

Artificial intelligence for enhanced diagnostic precision of prostate cancer

Agus Rizal Ardy Hariandy Hamid¹, Agnes Stephanie Harahap², Monik Ediana Miranda², Kahlil Gibran³, Nabila Husna Shabrina⁴



pISSN: 0853-1773 • eISSN: 2252-8083
<https://doi.org/10.13181/mji.0a.258312>
Med J Indones. 2025;34:189–200

Received: July 07, 2025

Accepted: September 22, 2025

Authors' affiliations:

¹Department of Urology, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jakarta, Indonesia, ²Department of Anatomical Pathology, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jakarta, Indonesia, ³Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia, ⁴Department of Computer Engineering, Universitas Multimedia Nusantara, Tangerang, Banten, Indonesia

Corresponding author:

Nabila Husna Shabrina
 Department of Computer Engineering,
 Universitas Multimedia Nusantara,
 Jalan Scientia Boulevard, Gading
 Serpong, Curug Sangereng, Kelapa Dua,
 Tangerang, Banten 15810, Indonesia
 Tel/Fax: +62-21-54220808
E-mail: nabila.husna@umn.ac.id

ABSTRACT

BACKGROUND Accurate diagnosis and grading of prostate cancer are essential for treatment planning. The role of artificial intelligence in prostate cancer intervention and diagnosis (RAPID) is a study aimed at developing artificial intelligence (AI) models to enhance diagnostic precision in prostate cancer by distinguishing malignant from non-cancerous histopathological findings.

METHODS Histopathological images were collected between 2023 and 2024 at the Department of Anatomical Pathology, Faculty of Medicine, Universitas Indonesia. The dataset included benign prostatic hyperplasia and prostate cancer cases. All slides were digitized and manually annotated by pathologists. Patch-based classification was performed using convolutional neural network and transformer-based models to differentiate malignant from non-malignant tissues.

RESULTS A total of 529 whole-slide images were processed, yielding 26,418 image patches for model training and testing. Deep learning models achieved strong performance in classification. Architectures including EfficientNetV2Bo, Xception, ConvNeXt-Tiny, and Vision Transformer (ViT) achieved near-perfect classification outcomes. EfficientNetV2Bo reached an AUC of 1.00 (95% CI: 1.00–1.00), sensitivity 0.99 (95% CI: 0.99–1.00), and specificity 1.00 (95% CI: 1.00–1.00). Xception and ConvNeXt-Tiny both achieved AUC 1.00 (95% CI: 1.00–1.00) with sensitivity and specificity of 1.00 (95% CI: 1.00–1.00). ViT performed strongly with AUC 0.999 (95% CI: 0.99–1.00), sensitivity 0.99 (95% CI: 0.99–0.99), and specificity 0.99 (95% CI: 0.99–0.99).

CONCLUSIONS RAPID demonstrated high potential as an AI-based diagnostic tool for prostate cancer, showing excellent accuracy in histopathological classification using the Indonesian dataset. These findings highlight the feasibility of deploying deep learning models to support diagnostic decision-making in clinical practice.

KEYWORDS artificial intelligence, computer-assisted diagnosis, computer-assisted image interpretation, pathology, prostatic neoplasms

Prostate cancer, the fourth most newly diagnosed cancer worldwide, accounted for 1.47 million new cases and 397,000 deaths in 2022, and remains the eighth leading cause of cancer mortality. It ranks as the fifth most prevalent cancer among Indonesian men, with 13,130 new cases (approximately 7.0%) in 2022.^{1–3} Prognosis in prostate cancer depends on several

factors, including age at diagnosis, tumor grade, tumor volume, and evidence of local invasion or metastasis.⁴ These disparities and prognostic challenges highlight the urgent need for early detection and optimized treatment strategies.

Distinguishing benign prostatic hyperplasia (BPH) from prostate cancer represents a key diagnostic

challenge, as both conditions commonly affect older men and share overlapping symptoms, particularly lower urinary tract symptoms (LUTS). Although BPH prevalence increases significantly with age—affecting >70% of men aged 60–69 years and >80% of men aged >70 years—prostate cancer incidence also rises with age and can be asymptomatic in early stages. Additionally, prostate size does not always correlate with LUTS severity, and regional variations further complicate diagnosis.⁵ The clinical and histopathological overlap between BPH and prostate cancer, along with the frequent coexistence of both conditions and interobserver variability in conventional histopathological assessment, further complicates diagnosis and treatment decisions.^{6,7} These limitations underscore the need for diagnostic tools that are more objective and accurate, providing decision support for pathologists in reliably distinguishing benign from malignant prostate conditions, while recognizing that pathological assessment remains the gold standard.

Given these ongoing diagnostic challenges, there is increasing interest in applying artificial intelligence (AI) and machine learning algorithms to enhance clinical and histopathological evaluation in prostate cancer.^{8–10} However, recent international surveys of urology healthcare providers revealed that although there is strong optimism regarding AI's role in diagnostics and treatment decision-making, clinical

validation remains a prerequisite for widespread adoption.¹¹ The role of artificial intelligence in prostate cancer intervention and diagnosis (RAPID) is an AI-based diagnostic framework developed at a leading national referral and teaching hospital in Indonesia. It integrates pathological data to support the development of AI models for prostate cancer screening, diagnosis, and prognosis. The unique histological and morphological characteristics of prostate cancer in individual patients necessitate personalized research approaches using samples and data derived from the Indonesian population. Therefore, in this study, we aimed to evaluate the RAPID diagnostic framework by developing and testing convolutional neural network (CNN) and transformer-based models to distinguish prostate cancer from benign histopathological findings in an Indonesian dataset, with the goal of enhancing diagnostic accuracy, reproducibility, and clinical decision support in prostate cancer management.

METHODS

This was a single-center, retrospective diagnostic accuracy study of stored histopathology slides in line with the biomedical image analysis challenges, standards for reporting of diagnostic accuracy studies artificial intelligence (STARD-AI), and strengthening

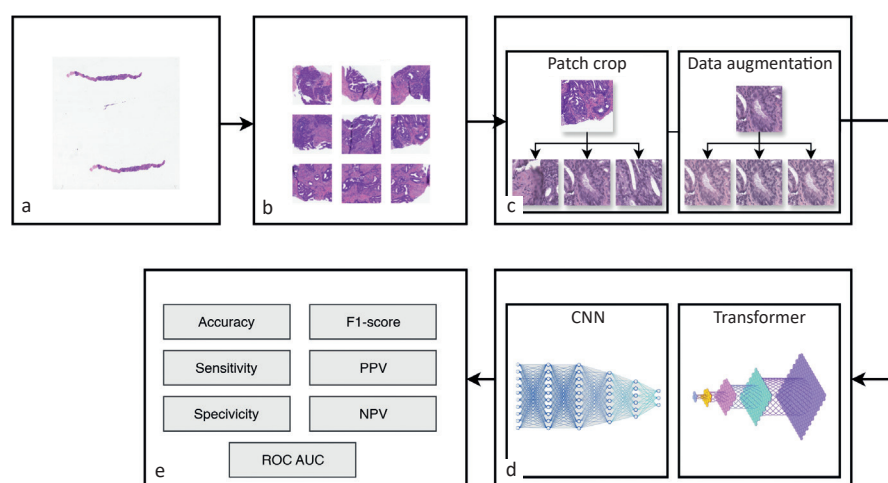


Figure 1. Workflow of the WSI-based prostate histopathology pipeline. (a) Dataset acquisition: a total of 529 WSIs scanned at $\times 40$ magnification; (b) region of interest extraction and labeling: pathologists manually delineated diagnostically relevant regions, saved as 2048×2048 pixel images, and annotated as malignant or non-malignant; (c) pre-processing and augmentation: extracted regions were patch-cropped into 512×512 pixels and underwent stain normalization and data augmentation; (d) model development: CNN, transformer models, and hybrid architectures were independently trained on the training set; (e) model evaluation: performance was assessed on a held-out test set using accuracy, sensitivity, specificity, PPV, NPV, F1-score, and AUC, each reported with 95% confidence interval. CNN=convolutional neural networks; NPV=negative predictive value; PPV=positive predictive value; ROC AUC=area under the receiver operating characteristic curve; WSI=whole-slide imaging

the reporting of observational studies in epidemiology (STROBE) guidelines. This study comprised several key stages, as illustrated in Figure 1. First, the patient slides were digitized using a whole-slide imaging (WSI) scanner.¹²⁻¹⁵ Subsequently, the digitized slides were preprocessed and the relevant features were extracted from the WSI files. Each extracted region was manually labeled by an expert pathologist. Subsequently, the labeled data were divided into smaller image patches and split into training, validation, and test sets. Data augmentation techniques were applied to balance the data and enhance model generalization. Finally, multiple deep-learning models were trained on the prepared dataset, and their performances were evaluated using a range of quantitative metrics. Potential confounders and effect modifiers, including stain variability, scanner or batch effects, tissue type, and inflammatory changes, have been recognized as factors influencing model performance.

Dataset collection

Data from patients diagnosed with prostate cancer and BPH were collected from the archive of the Department of Anatomical Pathology, Faculty of Medicine, Universitas Indonesia. Consecutive sampling was applied to include all cases from January 2023 to December 2024. This approach ensured that all consecutive eligible cases during the study period were included without additional selection beyond the predefined inclusion and exclusion criteria, thereby minimizing sampling bias. Inclusion criteria included history of core biopsy, transurethral resection, or prostatectomy. Cases with missing or poor-quality hematoxylin and eosin (H&E) slides, or those exhibiting marked inflammation or diagnosed as overt prostatitis, were excluded. All H&E slides were prepared with sufficient quality to allow digitization by the slide scanner. Slides that were blurred, unevenly stained, scattered, or otherwise inadequate for histopathological evaluation were excluded from the study. The scanner was calibrated on a monthly basis in accordance with the standard operating procedures established by the respective slide scanner company. Importantly, the non-tissue background was retained during digitization to preserve the natural appearance of each slide and maintain fidelity with routine histopathological practice. Several BPH cases included in the study demonstrated mild to moderate inflammation.

A board-certified pathologist with 10 years of experience and subspecialty training in uropathology (Pathologist 1) re-reviewed all H&E slides under a light microscope (Leica Microsystems, Germany) and scanned them using an Aperio GT450 whole-slide scanner (Leica Biosystems, USA), yielding a total of 555 WSIs. All slides were scanned at $\times 40$ magnification, corresponding to a spatial resolution of $0.26 \mu\text{m}/\text{pixel}$. The scanner was operated using the manufacturer's default international color consortium color profile, and automated color calibration was performed prior to each scanning session using an onboard reference. As part of the image acquisition workflow, quality control procedures were applied using the scanner's built-in algorithms. A second board-certified pathologist with 12 years of experience, without uropathology subspecialty (Pathologist 2), labeled the images as malignant or non-malignant. Labeling was performed according to the latest diagnostic criteria of WHO/ISUP classification.¹⁶ Equivocal cases requiring immunohistochemistry for definitive diagnosis were excluded, ensuring that only cases with clear diagnostic categories were retained. Pathologist 1 subsequently re-evaluated the categorizations or labeling made by Pathologist 2, and no discrepancies were identified.

Region extraction and labeling

All digitized WSIs were subjected to manual region of interest (ROI) extraction using ImageScope (Aperio Technologies, USA). For each slide, approximately 15–25 representative regions were selected by visual inspection, focusing on areas of diagnostic relevance. These regions were extracted at a fixed resolution of 2048×2048 pixels and saved in an uncompressed TIFF format to preserve image quality. Representative examples of extracted regions are shown in Figure 1b. This standardized patch extraction process ensures consistent image quality and dimensions for subsequent manual labeling and deep-learning model development.

The extracted ROIs were subsequently reviewed and labeled by board-certified pathologists who was blinded to patient metadata, clinical outcomes, and the assessments among pathologists to reduce potential bias during the annotation process. Then, each verified image was categorized into one of two classes, malignant or non-malignant, and organized into the corresponding directories. Accordingly, in this study, the AI classification output was limited to

these two categories and did not provide further sub-classification. The diagnostic criteria for acinar prostatic adenocarcinoma are divided into essential and desirable features. Malignant glands infiltrating the stroma, loss of basal cells, and nuclear features including enlargement and hyperchromasia are essential criteria. Prominent nucleoli, atypical luminal contents, and cytoplasmic features are desirable criteria. No missing labels remained in the dataset, and any unreadable or artefactual regions were excluded during verification to ensure only diagnostically interpretable images were included. Specific but not sensitive findings, such as perineural invasion, mucinous fibroplasia, and glomerulations, are also included in the desirable criteria. Ductal type adenocarcinoma with papillary structures and/or complex and cribriform glands lined with tall columnar pseudostratified cells is also included in the malignant type, but none matched the criteria in this case selection.¹⁶ Grading in this analysis was limited to discrimination between malignant and non-malignant lesions, and identification of prostatic adenocarcinoma only, without further sub-classification of histological patterns. ROIs were manually selected by the annotators to prevent exclusion of representative tumor and non-tumor areas. This subjective manual selection approach was applied consistently across all cases to minimize sampling bias and obtain consistent datasets.

Data preprocessing

In the data preprocessing stage, each categorized image was divided into smaller patches using the Patchify library.¹⁷ Specifically, each 2048 × 2048 pixel image was cropped into multiple non-overlapping 512 × 512 pixel patches (step size = 0). This patching strategy was employed to reduce computational complexity and memory demands during model training, given the high resolution of the original images. Then, the resulting patches were randomly partitioned into training, validation, and testing sets using a 70:15:15 split ratio. Figure 1C (left) shows representative examples of the extracted image patches used for the model training and evaluation.

To address the significant class imbalance, data augmentation was applied to the malignant class using StainLib and Albumentations.^{18,19} Each malignant patch was augmented 20-fold using color normalization based on the Macenko method,²⁰ which simulates realistic variations in histological staining.

This strategy substantially increased the number of malignant samples, bringing the dataset closer to class parity and thereby enhancing model fairness while reducing prediction bias toward the non-malignant class.

Model training

The evaluated architectures encompassed three major families of deep-learning models: CNNs, transformer-based models, and hybrid designs. CNNs are designed to learn local spatial features through convolutional operations and pooling layers, making them particularly effective in medical image analysis, where fine-grained morphological features are critical.¹⁷ These models are well-suited for histopathological classification tasks, as they can extract localized patterns related to glandular structures, nuclei, or stromal components. Transformer-based models, originally developed for natural language processing, rely on self-attention mechanisms to model long-range dependencies across input sequences or image patches. Histopathologically, transformers can capture global contextual relationships across large images, which is essential for accurately identifying patterns that may span broad tissue regions. Their ability to holistically weigh spatial features can enhance diagnostic precision, especially in complex or diffuse cases.²¹ Hybrid architectures aim to integrate the strengths of both CNNs and transformers. These models typically retain the efficient local feature extraction of CNNs while incorporating architectural elements from transformers, such as normalization strategies, attention-inspired block design, and modified activation functions, to improve the performance and generalizability of visual tasks.

Our findings were applied uniformly across all specimen types, including prostate biopsies, transurethral resection of the prostate samples (TURP), and radical prostatectomy specimens, thereby providing a generalizable assessment independent of specimen origin. Nonetheless, we recognize the potential for domain shift, as differences in tissue processing, staining protocols, and scanner calibration between specimen types may influence reproducibility and generalizability. Although this study did not specifically address domain adaptation, future work will aim to incorporate strategies to minimize the impact of domain shift and further refine prognostic stratification.

The CNN-based models evaluated in this study included ResNet50, ResNet50V2, DenseNet121, MobileNetV2, EfficientNetV2Bo, and Xception.^{22–26} Transformer-based models tested were the Vision Transformer (ViT)²⁷ and the Data-efficient Image Transformer (DeiT).²⁸ The hybrid architecture evaluated was ConvNeXt-Tiny,²⁹ which applies modern transformer design principles within a convolutional framework. Model training was performed using Google Colab Pro+ (Google LLC, USA), providing access to high-performance NVIDIA A100 GPUs and 83 GB of RAM. All models were implemented and trained using Keras version 3 (François Chollet & Google Research, USA), with consistent hyperparameter settings applied across all backbone architectures to ensure a fair comparison. The architectural details for each model are provided in the Supplementary Materials, and all the models were trained using the Adam optimizer with a fixed learning rate of 0.0001. The classification task was binary, and sparse categorical cross-entropy was used as the loss function. A batch size of 32 was applied during training over a maximum of 25 epochs, and early stopping was implemented to prevent overfitting. The early stopping criterion monitored validation loss with a patience value of five epochs. To ensure reproducibility, a random seed of 1,337 was applied across all training runs.

In this study, we aimed to develop a model with high sensitivity and specificity for detecting prostate carcinoma. In subsequent stages, internal validation will be performed using an in-house dataset derived from hospital cases, followed by external validation involving cases from other institutions assessed by multiple pathologists. These validation phases are planned for future investigations, along with the development of appropriate software tailored to the needs of daily clinical practice.

Model evaluation

A set of quantitative metrics was computed to evaluate the diagnostic performance of each model. These included accuracy, recall (sensitivity), precision, F1-score, positive predictive value (PPV), negative predictive value (NPV), and the area under the receiver operating characteristic curve (AUC ROC). Macro-averaged scores were used for accuracy, recall, precision, and F1-score to ensure equal weighting of the malignant and non-malignant

classes. A fixed decision threshold of 0.5 was applied across all models for classification. All performance metrics were reported with 95% confidence intervals (CI) and calculated via bootstrap resampling with 1,000 iterations. Statistical analysis was performed using Python v3.12.11 (Python Software Foundation, USA), with CI estimation and resampling conducted using the scikit-learn and scipy.stats libraries. Decision-curve analysis and calibration analysis were considered; however, these were not performed at the current stage because the models were newly trained and have yet to be internally and externally validated. These metrics provide a comprehensive assessment of each model's ability to accurately classify malignant and non-malignant histopathological features. This study was approved by the Ethics Committee of the Faculty of Medicine, Universitas Indonesia (No: KET-121/UN2.F1/ETIK/PPM.00.02/2025).

RESULTS

In this study, 555 WSIs were initially scanned. Of these, 26 WSIs (4.7%) were excluded due to unsuitable processing, extensive hemorrhage, marked prostatitis, or suboptimal image quality that did not meet display standards, resulting in 529 digitized WSIs subjected to manual ROI extraction using ImageScope. Following the verification process, 3,828 images were classified as non-malignant, while 195 images were identified as malignant. Furthermore, 26,418 image patches were generated for model development and evaluation. However, the final dataset exhibited significant class imbalance, with 1,242 patches labeled as malignant and 25,176 patches labeled as non-malignant. To address this imbalance, data augmentation techniques, including stain-based augmentation, were applied to increase the number of malignant samples and achieve a more balanced distribution between classes.

Model performance overview

Table 1 presents a comparative analysis of the model performance using accuracy, specificity, sensitivity, PPV, NPV, AUC ROC, and F1-score, with all metrics reported alongside their corresponding 95% CI. Among the evaluated models, Xception, EfficientNetV2Bo, and ConvNeXt-Tiny demonstrated the highest performance, each achieving perfect

Table 1. Test-set performance of CNN, transformer, and hybrid models for binary histopathology classification

Type	Model	Accuracy (CI)	Specificity (CI)	Sensitivity (CI)	PPV (CI)	NPV (CI)	AUC ROC (CI)	F1-score (CI)
CNN	Xception*	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
	EfficientNetV2B0*	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.99 (0.99–1.00)	1.00 (1.00–1.00)	0.99 (0.99–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)
	MobileNetV2	0.95 (0.94–1.00)	1.00 (1.00–1.00)	<0.01 (0.00–0.00)	1.00 (1.00–1.00)	0.50 (0.48–0.51)	0.64 (0.63–0.66)	0.33 (0.31–0.36)
	ResNet50	0.94 (0.92–0.96)	0.86 (0.85–0.87)	1.00 (1.00–1.00)	0.88 (0.87–89)	1.00 (1.00–1.00)	0.99 (0.99–1.00)	0.94 (0.92–0.96)
	ResNet50V2	0.65 (0.64–0.66)	0.33 (0.32–0.34)	0.96 (0.97–0.97)	0.59 (0.58–0.61)	0.89 (0.88–0.91)	0.72 (0.71–0.73)	0.60 (0.59–0.61)
	DenseNet121	0.60 (0.59–0.61)	0.95 (0.94–0.96)	0.25 (0.24–0.26)	0.83 (0.81–0.86)	0.55 (0.54–0.57)	0.76 (0.75–0.78)	0.54 (0.53–0.56)
Transformer	ViT	0.98 (0.98–0.99)	0.99 (0.99–0.99)	0.99 (0.98–0.99)	0.99 (0.99–0.99)	0.99 (0.98–0.99)	0.99 (0.99–1.00)	0.98 (0.98–0.99)
	DeiT	0.50 (0.50–0.50)	0 (0.00–0.00)	In 0	0.50 (0.50–0.50)	NaN	0.48 (0.48–0.52)	0.34 (0.33–0.34)
Hybrid [†]	ConvNeXT-Tiny*	1.00 (1.00–1.00)	0.99 (0.99–1.00)	1.00 (1.00–1.00)	0.99 (0.99–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)

AUC ROC=area under the receiver operating characteristic curve; CNN=convolutional neural networks; DeiT=data-efficient image transformer; NaN=not a number; NPV=negative predictive value; PPV=positive predictive value; ViT=vision transformer

*Best metric; [†]CNN architectures incorporating transformer-inspired design elements, in this study represented by ConvNeXT-Tiny

or near-perfect scores across all metrics, including an AUC ROC of 1.0 and an F1-score of 1, indicating robustness in distinguishing prostate cancer from benign cases within the dataset. ViT also performed exceptionally well, with an accuracy of 0.98, AUC ROC of 0.9992, and F1-score of 0.98, reflecting strong sensitivity and specificity in this classification task. No pre-specified subgroup analyses were performed; results reflect overall performance across the entire dataset.

In contrast, ResNet50V2, DenseNet121, and DeiT models showed lower performance compared with the other evaluated architectures. ResNet50V2 achieved an accuracy of 0.65 and AUC ROC of 0.7242, while DenseNet121 reported an accuracy of 0.60 and AUC ROC of 0.7657. The DeiT model showed the lowest overall performance, with an accuracy of 0.50 and an AUC ROC of 0.4852, indicating limited effectiveness for this diagnostic application. ResNet50 and MobileNetV2 demonstrated moderate performance, achieving high accuracies of 0.94 and 0.95, respectively; however, MobileNetV2's very low sensitivity suggests a tendency to miss malignant cases despite high overall accuracy.

In addition to performance metrics, the average end-to-end inference time per patch was measured. For 512 × 512 pixel input patches, the model achieved an average processing time of approximately 388 ms per patch, including data loading, preprocessing, and model inference.

The confusion matrices in Figure 2 complement the numerical metrics reported in Table 1, offering a visual representation of each model's performance and potential misclassification risk in real-world diagnostic settings. Each confusion matrix summarizes the number of true positives, true negatives, false positives, and false negatives on the held-out test set using a fixed classification threshold of 0.5. Class labels included non-malignant and malignant. Models, such as Xception, EfficientNetV2Bo, DeiT, and ConvNeXT-Tiny, achieved near-perfect classification, with negligible or zero misclassification errors. In contrast, models such as ResNet50, ResNet50V2, and DenseNet121 exhibited higher misclassification rates—particularly for false-positive and false-negative predictions—compared with the top-performing architectures. ViT presented a relatively balanced outcome with only minor

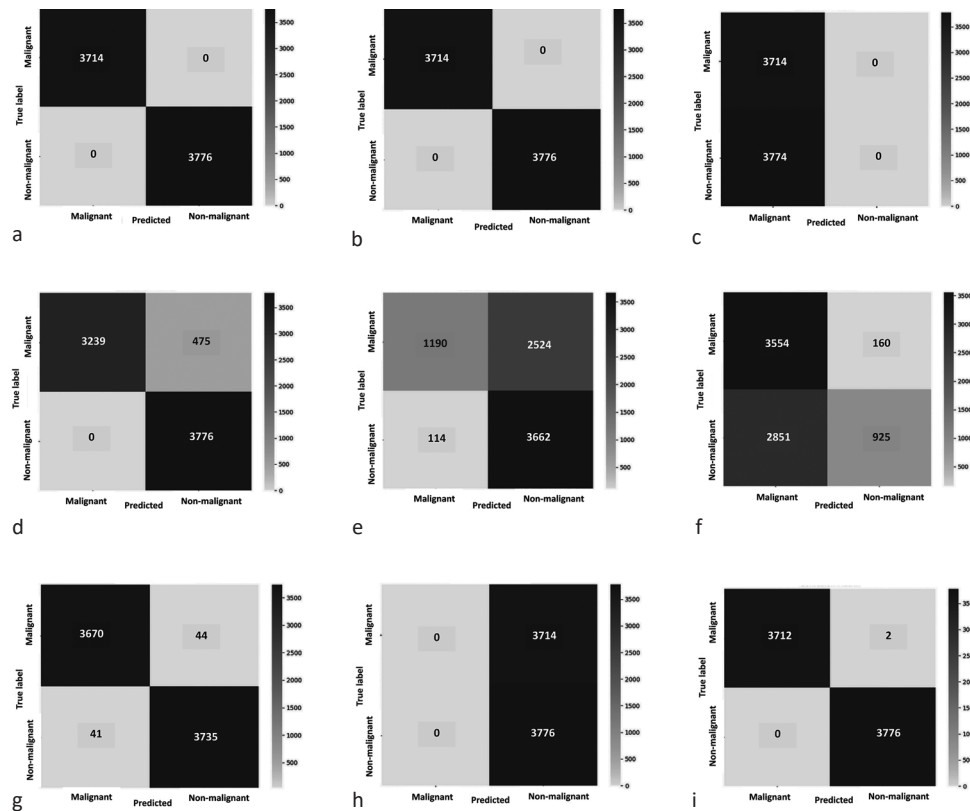


Figure 2. Confusion matrices illustrating the classification performance of each model on the held-out test set at a fixed threshold of 0.5. True labels are shown on the y-axis, predicted labels on the x-axis. Each cell represents the number of instances for the respective classification outcome: true positive (top left), true negative (bottom right), false positive (bottom left), false negative (top right). Darker shading indicates higher counts. (a) Xception; (b) EfficientNetV2Bo; (c) MobileNetV2; (d) ResNet50; (e) ResNet50V2; (f) DenseNet121; (g) ViT; (h) DeiT; (i) ConvNeXt Tiny. DeiT=data-efficient image transformer; ViT=vision transformer

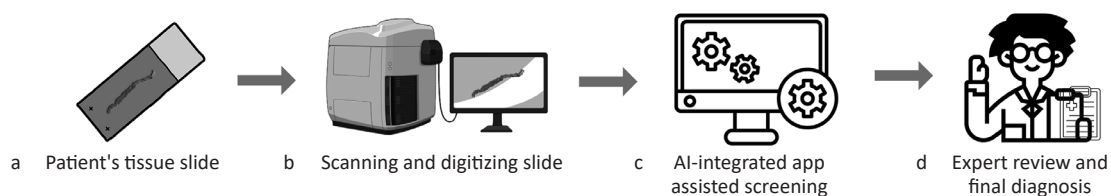


Figure 3. RAPID workflow for digital pathology diagnosis. (a) Patient's glass slide; (b) whole-slide scanning to generate a digital WSI file; (c) AI application performs assisted screening of tissue regions; (d) pathologist reviews AI outputs together with the WSI to confirm the final diagnosis. AI=artificial intelligence; RAPID=role of artificial intelligence in prostate cancer intervention and diagnosis; WSI=whole-slide image

errors. These visual patterns underscore differences in model behavior and robustness, particularly in handling borderline or ambiguous diagnostic cases.

DISCUSSION

As part of this initiative, the RAPID framework incorporates state-of-the-art deep-learning models specifically designed for image-based medical analysis. One of the most widely used architectures in this domain is the CNN, which is a deep-learning model specifically

designed for processing visual data, such as images or videos. Prior studies have shown the efficiency of this model in differentiating prostate cancer from BPH using transrectal ultrasound images.³⁰ In parallel with CNN advancements, transformer-based models have emerged as powerful alternatives for image analysis tasks in medical imaging, including digital pathology. Originally developed for natural language processing, transformers apply self-attention mechanisms that allow them to capture long-range dependencies and global contextual information more effectively than

traditional convolutional models.³¹ This architectural advantage makes transformers particularly promising for analyzing histopathological images, where spatial relationships and tissue architecture are critical for accurate classification.³² Recent studies have demonstrated the successful application of transformers in various cancer classification tasks, showing competitive or superior performance to CNNs.²¹

This study evaluated and compared the performance of several deep-learning architectures for distinguishing prostate cancer from BPH using high-resolution histopathological images. Unlike prior studies that often rely on a single model or publicly available datasets, our research leverages primary data collected directly from a top Indonesian referral hospital with numerous prostate cancer cases and systematically compares the performance of several state-of-the-art CNN and transformer-based models. This comparative approach enables a more comprehensive evaluation of model robustness and generalizability for distinguishing prostate cancer from BPH based on histopathological images.

The results demonstrated that multiple models achieved high diagnostic performance, with accuracy and AUC ROC values approaching or reaching 1.0. These findings underscore the promising role of AI in supporting the histopathological diagnosis of prostate lesions, consistent with previous studies.^{33,34} Among the architectures evaluated, three deep-learning models—Xception, EfficientNetV2Bo, and ConvNeXT-Tiny—demonstrated outstanding classification performance, achieving perfect scores across all evaluation metrics including accuracy, sensitivity, specificity, PPV, NPV, AUC ROC, and F1 score. These results reflect the high capacity of modern convolutional and hybrid architectures to capture subtle morphological differences in histopathological images of prostate tissues. The performance of Xception and EfficientNetV2 has been previously validated in medical image-classification tasks, showing robust feature extraction and generalization across datasets.²⁵ ConvNeXT-Tiny, a newer architecture combining convolutional principles with transformer-inspired design, has shown competitive results in image recognition benchmarks, and this study extends its utility to histopathology.

Among all the models tested, ViT stood out for its excellent balance between sensitivity (98.78%) and

specificity (98.94%), with an AUC ROC of 0.9992. Unlike CNNs, ViT employs a self-attention mechanism that allows it to capture long-range spatial dependencies across images. This characteristic is particularly beneficial in prostate histopathology, where the architectural distribution of glands and surrounding stroma plays a crucial role in diagnosis.^{23,31} The strong performance of ViT suggests that transformer-based models may offer distinct advantages over CNNs, particularly in complex classification tasks involving subtle morphological differences. High sensitivity reduces the risk of false-negative cancer diagnoses and improves patient outcomes, while high specificity minimizes the risk of overdiagnosis and overtreatment of benign conditions. Models, such as ViT, with near-perfect diagnostic performance, offer strong potential for integration into computer-aided diagnosis systems to assist pathologists in routine workflow.

In contrast, ResNet50V2 and DenseNet121 showed notably lower performance than the top-performing models, with AUC ROC values of 0.7242 and 0.7657, respectively. These models exhibited imbalanced sensitivity and specificity, suggesting difficulty in capturing the morphological diversity necessary for accurate classification. MobileNetV2, although achieving perfect specificity and PPV, had extremely low sensitivity (0.05%) and NPV (49.6%), resulting in a poor F1-score (0.33). This reflects a strong bias toward predicting non-malignant cases, likely owing to its lightweight architecture and limited feature extraction capability. The DeiT model performed the worst across nearly all metrics, with sensitivity and specificity values near zero, low AUC ROC (0.4852), and unreliable NPV. This result suggests a lack of stability and adaptability for this classification task, possibly due to insufficient domain-specific pretraining or architectural mismatch with patch-based histopathological inputs.

The use of a balanced patch-based dataset combined with extensive image augmentation likely contributed to the robust performance observed in several models. These preprocessing strategies enhance generalizability and mitigate class imbalance. However, it is important to acknowledge the inherent limitations of patch-based training, particularly in the context of clinical practice, where histopathological diagnosis is conducted on entire WSIs rather than isolated patches. As emphasized by Campanella et al,³⁵ clinical-grade AI systems require validation on WSI-

level datasets under realistic diagnostic conditions to ensure reliability and applicability.

Clinically, the variation in model performance highlights the importance of selecting AI tools capable of accurately distinguishing between malignant and benign prostate tissue. High-performing models, such as Xception, EfficientNetV2Bo, ConvNeXt-Tiny, and ViT, demonstrated superior ability to recognize significant histopathological features, including glandular architecture, nuclear morphology, and tissue organization relevant to prostate cancer diagnosis. Their greater diagnostic precision suggests substantial potential as decision-support systems, helping pathologists reduce interobserver variability and improve diagnostic reliability.

In contrast, poor-performing models, including ResNet50V2, DenseNet121, and MobileNetV2, indicate the challenge of deploying general image-classification networks onto complex histopathological data. Such findings highlight that not all AI architectures are equally suitable for pathology, and models must be selected carefully for clinical translation. Ultimately, integrating accurate and trustworthy AI models may facilitate earlier detection, improve grading consistency, and enhance patient management in prostate cancer.

The improvements observed in this study can be attributed to several factors. First, the use of advanced architectures capable of modeling complex morphological patterns and long-range spatial dependencies plays a key role. Transformer-based models, such as ViT, benefit from self-attention mechanisms, allowing them to capture global context critical in histopathology. Second, our data preprocessing pipeline, including stain normalization and augmentation, enhances the representation of malignant cases and reduces class imbalance. Third, the use of patch-based training preserves high-resolution diagnostic details while minimizing computational costs compared with WSI-level models. These findings suggest that architectures combining efficient convolutional strategies with global context modeling, such as transformers or hybrid models, are better suited for high-resolution pathology image analysis. However, translation into clinical practice still requires validation in full WSI contexts and implementation of interpretability mechanisms to ensure safe and trustworthy deployment.

The integration of AI tools, such as the RAPID framework, into clinical workflows offers substantial

potential to enhance prostate cancer diagnosis and management. In the broader context of pathology, the integration of high-performance deep-learning models, such as ViT, provides several practical advantages, including reducing interobserver variability, a well-documented issue in histopathological diagnosis.³⁶ By providing consistent and reproducible predictions based on learned morphological patterns, AI models serve as a valuable second opinion for pathologists, enhancing diagnostic confidence and accuracy. Moreover, such models can significantly reduce diagnostic turnaround time by pre-screening large volumes of slides, flagging suspicious areas for closer inspection, and triaging routine cases. This is particularly relevant in settings with a shortage of trained pathologists or high workload volume. As highlighted by Steiner et al,³⁷ AI-based assistance systems in pathology not only improve efficiency but also enable broader access to high-quality diagnostic services across institutions with varying resource levels. Such tools are especially valuable in regions with limited access to specialized uropathologists.

For oncology clinicians, timely and accurate histopathological reports are critical for determining treatment strategies, such as active surveillance versus curative therapy. AI-enhanced diagnostics may provide decision support for ambiguous cases, including atypical small acinar proliferation, borderline lesions, or early-stage tumors. Furthermore, integration with electronic medical records and radiological findings could enable AI to contribute to multidisciplinary decision-making, potentially improving patient stratification and treatment personalization.³⁸ To enable clinical adoption, AI systems should be embedded within existing digital pathology infrastructure. WSI scanners paired with AI modules can automate slide triage, generate alerts for atypical findings, and produce structured reports. These tools can also be implemented in remote pathology settings (telepathology), expanding access to expert-level diagnostics across underserved areas.^{36,39}

Ultimately, RAPID has the potential to enhance prostate cancer diagnostic workflows by providing decision support for pathologists and enabling clinicians to make timely, evidence-based decisions, as illustrated in Figure 3. The RAPID digital pathology pipeline begins with preparation of the patient's tissue slide, followed by scanning and digitization into a whole-slide image. Then, the digital image is analyzed through an AI-

integrated application that assists in screening and identifying potential pathological features. However, due to the multiplicity of models tested and the single-center derivation of this dataset, caution is warranted, as near-perfect results may not generalize to external, multi-institutional cohorts. Final diagnosis remains subject to expert pathologist oversight, integrating AI output with microscopic inspection and clinical judgment. Standard safety checkpoints, such as the ability to override AI predictions and perform confirmatory immunohistochemistry as needed, remain essential to maintain diagnostic accuracy and patient safety.

Despite these promising results, this study has some limitations. First, while the use of Indonesian pathology data provides important representation and contextual relevance, the dataset was derived from a single institution with a relatively limited sample size, which may constrain generalizability. Prostatitis cases with marked inflammatory changes were excluded to enhance the model's ability to distinguish malignant from non-malignant cases. Further study is needed to address this limitation. Additionally, the absence of blinded external pathology review, with both assessors from the same institution, may introduce bias. While data were split at the patient/slide level, near-perfect results warrant caution. Future multi-center studies will be required to verify robustness. External validation using independent multi-institutional datasets is essential to ensure adaptability and reliability. Prior studies have highlighted the importance of multi-center data in reducing overfitting and improving model performance across variations in staining protocols and slide scanners.^{35,36} Cross-center validation is critical to minimize institutional bias and increase confidence in clinical deployment.

Second, deep-learning models were developed using transfer learning with pretrained weights from models originally trained on general image datasets, such as ImageNet.⁴⁰ While transfer learning offers practical advantages in computational efficiency and faster convergence, it presents inherent limitations in histopathology. Pretrained feature extractors may not be optimally suited for capturing complex morphological patterns, staining variations, and textural characteristics specific to prostate tissue. Furthermore, the models were not subjected to extensive domain-specific fine-tuning or architectural

customization, which may have constrained their capacity to fully adapt to histological features.

Moreover, this study was limited by its focus on patch-level classification. Although patch-wise training improves efficiency and allows learning from localized features, real-world histopathological diagnosis is performed on entire WSIs. Therefore, the absence of WSI-level evaluation limits the assessment of practical clinical utility.

Future research should explore training models from scratch using large-scale, diverse pathology datasets or applying self-supervised learning techniques that more effectively capture domain-specific features. Additionally, future work should aim to incorporate explainable AI methods and extend model evaluation to full-slide classification. Expanding the dataset to include larger, more diverse cohorts from multiple centers would enhance the generalizability and clinical relevance of the RAPID framework. User-centered evaluations of AI tools in live diagnostic workflows are vital for assessing diagnostic performance, usability, and integration with pathologists' and urologists' routines.³⁷ Real-time testing can help identify operational barriers, optimize model interfaces, and assess the impact on diagnostic turnaround time, interobserver consistency, and clinical decision-making. Together, these efforts will support regulatory readiness and eventual translation of AI-based systems into routine prostate cancer diagnostics. Beyond technical validation, international surveys have emphasized that key enablers of AI adoption include regulatory approval, demonstration of clinical effectiveness, and training, while major barriers remain data privacy, accuracy, and ethical concerns.¹¹

In conclusion, the RAPID diagnostic system effectively distinguished prostate cancer from benign lesions in histopathological images using deep-learning models. Among the architectures evaluated, EfficientNetV2Bo, Xception, and ConvNeXt-Tiny achieved perfect performance, while ViT demonstrated excellent diagnostic balance. These results highlight the effectiveness of both convolutional and transformer-based models in identifying key morphological features in prostate tissues. The RAPID framework presents a data-efficient and scalable approach with the potential to enhance diagnostic reproducibility, provide decision support to pathologists, and improve prostate cancer management.

Conflict of Interest

Agus Rizal Ardy Hariandy Hamid is the editor-in-chief of this journal but was not involved in the review or decision-making process of the article.

Funding Sources

This research received no external funding. No funding agency was involved in the study design; in the collection, management, analysis, or interpretation of data; in manuscript preparation, review, or approval; or in the approval of the decision to publish the manuscript. The authors had complete access to all the data and accept responsibility for its integrity.

Acknowledgment

The authors thank PT Biogen Scientific for technical/material support with slide scanning. The authors also thank Anthony William Brian Iskandar, Darrin Ananda Nugraha, Taufiq Akmal Sungkar, John Christian, Salsa Billa As'syifa, Gilbert Zaini, Kang Heji Dian Pertivi, and Effie Ang Supono for their assistance in extracting histopathology slide images using ImageScope. Nabila Husna Shabrina acknowledges Universitas Multimedia Nusantara for institutional support provided during this study. The acknowledged parties had no involvement in the design of the study; collection of data, analysis, and interpretation; preparation, review, or approval of the manuscript; or in the decision to submit the manuscript.

REFERENCES

1. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al. Global cancer observatory: world [Internet]. International Agency for Research on Cancer; 2024. [cited 2025 Jul 8]. Available from: <https://gco.iarc.who.int/media/globocan/factsheets/populations/900-world-fact-sheet.pdf>.
2. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al. Global cancer observatory: prostate [Internet]. International Agency for Research on Cancer; 2024. [cited 2025 Jul 8]. Available from: <https://gco.iarc.who.int/media/globocan/factsheets/cancers/27-prostate-fact-sheet.pdf>.
3. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al. Global cancer observatory: Indonesia [Internet]. International Agency for Research on Cancer; 2024. [cited 2025 Jul 8]. Available from: <https://gco.iarc.who.int/media/globocan/factsheets/populations/360-indonesia-fact-sheet.pdf>.
4. Ng M, Leslie SW, Baradhi KM. Benign prostatic hyperplasia. [Update 2023 Jun 12] In: StatPearls [Internet]. Treasure Island: StatPearls Publishing; 2025. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK558964/>.
5. Clark R, Vesprini D, Narod SA. The effect of age on prostate cancer survival. *Cancers*. 2022;14(17):4149.
6. Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kusonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol*. 2016;50(6):420–4.
7. Oyama T, Allsbrook WC Jr, Kurokawa K, Matsuda H, Segawa A, Sano T, et al. A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. *Arch Pathol Lab Med*. 2005;129(8):1004–10.
8. Pantanowitz L, Quiroga-Garza GM, Bien L, Heled R, Laifenfeld D, Linhart C, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health*. 2020;2(8):e407–16.
9. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol*. 2019;16(7):391–403.
10. Riaz IB, Harmon S, Chen Z, Naqvi SAA, Cheng L. Applications of artificial intelligence in prostate cancer care: a path to enhanced efficiency and outcomes. *Am Soc Clin Oncol Educ Book*. 2024;44(3):e438516.
11. Ho YT, Dhalas RR, Zohair M, Deb S, Shoaib M, Elmer S, et al. Artificial intelligence in urology—a survey of urology healthcare providers. *Soc Int Urol J*. 2025;6(4):53.
12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296.
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
14. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open*. 2021;11(6):e047709.
15. Maier-Hein L, Reinke A, Kozubek M, Martel AL, Arbel T, Eisenmann M, et al. BIAS: transparent reporting of biomedical image analysis challenges. *Med Image Anal*. 2020;66:101796.
16. Netto GJ, Amin MB, Berney DM, Compérat EM, Gill AJ, Hartmann A, et al. The 2022 World Health Organization classification of tumors of the urinary system and male genital organs—Part B: prostate and urinary tract tumors. *Eur Urol*. 2022;82(5):469–82.
17. Wu W. patchify, version 0.2.3 [Internet]. Python Package Index; 2021. [cited 2025 Jul 8]. Available from: <https://pypi.org/project/patchify/>.
18. Otálora S, Marini N, Podareanu D, Hekster R, Tellez D, Van Der Laak J, et al. Stainlib: a python library for augmentation and normalization of histopathology H&E images. *bioRxiv* (Cold Spring Harbor Laboratory). 2022.
19. Iglavikov VI. alumentations, version 2.0.8 [Internet]. Python Package Index; 2025 [cited 2025 Jul 8]. Available from: <https://pypi.org/project/alumentations/>.
20. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*; 2009 Jun 28–Jul 1; Boston, USA. Piscataway: IEEE; 2009. p. 1107–10.
21. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. *Evol Intell*. 2022;15(1):1–22.
22. Atabansi CC, Nie J, Liu H, Song Q, Yan L, Zhou X. A survey of transformer applications for histopathological image analysis: new developments and future directions. *Biomed Eng Online*. 2023;22(1):96.
23. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016 Jun 27–30; Las Vegas, USA. Piscataway: IEEE; 2016. p. 770–8.
24. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 Jul 21–26; Honolulu, USA. Piscataway: IEEE; 2017. p. 2261–9.
25. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [Internet]. *arXiv preprint arXiv:1704.04861*; 2017 [cited 2025 Jul 8]. Available from: <https://arxiv.org/abs/1704.04861>.
26. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks [Internet]. *arXiv preprint arXiv:1905.11946*; 2019 [cited 2025 Jul 8]. Available from: <https://arxiv.org/abs/1905.11946>.
27. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 Jul 21–26; Honolulu, USA. Piscataway: IEEE; 2017. p. 1800–7.
28. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words:

- transformers for image recognition at scale [Internet]. arXiv preprint arXiv:2010.11929; 2020 [cited 2025 Jul 8]. Available from: <https://arxiv.org/abs/2010.11929>.
29. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H. Training data-efficient image transformers & distillation through attention [Internet]. arXiv preprint arXiv:2012.12877; 2020 [cited 2025 Jul 8]. Available from: <https://arxiv.org/abs/2012.12877>.
 30. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 19–24; New Orleans, USA. Piscataway: IEEE; 2022. p. 11966–76.
 31. Huang TL, Lu NH, Huang YH, Twan WH, Yeh LR, Liu KY, et al. Transfer learning with CNNs for efficient prostate cancer and BPH detection in transrectal ultrasound images. *Sci Rep*. 2023;13(1):21849.
 32. Lin T, Wang Y, Liu X, Qiu X. A survey of transformers. *AI Open*. 2022;3:111–32.
 33. Xu H, Xu Q, Cong F, Kang J, Han C, Liu Z, et al. Vision transformers for computational histopathology. *IEEE Rev Biomed Eng*. 2024;17:63–79.
 34. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol*. 2020;21(2):233–41.
 35. Otálora S, Marini N, Müller H, Atzori M. Combining weakly and strongly supervised learning improves strong supervision in Gleason pattern classification. *BMC Med Imaging*. 2021;21(1):77.
 36. Chaurasia AK, Harris HC, Toohey PW, Hewitt AW. A generalised vision transformer-based self-supervised model for diagnosing and grading prostate cancer using histological images. *Prostate Cancer Prostatic Dis*. 2025;28(1):1–9.
 37. Campanella G, Hanna MG, Geneslaw L, Mirafior A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25(8):1301–9.
 38. Bulten W, Kartasalo K, Chen PC, Ström P, Pinckaers H, Nagpal K, et al. PANDA challenge consortium. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med*. 2022;28(1):154–63.
 39. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am J Surg Pathol*. 2018;42(12):1636–46.
 40. Wang X, Yang S, Zhang J, Wang M, Zhang J, Huang J, et al. TransPath: Transformer-based self-supervised learning for histopathological image classification. In: de Bruijne M, et al, editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science*. Cham: Springer; 2021. p. 186–96.
 41. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2009 Jun 20–25; Miami, USA. Piscataway: IEEE; 2009. p. 248–55.