved in training. To further explain this phenomenon, Table 3 listed the AUC, accuracy, sensitivity, and specificity of these models on the testing cohort. It can be observed that when selecting one WSI per patient for training, the MIL model exhibited higher accuracy and specificity values, and when using multi-WSIs per patient for training, the MIL model exhibited higher AUC and sensitivity values. For each patient, it seemed that several WSIs contained more crucial information and are assigned higher weights in the final decision. It also indicated that using all WSIs per patient for training led to the model achieving the highest AUC and moderate levels of accuracy, sensitivity, and specificity. Therefore, we concluded that increasing the number of WSIs for each patient can improve the generalization performance of the MIL model, and features aggregated from all 5 WSIs of each patient were used for model training.

The impact of clinical features on prediction performance

The ROC curves of various classifiers trained using pathology images and clinical features on the testing

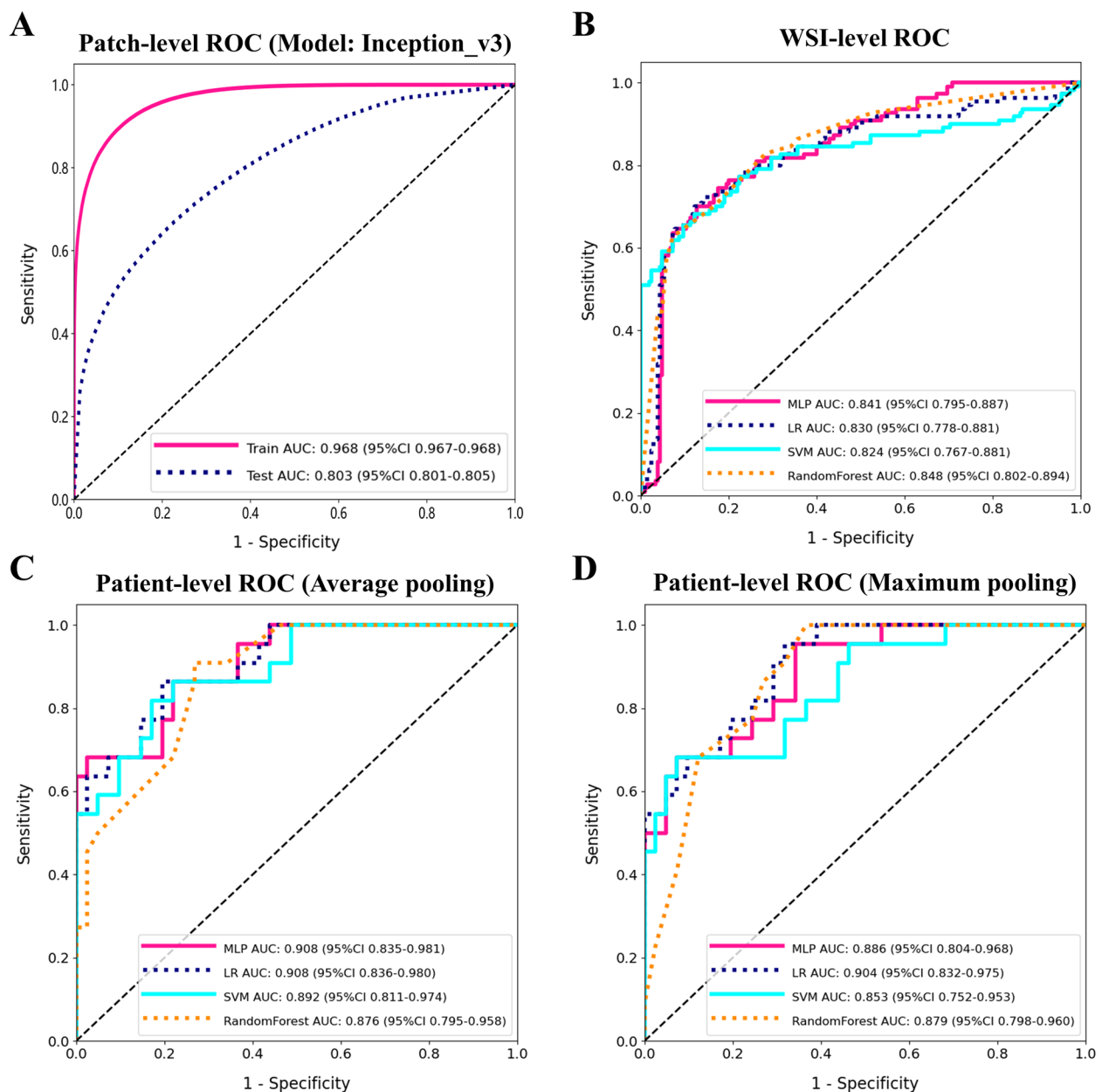


Fig. 2 BCR prediction system efficacy at the patch-level, WSI-level and patient-level. **A** the patch-level ROC curves in the training and testing cohorts. Inception_v3 architecture was selected to build patch prediction model. **B** the WSI-level ROC curves in the testing cohort. **C** the patient-level ROC curves in testing cohort. The patient-level features were aggregated by average pooling operation. **D** the patient-level ROC curves in the testing cohort. The patient-level features were aggregated by maximum pooling operation

cohort were shown in Fig. 5A. The clinical information included patient age, PSA value, primary Gleason score, and secondary Gleason score. The ROC curves of classifiers trained by clinical information were shown in Fig. 5B. It could be observed that, compared to clinical features, the pathological image features extracted using CNN and MIL methods could significantly enhance the model efficacy. The MLP classifier

trained on combined pathological and clinical features achieved the highest AUC value in this study, reaching 0.911(95%CI: 0.840–0.982). The corresponding values of accuracy, sensitivity, specificity and F1-score for various classifiers were shown in Table 4. For all classifiers trained by pathological and clinical features, decision curve analyses demonstrated good clinical benefits, as shown in Fig. 5C-E.

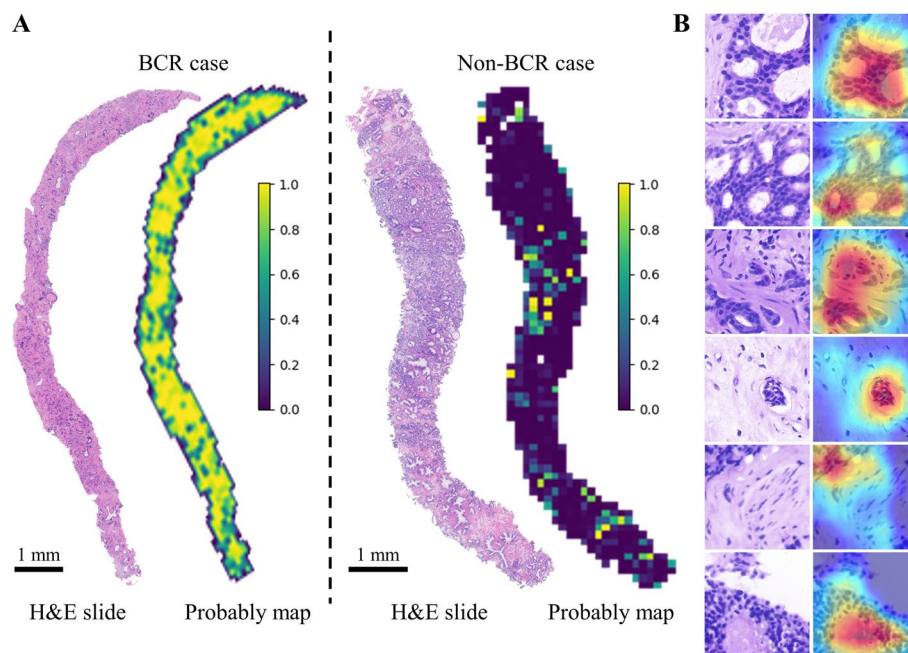


Fig. 3 Visualization of BCR prediction results. **A** H&E slides and probability maps for BCR and non-BCR cases. Color bars show the BCR probability of each patch. **B** The utilization of Grad-CAM in displaying the activation of the last convolutional layer for patch-level prediction model

Discussion

PCa is the leading cause of cancer-associated disability due to the negative effects of over-treatment and under-treatment, and it is also a major cause of cancer death in men [28, 29]. Radical prostatectomy is pivotal in the treatment of PCa, and its procedure directly affects the prognosis of patients. The BCR of PCa indicates tumor progression and serves as a crucial basis for formulating treatment procedures. This study has developed a preoperative BCR prediction system for PCa, aiming to provide valuable reference for guiding the course of radical prostatectomy procedures.

Since it is difficult to annotate regions or patches related to BCR on pathological slides, we assigned each WSI a single overarching label, instead of manually annotating each region or patch of a slide, which means the cropped patches of each WSI shared the label of the corresponding WSI. The DL models then use WSI-level annotation information to identify regions of interest or use this to classify the disease state of the slide. This approach, combining DL and the MIL pipelines, has demonstrated promising performance in tumor region identification [30], Gleason scoring [5], tumor purity prediction [31], and morphological feature segmentation [32] of PCa. Thus, we constructed a BCR prediction system for PCa biopsy tissue based on the DL and the MIL models, and demonstrated enhanced performance compared to current systems.

Several BCR prediction systems for PCa have been developed, and most of them are trained using clinical variables, radiological variables [33–43], or macroscopic histological variables [44–55] such as Gleason score, quantitative nuclear grade, and seminal vesical invasion. With the development of AI, DL models that can obtain microscopic information and extract high-dimensional features have shown excellent performance in many medical tasks. Eminaga et al. [56] and Pinckaers et al. [22] trained DL models using H&E-stained tissue microarrays (TMA) to predict BCR. Eminaga et al. constructed the DL model using the PlexusNET and grid search methods, and the AUC on the testing cohort was 0.71 (95% CI: 0.67–0.75). Pinckaers et al. constructed the DL model based on Resnet50 backbone, and hazard ratios in univariate and multivariate analyses were 5.78 (95% CI: 2.44–13.72; $p < 0.005$) and 3.02 (95% CI: 1.10–8.29; $p = 0.03$), respectively. Huang et al. [57] trained a DL model using WSIs of radical prostatectomy specimens to predict BCR, and the AUC on the testing cohort was 0.78. Our system trained a DL model based on the InceptionV3 backbone using WSIs of preoperative biopsy tissue, and its AUC value at WSI level in the testing cohort was observed to reach 0.848 (95% CI: 0.802–0.894).

MIL is able to improve the performance of DL models [58, 59]. MIL is a machine learning paradigm where training data is grouped into bags, and the labels of the bags are determined by the labels of their instances.

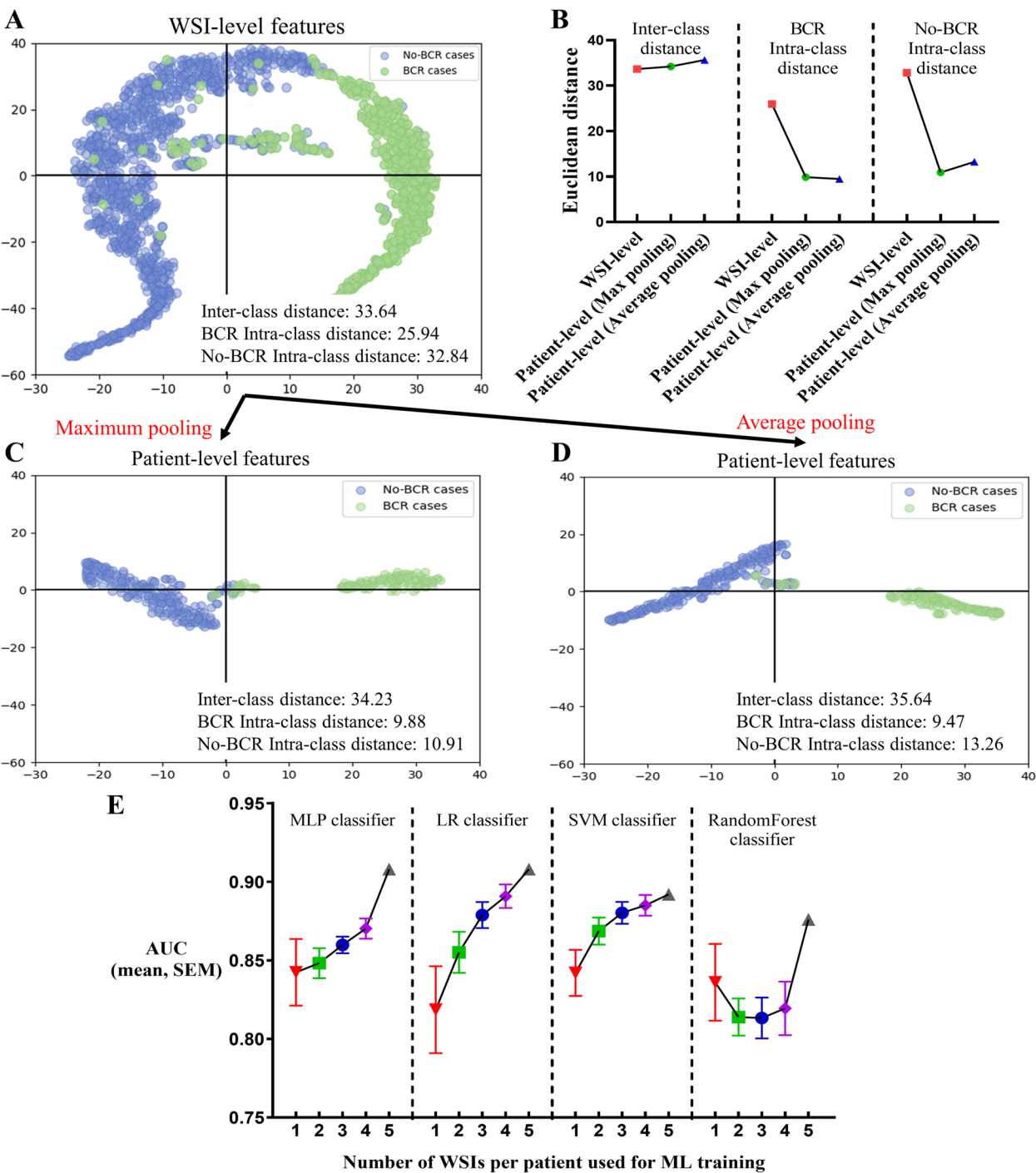


Fig. 4 Two-dimensional t-SNE projection of WSI-level and patient-level feature space. **A** t-SNE Visualization of WSI-level features. Original WSI-level features are extracted using the MIL model. **B** The inter-class and intra-class distances of BCR and No-BCR cases were calculated using two-dimensional t-SNE projection to evaluate the distribution of feature space. **C** t-SNE Visualization of Patient-level features which were aggregated from WSI-level features using average pooling. **D** t-SNE Visualization of Patient-level features which were aggregated from WSI-level features using maximum pooling. **E** The efficacy of models trained with different numbers of WSIs per patient has been evaluated by the AUC values of the testing cohort. Patient-level features are aggregated from WSI-level features using average pooling, and used for training classifiers (MLP, LR, SVM and RandomForest). All combinations of various numbers of WSIs for each patient are traversed. SEM: Standard Error of the Mean

Table 3 The performance of models trained with varying numbers of WSIs per patient

Models	Number of WSIs	AUC (mean, SEM)	Accuracy (mean, SEM)	Sensitivity (mean, SEM)	Specificity (mean, SEM)
MLP	1	0.842(± 0.021)	0.800(± 0.009)	0.718(± 0.049)	0.844(± 0.031)
	2	0.848(± 0.010)	0.787(± 0.011)	0.777(± 0.038)	0.793(± 0.032)
	3	0.860(± 0.005)	0.784(± 0.009)	0.837(± 0.030)	0.756(± 0.025)
	4	0.870(± 0.006)	0.775(± 0.008)	0.855(± 0.027)	0.732(± 0.026)
	5	0.908	0.857	0.636	0.976
LR	1	0.8186(± 0.028)	0.822(± 0.009)	0.6998(± 0.060)	0.8878(± 0.036)
	2	0.8571(± 0.014)	0.814(± 0.006)	0.7773(± 0.032)	0.8341(± 0.021)
	3	0.8789(± 0.008)	0.811(± 0.009)	0.8319(± 0.030)	0.8001(± 0.025)
	4	0.8914(± 0.008)	0.794(± 0.009)	0.8818(± 0.027)	0.7462(± 0.026)
	5	0.908	0.81	0.818	0.805
SVM	1	0.842(± 0.015)	0.816(± 0.015)	0.709(± 0.045)	0.873(± 0.041)
	2	0.869(± 0.008)	0.802(± 0.010)	0.800(± 0.035)	0.802(± 0.030)
	3	0.880(± 0.007)	0.797(± 0.011)	0.759(± 0.029)	0.817(± 0.029)
	4	0.885(± 0.007)	0.797(± 0.018)	0.773(± 0.056)	0.810(± 0.057)
	5	0.892	0.81	0.773	0.829
Random Forest	1	0.836(± 0.024)	0.794(± 0.021)	0.673(± 0.033)	0.859(± 0.034)
	2	0.814(± 0.012)	0.732(± 0.015)	0.773(± 0.032)	0.710(± 0.032)
	3	0.813(± 0.013)	0.740(± 0.009)	0.741(± 0.039)	0.739(± 0.025)
	4	0.819(± 0.017)	0.749(± 0.023)	0.736(± 0.049)	0.756(± 0.051)
	5	0.876	0.778	0.864	0.732

SEM Standard error of the mean

Although MIL is widely adopted in computer vision, its utilization in prostate histology images remains scarce. The main premise of most weakly-supervised methods for histological analysis requires pooling the features from WSI patches, under the MIL framework [31]. In this study, we combined MIL with histogram methods to extract prediction probabilities features of all patches in each WSI as the WSI-level features, and then, we aggregated the WSI-level features to obtain the patient-level features using pooling operations. Each step here improved the performance of the system. Through t-SNE dimensionality reduction, we found that the reason why the AUC value after the average pooling operation was higher than after the maximum pooling operation was that average pooling generated larger inter-class distances. This could be because the average pooling method better represents the overall trend of BCR in patients, hence exhibiting superior performance.

The pathology diagnosis of PCa needle biopsy is the gold standard for confirming PCa before radical prostatectomy. Increasing the core number of prostate biopsies

can enhance the cancer detection rate, and related studies have suggested an optimal number of 10–12 cores [60, 61]. In this study, experienced pathologists excluded WSI without cancerous regions and then selected five WSIs for each patient. We examined the impact of the number of WSIs on BCR prediction and observed that increasing the number of WSIs improved the overall performance of the model, a trend also reported in other literatures [62, 63]. Further research revealed that this was because the features of individual WSIs for each patient exhibited higher weights in the model, suggesting that these WSIs contained information more closely related to the possibility of recurrence. The phenomenon of several high-weight WSIs did not become diluted with the increase in the number of WSIs in our system. Therefore, we believe that increasing the number of WSIs for each patient can improve the generalization performance of BCR prediction models for PCa.

We performed visualization of high-risk BCR areas in prostate biopsy tissue slides, and found DL features were significantly correlated with pathological findings, indicating the interpretability of DL models based on WSI. Using

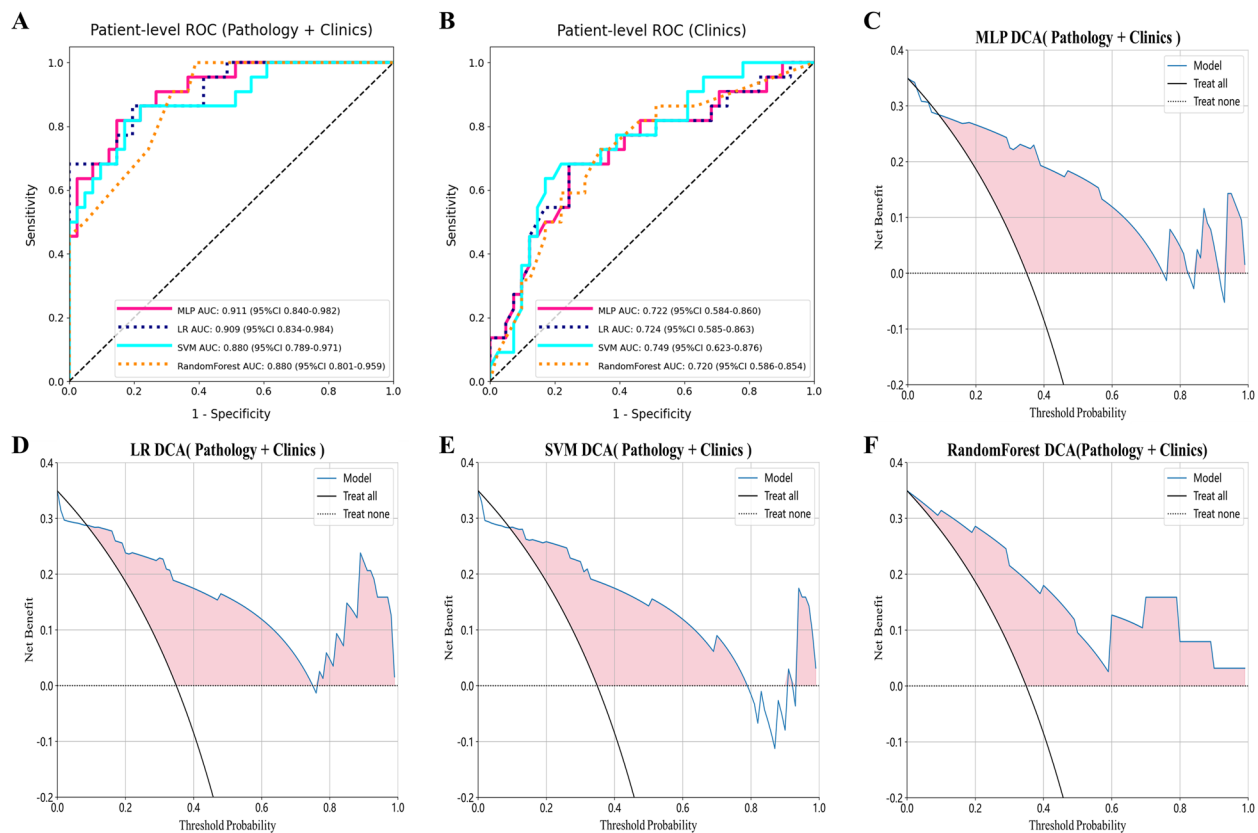


Fig. 5 BCR prediction system evaluation. **A** the ROC curves for the BCR prediction system trained with pathological and clinical features. **B** the ROC curves for the BCR prediction system trained with clinical features. **C–F** Decision curve analyses (DCA) among different machine learning methods in the testing cohort of the BCR prediction system trained with pathological and clinical features. **C** MLP method. **D** LR method. **E** SVM method. **F** RandomForest method

the Grad-CAM-based images, we observed that in the tumor region, the sieve arrangement area and suspicious vascular gap infiltration area indicated a higher likelihood of BCR, consistent with the current consensus [64–67]. Areas of neural invasion, necrosis, and sharp-angle glandular areas also indicated a higher likelihood of BCR. In the non-tumor areas, gland morphologies differed in non-cancer regions between patients who experienced recurrence and those who did not. Recent studies based on prostate digital pathology have reported similar findings [68, 69].

It should be clarified that the proposed system refrained from using the tumor-stage (T-stage) information as system parameters in view of the following considerations, notwithstanding the fact that it had been collected beforehand. Foremost, we employed clinical indices that could be obtained before radical prostatectomy as system parameters so that the proposed system could predict the BCR risk immediately after obtaining either preoperative biopsy WSI or postoperative WSI. However, in many cases, the pathology tumor stage (pT-stage) is reported after radical prostatectomy has been

completed. Secondly, we preferred to select the clinical variables that could be automatically obtained or generated as the system inputs, such as the Gleason score which can be provided by the PAIGE Prostate AI product after scanning the H&E-stained slides. However, the T-stage information still relies on surgeons' provision, which restricts the system's independent diagnostic capability. Furthermore, we also attempted to construct a system with T-stage information but it failed to improve the performance (Supplementary Table 1). Initially, the clinical tumor stage (cT-stage) was collected, but the preponderance of cases presented with a T2 stage, rendering it statistically inconsequential. Subsequently, the more accurate pT-stage was used, but it was excluded during feature selection process, which was carried out using Pearson's correlation coefficient and LASSO (Supplementary Fig. 1). Therefore, after weighing the improvement of system efficacy brought by the T-stage (cT and pT) information against the costs that the system has to bear, we did not include the T stage as a system input parameter.

Table 4 The performance of the system combining pathological and clinical features

Model	Group	Accuracy	AUC	Sensitivity	Specificity
MLP	Pathology + Clinics	0.825	0.911	0.773	0.854
	Pathology	0.857	0.908	0.636	0.976
	Clinics	0.714	0.722	0.636	0.756
LR	Pathology + Clinics	0.873	0.909	0.636	1
	Pathology	0.81	0.908	0.818	0.805
	Clinics	0.714	0.724	0.636	0.756
SVM	Pathology + Clinics	0.81	0.88	0.773	0.829
	Pathology	0.81	0.892	0.773	0.829
	Clinics	0.746	0.749	0.591	0.829
RandomForest	Pathology + Clinics	0.73	0.88	0.909	0.634
	Pathology	0.778	0.876	0.864	0.732
	Clinics	0.683	0.72	0.636	0.707

However, this research has limitations. All patients selected for this research had at least five biopsy cores containing PCa tissue, which introduced selection bias in patient inclusion. In addition, this research collected a retrospective cohort, and the prospective design would strengthen the findings of the study.

Conclusions

In summary, this study developed a DL system using digital pathology slides from prostate multi-core needle biopsies to predict PCa recurrence before radical prostatectomy. Due to the uncertainty in obtaining cancer tissue in biopsies, the proposed system can accept any number of WSIs as input. However, it should be noted that the system’s performance tends to decline when the number of inputs WSIs decreases. The predictive result can provide valuable reference for guiding radical prostatectomy procedures, such as considering more aggressive treatment approaches or hormone and immune therapy for high-risk patients. The system has demonstrated satisfactory performance in the testing cohort and the potential to produce favorable clinical benefits.

Abbreviations

PCa	Prostate cancer
PSA	Prostate-specific antigen
BCR	Biochemical recurrence

WSI	Whole slide image
CNN	Convolutional neural network
DL	Deep learning
MIL	Multiple instance learning
ML	Machine learning
H&E	Hematoxylin and Eosin
TMUCH	Tianjin Medical University Cancer Institute and Hospital
TBH	Tianjin Baodi Hospital
SD	Standard deviation
PALHI	Patch likelihood histogram
BoW	Bag of words
TF-IDF	Term frequency-inverse document frequency
MLP	Multilayer perceptron
LR	Logistic regression
SVM	Support vector machine
ROC	Receiver operating characteristic
AUC	Area under the receiver operating characteristic curve
SEM	Standard error of the mean
t-SNE	T-distributed stochastic neighbor embedding
DCA	Decision curve analysis
Grad-CAM	Gradient-weighted class activation mapping

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-025-13628-9>.

Supplementary Material 1.

Acknowledgements

Not applicable.

Authors’ contributions

L.C. was involved in design of experiments, analysis of results and wrote the manuscript. A.Z., and P.Z. were involved in programming the experimental setup, training the model. R.H., A.Z. and N.L. performed data collection, and analysis of results. L.C., L.L. and W.C. read and interpreted pathological images. W.C. and N.L. supervised the work. All authors reviewed the manuscript and agree with its content.

Funding

This study was funded by National Natural Science Foundation of China (grant number 82002813, 82202909), Tianjin Health Technology Project (grant number TJWJ2021QN010) and Tianjin Key Medical Discipline (Pathology) Construction Project (grant number TJYXZDXK-012A).

Data availability

The datasets analyzed during the current study were used with institutional permission through IRB approval for the current study and are thus not publicly available, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This multicenter retrospective study was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (No. Ek2020074), and a waiver of informed consent was granted.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 August 2024 Accepted: 3 February 2025
Published online: 10 February 2025

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–49.
- Amling CL, Blute ML, Bergstralh EJ, Seay TM, Slezak J, Zincke H. Long-term hazard of progression after radical prostatectomy for clinically localized prostate cancer: continued risk of biochemical failure after 5 years. *J Urol*. 2000;164(1):101–5.
- Han M, Partin AW, Pound CR, Epstein JI, Walsh PC. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy: the 15-year Johns Hopkins experience. *Urol Clin North Am*. 2001;28(3):555–65.
- Freedland SJ, Humphreys EB, Mangold LA, Eisenberger M, Dorey FJ, Walsh PC, Partin AW. Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *J Am Med Assoc*. 2005;294(4):433–9.
- Mun Y, Paik I, Shin S-J, Kwak T-Y, Chang H. Yet another automated Gleason grading system (YAAGGS) by weakly supervised deep learning. *NPJ Digit Med*. 2021;4(1):99.
- Epstein JI, Lecksell K, Carter HB. Prostate cancer sampled on sextant needle biopsy: significance of cancer on multiple cores from different areas of the prostate. *Urology*. 1999;54(2):291–4.
- Eichler K, Hempel S, Wilby J, Myers L, Bachmann LM, Kleijnen J. Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review. *J Urol*. 2006;175(5):1605–12.
- Cosma G, Acampora G, Brown D, Rees RC, Khan M, Pockley AG. Prediction of pathological stage in patients with prostate cancer: a neuro-fuzzy model. *PLoS One*. 2016;11(6):e0155856.
- Nagpal K, Foote D, Liu Y, Chen P-HC, Wulczyn E, Tan F, Olson N, Smith JL, Mohtashamian A, Wren JH. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med*. 2019;2(1):48.
- Esteve A, Feng J, van der Wal D, Huang S-C, Simko JP, DeVries S, Chen E, Schaeffer EM, Morgan TM, Sun Y. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med*. 2022;5(1):71.
- Chen N, Zhou Q. The evolving Gleason grading system. *Chin J Cancer Res*. 2016;28(1):58.
- Allsbrook WC Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, Bostwick DG, Humphrey PA, Jones EC, Reuter VE. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol*. 2001;32(1):74–80.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74.
- Sakamoto T, Furukawa T, Lami K, Pham HHN, Uegami W, Kuroda K, Kawai M, Sakanashi H, Cooper LAD, Bychkov A. A narrative review of digital pathology and artificial intelligence: focusing on lung cancer. *Transl Lung Cancer Res*. 2020;9(5):2255.
- Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer*. 2021;124(4):686–96.
- Rabilloud N, Allaume P, Acosta O, De Crevoisier R, Bourgade R, Loussouarn D, Rioux-Leclercq N, Khene Z-e, Mathieu R, Bensalah K. Deep learning methodologies applied to digital pathology in prostate cancer: a systematic review. *Diagnostics*. 2023;13(16):2676.
- Perincheri S, Levi AW, Celli R, Gershkovich P, Rimm D, Morrow JS, Rothrock B, Raciti P, Klimstra D, Sinard J. An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. *Mod Pathol*. 2021;34(8):1588–95.
- da Silva LM, Pereira EM, Salles PG, Godrich R, Ceballos R, Kunz JD, Casson A, Viret J, Chandarlapaty S, Ferreira CG. Independent real-world application of a clinical-grade automated prostate cancer detection system. *J Pathol*. 2021;254(2):147–58.
- Arvaniti E, Fricker KS, Moret M, Rupp N, Hermanns T, Fankhauser C, Wey N, Wild PJ, Rueschoff JH, Claassen M. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci Rep*. 2018;8(1):12054.
- Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, Kamoun A, Sefta M, Toldo S, Zaslavskiy M. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun*. 2020;11(1):3877.
- Weitz P, Wang Y, Kartasalo K, Egevad L, Lindberg J, Grönberg H, Eklund M, Rantalainen M. Transcriptome-wide prediction of prostate cancer gene expression from histopathology images using co-expression-based convolutional neural networks. *Bioinformatics*. 2022;38(13):3462–9.
- Pinckaers H, van Ipenburg J, Melamed J, De Marzo A, Platz EA, van Ginneken B, van der Laak J, Litjens G. Predicting biochemical recurrence of prostate cancer with artificial intelligence. *Commun Med (Lond)*. 2022;2(1):64.
- Farrokh M, Kumar N, Gann PH, Greiner R. Learning to predict prostate cancer recurrence from tissue images. *J Pathol Inform*. 2023;15:100344.
- Tsuneki M, Abe M, Ichihara S, Kanavati F. Inference of core needle biopsy whole slide images requiring definitive therapy for prostate cancer. *BMC Cancer*. 2023;23(1):11.
- Cao R, Yang F, Ma S-C, Liu L, Zhao Y, Li Y, Wu D-H, Wang T, Lu W-J, Cai W-J. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer. *Theranostics*. 2020;10(24):11080.
- Escalante HJ, Ponce-López V, Escalera S, Baró X, Morales-Reyes A, Martínez-Carranza J. Evolving weighting schemes for the bag of visual words. *Neural Comput Appl*. 2017;28:925–39.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
- Carroll PH, Mohler JL. NCCN guidelines updates: prostate cancer and prostate cancer early detection. *J Natl Compr Canc Netw*. 2018;16(5):620–3.
- Ward EM, Sherman RL, Henley SJ, Jemal A, Siegel DA, Feuer EJ, Firth AU, Kohler BA, Scott S, Ma J. Annual report to the nation on the status of cancer, featuring cancer in men and women age 20–49 years. *J Natl Cancer Inst*. 2019;111(12):1279–97.
- Campanella G, Hanna MG, Geneslaw L, Mirafior A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301–9.
- Brendel M, Getseva V, Al Assaad M, Sigouros M, Sigaras A, Kane T, Khosravi P, Mosquera JM, Elemento O, Hajirasouliha I. Weakly-supervised tumor purity prediction from frozen H&E stained slides. *EBioMedicine*. 2022;80:104067.
- Bukowy JD, Foss H, McGarry SD, Lowman AK, Hurrell SL, Iczkowski KA, Banerjee A, Bobholz SA, Barrington A, Dayton A. Accurate segmentation of prostate cancer histomorphometric features using a weakly supervised convolutional neural network. *J Med Imaging*. 2020;7(5):057501.
- Lee HW, Kim E, Na I, Kim CK, Seo SI, Park H. Novel multiparametric magnetic resonance imaging-based deep learning and clinical parameter integration for the prediction of long-term biochemical recurrence-free survival in prostate cancer after radical prostatectomy. *Cancers*. 2023;15(13):3416.
- Hou Y, Jiang K-W, Wang L-L, Zhi R, Bao M-L, Li Q, Zhang J, Qu J-R, Zhu F-P, Zhang Y-D. Biopsy-free AI-aided precision MRI assessment in prediction of prostate cancer biochemical recurrence. *Br J Cancer*. 2023;129(10):1625–33.
- Shiradkar R, Ghose S, Mahran A, Li L, Hubbard I, Fu P, Tirumani SH, Ponsky L, Purysko A, Madabhushi A. Prostate surface distension and tumor texture descriptors from pre-treatment MRI are associated with biochemical recurrence following radical prostatectomy: preliminary findings. *Front Oncol*. 2022;12:841801.
- Papp L, Spielvogel C, Grubmüller B, Grahovac M, Krajnc D, Ecsedi B, Saresghi RA, Mohamad D, Hamboeck M, Rausch I. Supervised machine learning enables non-invasive lesion characterization in primary prostate cancer with [68 Ga] Ga-PSMA-11 PET/MRI. *Eur J Nucl Med Mol Imaging*. 2021;48:1795–805.
- Ekşi M, Evren İ, Akkaş F, Arıkan Y, Özdemir O, Özlü DN, Ayten A, Sahin S, Tuğcu V, Taşçı AI. Machine learning algorithms can more efficiently predict biochemical recurrence after robot-assisted radical prostatectomy. *Prostate*. 2021;81(12):913–20.
- Park S, Byun J, Woo JY. A machine learning approach to predict an early biochemical recurrence after a radical prostatectomy. *Appl Sci*. 2020;10(11):3854.
- Zhang Y-D, Wang J, Wu C-J, Bao M-L, Li H, Wang X-N, Tao J, Shi H-B. An imaging-based approach predicts clinical outcomes in prostate cancer

- through a novel support vector machine classification. *Oncotarget*. 2016;7(47):78140.
40. Goyal NK, Kumar A, Acharya RL, Dwivedi US, Trivedi S, Singh PB, Singh T. Prediction of biochemical failure in localized carcinoma of prostate after radical prostatectomy by neuro-fuzzy. *Indian Journal of Urology*. 2007;23(1):14–7.
 41. Poulakis V, Witzsch U, De Vries R, Emmerlich V, Meves M, Altmannsberger H-M, Becht E. Preoperative neural network using combined magnetic resonance imaging variables, prostate specific antigen, and Gleason score to predict prostate cancer recurrence after radical prostatectomy. *Eur Urol*. 2004;46(5):571–8.
 42. Yan Y, Shao L, Liu Z, He W, Yang G, Liu J, Xia H, Zhang Y, Chen H, Liu C. Deep learning with quantitative features of magnetic resonance images to predict biochemical recurrence of radical prostatectomy: A multi-center study. *Cancers*. 2021;13(12):3098.
 43. Wong NC, Lam C, Patterson L, Shayegan B. Use of machine learning to predict early biochemical recurrence after robot-assisted prostatectomy. *BJU Int*. 2019;123(1):51–7.
 44. Liu J, Zhang H, Woon DT, Perera M, Lawrentschuk N. Predicting Biochemical Recurrence of Prostate Cancer Post-Prostatectomy Using Artificial Intelligence: A Systematic Review. *Cancers*. 2024;16(21):3596.
 45. Leo P, Chandramouli S, Farre X, Elliott R, Janowczyk A, Bera K, Fu P, Janaki N, El-Fahmawi A, Shahait M. Computationally derived cribriform area index from prostate cancer hematoxylin and eosin images is associated with biochemical recurrence following radical prostatectomy and is most prognostic in gleason grade group 2. *Eur Urol Focus*. 2021;7(4):722–32.
 46. Potter SR, Miller MC, Mangold LA, Jones KA, Epstein JI, Veltri RW, Partin AW. Genetically engineered neural networks for predicting prostate cancer progression after radical prostatectomy. *Urology*. 1999;54(5):791–5.
 47. Kim J-K, Hong S-H, Choi I-Y. Partial correlation analysis and Neural-Network-Based prediction model for biochemical recurrence of prostate cancer after radical prostatectomy. *Appl Sci*. 2023;13(2):891.
 48. Tan YG, Fang AH, Lim JK, Khalid F, Chen K, Ho HS, Yuen JS, Huang HH, Tay KJ. Incorporating artificial intelligence in urology: Supervised machine learning algorithms demonstrate comparative advantage over nomograms in predicting biochemical recurrence after prostatectomy. *Prostate*. 2022;82(3):298–305.
 49. Sargos P, Leduc N, Giraud N, Gandaglia G, Roumiguié M, Ploussard G, Rozet F, Soulié M, Mathieu R, Artus PM. Deep neural networks outperform the CAPRA score in predicting biochemical recurrence after prostatectomy. *Front Oncol*. 2021;10:67923.
 50. Park J, Rho MJ, Moon HW, Kim J, Lee C, Kim C-S, Jeon SS, Kang M, Lee JY. Dr. Answer Ai for prostate cancer: predicting biochemical recurrence following radical prostatectomy. *Technol Cancer Res Treat*. 2021;20:15330338211024660.
 51. Lee SJ, Yu SH, Kim Y, Kim JK, Hong JH, Kim C-S, Seo SI, Byun S-S, Jeong CW, Lee JY. Prediction system for prostate cancer recurrence using machine learning. *Appl Sci*. 2020;10(4):1333.
 52. Hu X-H, Cammann H, Meyer H-A, Jung K, Lu H-B, Leva N, Magheli A, Stephan C, Busch J. Risk prediction models for biochemical recurrence after radical prostatectomy using prostate-specific antigen and Gleason score. *Asian J Androl*. 2014;16(6):897–901.
 53. Porter C, O'Donnell C, Crawford ED, Gamito EJ, Errejon A, Genega E, Sotelo T, Tewari A. Artificial neural network model to predict biochemical failure after radical prostatectomy. *Mol Urol*. 2001;5(4):159–62.
 54. Han M, Snow PB, Epstein JI, Chan TY, Jones KA, Walsh PC, Partin AW. A neural network predicts progression for men with Gleason score 3+ 4 versus 4+ 3 tumors after radical prostatectomy. *Urology*. 2000;56(6):994–9.
 55. Sandeman K, Blom S, Koponen V, Manninen A, Juhila J, Rannikko A, Roponen T, Mirtti T. AI model for prostate biopsies predicts cancer survival. *Diagnostics*. 2022;12(5):1031.
 56. Eminaga O, Saad F, Tian Z, Wolffgang U, Karakiewicz PI, Ouellet V, Azzi F, Spieker T, Helmke BM, Graefen M. Artificial intelligence unravels interpretable malignancy grades of prostate cancer on histology images. *NPJ Imaging*. 2024;2(1):6.
 57. Huang W, Randhawa R, Jain P, Hubbard S, Eickhoff J, Kummar S, Wilding G, Basu H, Roy R. A novel artificial intelligence-powered method for prediction of early recurrence of prostate cancer after prostatectomy and cancer drivers. *JCO Clin Cancer Inform*. 2022;6:e2100131.
 58. Lu MY, Williamson DF, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. 2021;5(6):555–70.
 59. Silva-Rodríguez J, Schmidt A, Sales MA, Molina R, Naranjo V. Proportion constrained weakly supervised histopathology image classification. *Comput Biol Med*. 2022;147:105714.
 60. Presti JC. Prostate biopsy: how many cores are enough? *Urol Oncol*. 2003;21(2):135–40.
 61. Bjurlin MA, Wysock JS, Taneja SS. Optimization of prostate biopsy: review of technique and complications. *Urol Clin*. 2014;41(2):299–313.
 62. Cai X, Zhang H, Wang Y, Zhang J, Li T. Digital pathology-based artificial intelligence models for differential diagnosis and prognosis of sporadic odontogenic keratocysts. *Int J Oral Sci*. 2024;16(1):16.
 63. Volinsky-Fremont S, Horeweg N, Andani S, Barkey Wolf J, Lafarge MW, de Kroon CD, Ørtoft G, Høgdaal E, Dijkstra J, Jobsen JJ. Prediction of recurrence risk in endometrial cancer with multimodal deep learning. *Nat Med*. 2024;30(7):1962–73.
 64. Iczkowski KA, Torkko KC, Kotnis GR, Storey Wilson R, Huang W, Wheeler TM, Abeyta AM, La Rosa FG, Cook S, Werahera PN. Digital quantification of five high-grade prostate cancer patterns, including the cribriform pattern, and their association with adverse outcome. *Am J Clin Pathol*. 2011;136(1):98–107.
 65. Dong F, Yang P, Wang C, Wu S, Xiao Y, McDougal WS, Young RH, Wu C-L. Architectural heterogeneity and cribriform pattern predict adverse clinical outcome for Gleason grade 4 prostatic adenocarcinoma. *Am J Surg Pathol*. 2013;37(12):1855–61.
 66. Siadat F, Sykes J, Zlotta AR, Aldaoud N, Egawa S, Pushkar D, Kuk C, Bristow RG, Montironi R, Van Der Kwast T. Not all Gleason pattern 4 prostate cancers are created equal: a study of latent prostatic carcinomas in a cystoprostatectomy and autopsy series. *Prostate*. 2015;75(12):1277–84.
 67. Choy B, Pearce SM, Anderson BB, Shalhav AL, Zagaja G, Eggner SE, Paner GP. Prognostic significance of percentage and architectural types of contemporary Gleason pattern 4 prostate cancer in radical prostatectomy. *Am J Surg Pathol*. 2016;40(10):1400–6.
 68. Duenweg SR, Brehler M, Lowman AK, Bobholz SA, Kyereme F, Winiarz A, Nath B, Iczkowski KA, Jacobsohn KM, LaViolette PS. Quantitative histomorphometric features of prostate cancer predict patients who biochemically recur following prostatectomy. *Lab Invest*. 2023;103(12):100269.
 69. Reeves FA, Battye S, Roth H, Peters JS, Hovens C, Costello AJ, Corcoran NM. Prostatic nerve subtypes independently predict biochemical recurrence in prostate cancer. *J Clin Neurosci*. 2019;63:213–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.