An artificial intelligence model for detecting pathological lymph node metastasis in prostate cancer using whole slide images: a retrospective, multicentre, diagnostic study

Shaoxu Wu,[a,b,c,j] Yun Wang,[a,j] Guibin Hong,[a,j] Yun Luo,[d,j] Zhen Lin,[e,j] Runnan Shen,[a] Hong Zeng,[f] Abai Xu,[g] Peng Wu,[h] Mingzhao Xiao,[i] Xiaoyang Li,[d] Peng Rao,[g] Qishen Yang,[h] Zhengyuan Feng,[h] Quanhao He,[i] Fan Jiang,[a] Ye Xie,[a] Chengxiao Liao,[a] Xiaowei Huang,[e] Rui Chen,[e] and Tianxin Lin[a,b,c,*]

[a]Department of Urology, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou, China
[b]Guangdong Provincial Key Laboratory of Malignant Tumour Epigenetics and Gene Regulation, Guangdong-Hong Kong Joint Laboratory for RNA Medicine, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou, China
[c]Guangdong Provincial Clinical Research Centre for Urological Diseases, Guangzhou, China
[d]Department of Urology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[e]CellsVision Medical Technology Services Co., Ltd., Guangzhou, China
[f]Department of Pathology, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou, China
[g]Department of Urology, Zhujiang Hospital, Southern Medical University, Guangzhou, China
[h]Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou, China
[i]Department of Urology, First Affiliated Hospital of Chongqing Medical University, Chongqing Medical University, Chongqing, China

eClinicalMedicine 2024;71: 102580
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102580

Summary

Background The pathological examination of lymph node metastasis (LNM) is crucial for treating prostate cancer (PCa). However, the limitations with naked-eye detection and pathologist workload contribute to a high missed-diagnosis rate for nodal micrometastasis. We aimed to develop an artificial intelligence (AI)-based, time-efficient, and high-precision PCa LNM detector (ProCaLNMD) and evaluate its clinical application value.

Methods In this multicentre, retrospective, diagnostic study, consecutive patients with PCa who underwent radical prostatectomy and pelvic lymph node dissection at five centres between Sep 2, 2013 and Apr 28, 2023 were included, and histopathological slides of resected lymph nodes were collected and digitised as whole-slide images for model development and validation. ProCaLNMD was trained at a dataset from a single centre (the Sun Yat-sen Memorial Hospital of Sun Yat-sen University [SYSMH]), and externally validated in the other four centres. A bladder cancer dataset from SYSMH was used to further validate ProCaLNMD, and an additional validation (human-AI comparison and collaboration study) containing consecutive patients with PCa from SYSMH was implemented to evaluate the application value of integrating ProCaLNMD into the clinical workflow. The primary endpoint was the area under the receiver operating characteristic curve (AUROC) of ProCaLNMD. In addition, the performance measures for pathologists with ProCaLNMD assistance was also assessed.

Findings In total, 8225 slides from 1297 patients with PCa were collected and digitised. Overall, 8158 slides (18,761 lymph nodes) from 1297 patients with PCa (median age 68 years [interquartile range 64–73]; 331 [26%] with LNM) were used to train and validate ProCaLNMD. The AUROC of ProCaLNMD ranged from 0.975 (95% confidence interval 0.953–0.998) to 0.992 (0.982–1.000) in the training and validation datasets, with sensitivities > 0.955 and specificities > 0.921. ProCaLNMD also demonstrated an AUROC of 0.979 in the cross-cancer dataset. ProCaLNMD use triggered true reclassification in 43 (4.3%) slides in which micrometastatic tumour regions were initially missed by pathologists, thereby correcting 28 (8.5%) missed-diagnosed cases of previous routine pathological reports. In the human-AI comparison and collaboration study, the sensitivity of ProCaLNMD (0.983 [0.908–1.000]) surpassed that of two junior pathologists (0.862 [0.746–0.939], $P = 0.023$; 0.879 [0.767–0.950], $P = 0.041$) by 10–12% and showed no difference to that of two senior pathologists (both 0.983 [0.908–1.000], both $P > 0.99$). Furthermore, ProCaLNMD significantly boosted the diagnostic sensitivity of two junior pathologists (both $P = 0.041$) to the level of senior pathologists (both $P > 0.99$), and substantially reduced the four pathologists' slide reviewing time (–31%, $P < 0.0001$; –34%, $P < 0.0001$; –29%, $P < 0.0001$; and –27%, $P = 0.00031$).

*Corresponding author. Department of Urology, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, 107th Yanjiangxi Road, Guangzhou 510120, China.
E-mail address: lintx@mail.sysu.edu.cn (T. Lin).
[j]These authors contributed equally.

Interpretation ProCaLNMD demonstrated high diagnostic capabilities for identifying LNM in prostate cancer, reducing the likelihood of missed diagnoses by pathologists and decreasing the slide reviewing time, highlighting its potential for clinical application.

Funding National Natural Science Foundation of China, the Science and Technology Planning Project of Guangdong Province, the National Key Research and Development Programme of China, the Guangdong Provincial Clinical Research Centre for Urological Diseases, and the Science and Technology Projects in Guangzhou.

Keywords: Artificial intelligence; Prostate cancer; Lymph node metastasis; Digital pathology; Multicentre validation

Research in context

Evidence before this study

We conducted a search on PubMed, spanning from the database's inception to Dec 31, 2023, using search terms "artificial intelligence" or "AI", "deep learning" or "DL", "machine learning" or "ML", "lymph node metastasis" or "LNM", and "whole slide image" or "WSI", without language restrictions. Several studies have developed artificial intelligence (AI) models to detect lymph node metastasis (LNM) in various types of cancer, including breast, gastric, colorectal, bladder, and melanoma cancers, using histological slide images. However, no study has reported the application of AI for detecting pathological LNM in prostate cancer.

Added value of this study

In this multicentre study, we developed and validated a prostate cancer lymph node metastasis detector (ProCaLNMD), which is powered by AI and can detect prostate cancer LNM with a sensitivity higher than that of initial routine pathological reports in each centre. To our knowledge, this is the first study to develop an AI-based diagnostic model for detecting pathological LNM in prostate cancer, and it includes most cases of any study that has applied AI for pathological LNM detection. ProCaLNMD showed promising performance in multicentre validation, significantly shortened the slide reviewing time of pathologists, and reduced the missed diagnosis rate of LNM positive patients with prostate cancer by approximately 7.3%.

Implications of all the available evidence

The current process of pathological assessment for LNM in clinical practice is challenging, as it is time-consuming, requires great effort, and carries the risk of overlooking micrometastases. The AI model we developed can be used as an effective and reliable diagnostic tool for LNM detection in prostate cancer. The high sensitivity, reliability, and time-saving properties of ProCaLNMD suggest that it has the potential to significantly reduce the number of missed diagnoses and enhance the efficiency of pathologists.

Introduction

Prostate cancer (PCa) is the second most common cancer globally and the fifth leading cause of cancer-related deaths among men.[1] Lymph node metastasis (LNM), confirmed by pathological examination, occurs in 12–17% of PCa cases[2,3] and is a significant indicator of poor prognosis.[4,5] Patients with PCa and LNM can benefit from adjuvant hormonal therapy and adjuvant radiotherapy,[3,6] emphasising the importance of accurately evaluating LNM postoperatively.

Pathological examination has long been considered as the gold standard for evaluating LNM. However, challenges have arisen in the pathological assessment of LNM in patients with PCa. The high missed-diagnosis rate of nodal micrometastasis (tumour size $\leq$ 2 mm) is a primary challenge. Approximately 8.8–13% of PCa cases with nodal micrometastasis only (pN1mi) have been reported to be missed by routine pathological examination,[4,7] and the prognosis of these neglected patients is significantly worse than that of the LNM-negative (pN-) patients.[7] This challenge can be attributed to the limitations of naked-eye detection. The routine procedure for the pathological LNM evaluation involves manually reviewing histopathological slides by pathologists, which is time-consuming and laborious, with the risk of missing micrometastatic cancer clusters. The second challenge is the critical shortage of pathologists in developing countries.[8] In China, there is less than one pathologist per 100,000 population,[8,9] which intensifies this difficulty. Serial sectioning and immunohistochemistry (IHC) can reduce missed diagnosis of PCa nodal micrometastasis,[4,7] but this increases pathologist workload by approximately 19 times.[10] The shortage of pathologists hinders the routine application of serial sectioning and IHC, thereby increasing the risk of missing nodal micrometastases. The third challenge is associated with easily confused PCa subtypes. The use of neoadjuvant hormonal therapy (NHT) is common in high-risk PCa cases. After NHT, the PCa cells undergo regressive changes,[11] making it difficult to distinguish them from normal lymph tissues. This challenge is also attributed to the foam-like variant subtype of prostatic

acinar adenocarcinoma, which is common and characterised by loose cytoplasm and no significant nuclear atypia.[12,13] Therefore, it is easily confused with normal cells in the lymphatic sinuses, particularly in micrometastasis cases.

To address these challenges, artificial intelligence (AI)-based automated detection models are expected. In recent years, digital pathology has made considerable progress,[14] and AI models based on digitised histopathological slides have been widely used to help improve the workflow of pathological diagnosis.[9,15] Several studies[16–23] have focused on building AI-based models to identify pathological LNM in various types of cancer, including breast, gastric, colorectal, bladder, and melanoma cancers, with satisfactory capabilities. In our study on developing a model for bladder cancer LNM detection,[20] we applied it to identify LNM in PCa in a single-centre, small sample size dataset. Although the area under the receiver operating characteristic curve (AUC) reached 0.922, it was not enough ideal for improving clinical diagnosis in the real world. Therefore, further studies are warranted. To the best of our knowledge, no study has reported the development of an AI-based model to assist in the postoperative evaluation of PCa LNM.

In this study, we developed an AI-based model that analyses whole-slide images (WSIs) to automatically identify metastatic PCa clusters within lymph nodes, with the aim of improving the accuracy and efficiency of the clinical workflow of PCa LNM evaluation. We validated the diagnostic performance of the model using independent multicentre datasets, including more than 12,000 WSIs of 1823 patients from five large medical centres.

Methods

Study design and participants

To develop and validate the AI-based model (Prostate Cancer Lymph Node Metastasis Detector, known as ProCaLNMD), we retrospectively reviewed the data of consecutive patients who underwent radical prostatectomy (RP) and pelvic lymph node dissection (PLND) at five medical centres in China, including Sun Yat-sen Memorial Hospital of Sun Yat-sen University (SYSMH, between Sep 2, 2013, and May 27, 2022), the Third Affiliated Hospital of Sun Yat-sen University (SYUTH, between Dec 10, 2013, and Jun 24, 2022), Nanfang Hospital of Southern Medical University (NFH, between Jan 30, 2018, and Mar 23, 2022), Zhujiang Hospital of Southern Medical University (ZJH, between Feb 20, 2018, and Apr 24, 2022), and the First Affiliated Hospital of Chongqing Medical University (CMUFH, between Jul 5, 2019, and May 6, 2022). We collected histopathological slides of the resected lymph nodes and comprehensive clinical data from these patients. To evaluate ProCaLNMD's ability in detecting LNM for other type of cancer, SYSMH bladder cancer dataset[20] was used as cross-cancer validation dataset. In addition, to further confirm model validity, we conducted an additional study at SYSMH including eligible consecutive patients who underwent RP and PLND between May 30, 2022, to Apr 28, 2023. The geographic locations of involved five centres was shown in the Appendix p 15.

Ethics statement

This study was conducted in accordance with the Declaration of Helsinki and was reported according to the Standards for Reporting Diagnostic Accuracy (STARD, Supplementary table). The research ethics committee of SYSMH approved this study (approval number: SYSKY-2022-162-01) and waived the need for obtaining informed consent due to the retrospective nature of the study.

Development and multicentre validation of ProCaLNMD

The clinical data of the patients, including age, PCa pT stage, pN stage, histological type, Gleason score (International Society of Urological Pathology grade group), and whether NHT had been received, were obtained from the medical record archives of the five centres. The haematoxylin and eosin (H&E)-stained histopathological slides of surgically resected lymph nodes were obtained from the pathology department's archive and scanned as WSIs using two types of digital slide scanners (KF-PRO-120 and SQS-600P) with 20 × objective lenses. The KF-PRO-120 scanner (Ningbo, China) with a specimen-level pixel size of 0.254 μm × 0.254 μm was used to scan slides from SYSMH and CMUFH, while the SQS-600P scanner (Shenzhen, China) with a specimen-level pixel size of 0.206 μm × 0.206 μm was used to scan slides from SYUTH, NFH, and ZJH. Any slides with quality issues, such as bubbles under the cover slide, improper staining, extreme fading, tissue folding, or out-of-focus during scanning, were excluded. It was ensured that every enrolled patient had complete clinical and pathological data collected and no missing data was found.

Patients from SYSMH were divided into training and internal validation datasets in a ratio of 7:3 by setting a cutoff point for the surgery date (Sep 30, 2019). Patients from SYUTH, NFH, ZJH, and CMUFH were assigned to four independent external validation datasets (Fig. 1).

The reference standard for WSIs was determined by the consensus of two senior pathologists, each with >15 years of experience, recruited from SYSMH. This was done because a high missed-diagnosis rate of routine pathological reports for nodal micrometastasis has been reported.[4,7] The two experts independently reviewed and classified the WSIs as positive (with LNM) or negative (without LNM), and if there was any inconsistency in classification between them or any of them required IHC, the final decision was made using IHC. Either

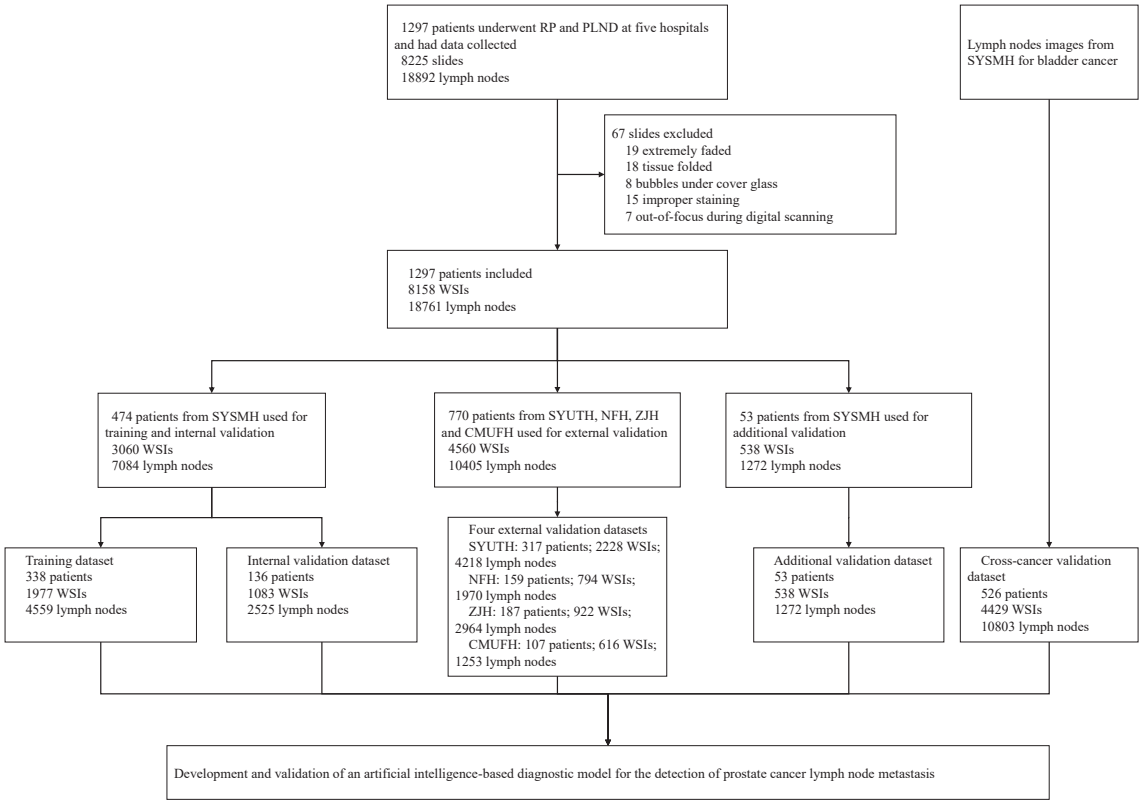

***Fig. 1:* Study profile.** SYSMH = Sun Yat-sen Memorial Hospital of Sun Yat-sen University. SYUTH = The Third Affiliated Hospital of Sun Yat-sen University. NFH = Nanfang Hospital of Southern Medical University. ZJH = Zhujiang Hospital of Southern Medical University. CMUFH = The First Affiliated Hospital of Chongqing Medical University. RP = radical prostatectomy. PLND = pelvic lymph node dissection. WSIs = whole-slide images.

pathologist was blind to the clinical information of the patients. The reference standard was constructed before developing the AI model.

We used DeepLabv3+[24] as the segmentation model framework and RegNet_Y40[25] as the encoder within this framework. Python software (version 3.7.0) was used to develop the model. The encoder was initialised using ImageNet-pretrained parameters. The two senior pathologists, who created the reference standard, drew pixel-level annotation to train the model (Fig. 2), and further details can be found in the Appendix (pp 2, 8). To train the model, Radam optimizer[26] with a weight decay of 0.0001 and a learning rate of 0.0001 was used. CosineAnnealingWarmupRestarts (available at https://github.com/katsura-jp/pytorch-cosine-annealing-with-warmup) was employed for learning rate scheduling, with first_cycle_steps = 20, max_lr = 0.0001, min_lr = 0.00001, and warmup_steps = 5. Focal Loss[27] was used as the loss function with $\alpha = 0.75$ and $\gamma = 2$. We trained with a batch size of 6 for 100 epochs, generating 1200 samples per epoch on two GeForce RTX 3090 GPUs. The final model was selected based on the highest mean intersection over union in the segmentation validation set (Appendix p 6). During the validation phase, WSIs were slid into numerous patch images, and invalid regions devoid of tissues were excluded. ProCaLNMD then made predictions on all valid patches, yielding $1 \times 2048 \times 2048$ result masks as 8-bit grayscale images with pixel values ranging from 0 to 255.

The WSI-level classification was accomplished through a thresholding strategy (Appendix p 7), and a heatmap was generated by colouring pixels based on the output values. Low values were represented in blue, while high values were represented in red (Fig. 2, Appendix p 8). Detailed descriptions of data augmentation, the strategy to handle data imbalance and hard samples, and binary classification strategy with two threshold variables are provided in the Appendix (pp 2–4). The flowchart of ProCaLNMD development is shown in Fig. 2.

Additional validation: human-AI comparison and collaboration

To compare the diagnostic performance of ProCaLNMD with that of pathologists and evaluate the effectiveness of ProCaLNMD in assisting pathologists in reviewing

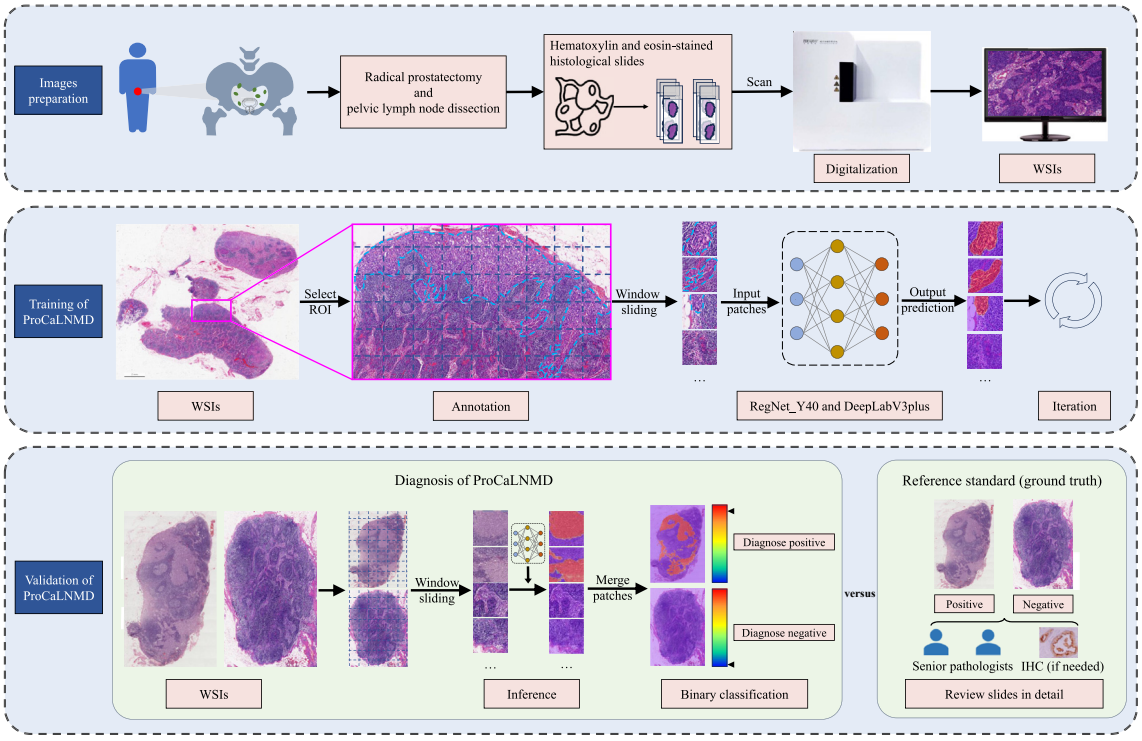

Fig. 2: **Flowchart of development of ProCaLNMD.** WSI = whole-slide image. ROI = region of interest. AI = artificial intelligence. IHC = immunohistochemistry. ProCaLNMD = prostate cancer lymph node metastasis detector.

slides, an additional validation was conducted on 53 consecutive patients from SYSMH. Four pathologists (two seniors with 7 years of experience, and two juniors with 3 years of experience) were recruited from SYSMH to review histopathological slides of resected lymph nodes. The slides were collected, quality-controlled, and assigned for review, following the routine pathological assessment process. On each working day, each pathologist was required to review all slides of four patients, with two being reviewed in morning and the remaining in afternoon. The order in which the 53 patients presented was randomly generated and unique to each pathologist, and they were blinded to any information regarding the patients they were reviewing. Next, the slides from all 53 patients were scanned as WSIs using a KF-PRO-120 scanner with 20 × objective lenses and submitted to ProCaLNMD for independent diagnosis. After a 4-week washout period, the pathologists reviewed the WSIs again, this time with the assistance of ProCaLNMD, which highlighted suspicious areas of cancer clusters. The order in which the patients were presented to the pathologists in the AI-assisted mode was decided randomly and differed from that in the non-AI-assisted mode. The time taken by each pathologist to review each patient in both two modes was recorded.

Prior to reviewing, the reference standard was constructed in the same manner as that in the multicentre validation. The four pathologists were blinded to the reference standard during the study.

Outcomes

The primary endpoint of this study was the AUC of ProCaLNMD. Secondary endpoints were sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Additionally, the study compared the performance of pathologists with and without the use of ProCaLNMD. The present study was exploratory and no multiplicity adjustment was applied for multiple endpoints.

Statistical analysis

All statistical analyses were performed using R software (version 4.3.1). No ex-ante sample size calculation was performed due to the exploratory nature of the present study and the data-driven nature of AI. The ggplot2 package (version 3.4.2) was used to draw the violin plot. We employed the receiver operating characteristic (ROC) curve to demonstrate the diagnostic ability of ProCaLNMD, and calculated the AUC using the pROC package (version 1.18.2). Threshold values for the binary classification of ProCaLNMD were based on the optimal F_2 score. The 95% confidence intervals (CI) for sensitivity, specificity, PPV, and NPV were calculated using the Clopper-Pearson method because multiple indexes were close to 1. We used the paired χ^2 test (McNemar's

test) to compare the diagnostic sensitivity of different operators (ProCaLNMD, different pathologists, and pathologists with and without AI assistance). We used the Wilcoxon signed rank test to analyse the difference in the reviewing time of pathologists per patient in the non-AI-assisted and AI-assisted modes. All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant.

Role of the funding source

The funders were not involved in the study design, data collection, data analysis, data interpretation, or manuscript writing. All authors reviewed the manuscript, approved the submitted version, had full access to all the data reported in the study, and had final responsibility for the decision to submit the manuscript for publication.

Results

In this retrospective, multicentre, diagnostic study, 8225 slides from 1297 patients with PCa (median age 68 [interquartile range 64–73] years; 331 [26%] with LNM) were collected. After quality control, 67 (0.8%) slides were excluded due to low quality (19, extremely faded; 18, tissue folded; 15, improper staining; 8, bubbles under cover glass; and 7, out-of-focus during scanning). There was no missing data in the present study. For the development of ProCaLNMD, 1977 WSIs of 338 patients from SYSMH (between Sep 2, 2013, and Sep 30, 2019) were included in the training dataset, and 1083 WSIs of 136 patients from SYSMH (between Oct 1, 2019 and May 27, 2022) were included in the internal validation dataset. The external validation datasets included 2228 WSIs of 317 patients from SYUTH, 794 WSIs of 159 patients from NFH, 922 WSIs of 187 patients from ZJH, and 616 WSIs of 107 patients from CMUFH. To further validate ProCaLNMD, 538 WSIs of 53 patients from SYSMH (between May 30, 2022, and Apr 28, 2023) were included in the additional validation dataset. In addition, 4429 WSIs of 526 patients with bladder cancer from SYSMH (between Jan 1, 2013, and Dec 31,2021) were included to validate ProCaLNMD's ability in detecting LNM in other type of cancer.

Overall, 8158 WSIs of 1297 patients (18,761 lymph nodes) with PCa from five independent medical centres were used to develop and validate ProCaLNMD. The study profile is presented in Fig. 1, and the characteristics of patients with PCa are listed in Table 1.

The AUC of ProCaLNMD varied from 0.975 (95% CI: 0.953–0.998) to 0.992 (0.982–1.000) in the training, internal validation, and external validation datasets (Fig. 3A). As the threshold values were determined based on the optimal F_2 score, ProCaLNMD sensitivity reached 0.986 (139/141, 0.950–0.998), 0.974 (75/77, 0.909–0.997), 0.978 (312/319, 0.955–0.991), 0.982 (167/170, 0.949–0.996), 0.957 (135/141, 0.910–0.984), 0.955 (84/88, 0.888–0.987), and 0.983 (57/58, 0.908–1.000) in the training, internal validation, SYUTH, NFH, ZJH, CMUFH, and additional validation datasets, respectively. Meanwhile, the specificities in all seven PCa datasets were >0.920 (Table 2). In combined all external validation datasets, ProCaLNMD demonstrated an AUC of 0.983 (0.977–0.990, Appendix p 14), a sensitivity of 0.972 (698/718, 0.957–0.983), and a specificity of 0.932 (3582/3842, 0.924–0.940, Table 2).

The diagnostic performance of ProCaLNMD was validated in various subgroups. For subgroups of patients with different pT stages, Gleason scores, and those who received NHT before surgery, ProCaLNMD AUCs ranged from 0.953 to 1.000, 0.972 to 1.000, and 0.962 to 0.988, respectively, in the five centres (Fig. 3B). For subgroups of pN + patients with a foam-like variant subtype, ProCaLNMD exhibited a sensitivity comparable to that of pathological reports (0.967 vs 0.958, $P = 0.80$, Appendix p 10). The output heatmaps showed that ProCaLNMD was able to accurately identify nodal metastatic cancer clusters of the foam-like subtype (Appendix p 10) and those with degenerative changes after NHT (Appendix p 11). Additionally, ProCaLNMD demonstrated an AUC of 0.979 in identifying bladder cancer LNM, with a sensitivity of 0.963 and a specificity of 0.955 (Appendix p 16).

In the additional validation, ProCaLNMD displayed a sensitivity of 0.983 (57/58, 95% CI 0.908–1.000), comparable to that of the two senior pathologists (both 0.983, 57/58, 0.908–1.000, $P > 0.99$) and higher than that of junior pathologist 1 (0.862, 50/58, 0.746–0.939, $P = 0.023$) and junior pathologist 2 (0.879, 51/58, 0.767–0.950, $P = 0.041$). In the human-AI collaboration phase, ProCaLNMD helped reduce the number of missed positive slides by 6 (6/8, 75%) for junior pathologist 1, 6 (6/7, 85.7%) for junior pathologist 2, 1 (1/1, 100%) for senior pathologists 1, and 1 (1/1, 100%) for senior pathologist 2 (Appendix p 12). ProCaLNMD use improved the sensitivity exhibited by pathologists (junior pathologist 1: from 0.862 without AI assistance to 0.966 with AI assistance, +10.4%, $P = 0.041$; junior pathologist 2: from 0.879 to 0.983, +10.4%, $P = 0.041$; senior pathologists 1 & 2: both from 0.983 to 1.000, +1.7%, $P > 0.99$; Fig. 3C and Table 2). With ProCaLNMD assistance, the sensitivity of junior pathologists 1 (0.966, 56/58, 0.881–0.996) and 2 (0.983, 57/58, 0.908–1.000) were both comparable (both $P > 0.99$) to that of senior pathologists without AI assistance (both 0.983, 57/58, 0.908–1.000). The reclassification tables showed that all four pathologists had a net reclassification index > 0 (Appendix p 12), indicating improved performance with AI assistance. In addition, with the assistance of ProCaLNMD, the mean reviewing time per patient was significantly reduced for all four pathologists (for junior 1, from 791s without assistance to 542s with assistance, –31.4%, $W = 2231$, $P < 0.0001$; for junior 2, from 753s to 501s, –33.5%, $W = 2102$, $P < 0.0001$; for senior 1, from 584s to 414s, –29.1%, $W = 2057$,

	Training dataset (n = 338)	Internal validation dataset (n = 136)	External validation dataset 1 (n = 317)	External validation dataset 2 (n = 159)	External validation dataset 3 (n = 187)	External validation dataset 4 (n = 107)	Additional validation dataset (n = 53)
Age, years	67 (63–72)	67 (63–72)	69 (64–74)	68 (64–74)	69 (64–73)	69 (65–73)	69 (64–72)
pT stage							
pT2	182 (54%)	49 (36%)	136 (43%)	75 (47%)	104 (56%)	48 (45%)	20 (38%)
pT3a	60 (18%)	36 (26%)	67 (21%)	16 (10%)	38 (20%)	32 (30%)	13 (25%)
pT3b	91 (27%)	50 (37%)	112 (35%)	58 (36%)	38 (20%)	27 (25%)	18 (34%)
pT4	5 (1%)	0	0	7 (4%)	5 (3%)	0	1 (2%)
Others[a]	0	1 (1%)	2 (1%)	3 (2%)	2 (1%)	0	1 (2%)
pN stage[b]							
pN0	276 (82%)	105 (77%)	224 (71%)	111 (70%)	134 (72%)	79 (74%)	37 (70%)
pN1mi	23 (7%)	7 (5%)	29 (9%)	12 (8%)	24 (13%)	8 (7%)	5 (9%)
pN1	39 (12%)	24 (18%)	64 (20%)	36 (23%)	29 (16%)	20 (19%)	11 (21%)
Histological type							
Acinar adenocarcinoma	325 (96%)	129 (95%)	313 (99%)	154 (97%)	184 (98%)	106 (99%)	51 (96%)
Others	13 (4%)	7 (5%)	4 (1%)	5 (3%)	3 (2%)	1 (1%)	2 (4%)
Gleason score (ISUP grade group)							
1	45 (13%)	5 (4%)	34 (11%)	10 (6%)	18 (10%)	0	0
2	72 (21%)	30 (22%)	54 (17%)	30 (19%)	31 (17%)	14 (13%)	7 (13%)
3	82 (24%)	40 (29%)	60 (19%)	24 (15%)	27 (14%)	28 (26%)	14 (26%)
4	42 (12%)	17 (13%)	55 (17%)	18 (11%)	46 (25%)	17 (16%)	6 (11%)
5	59 (17%)	26 (19%)	104 (33%)	71 (45%)	60 (32%)	46 (43%)	12 (23%)
Others[c]	38 (11%)	18 (13%)	10 (3%)	6 (4%)	5 (3%)	2 (2%)	14 (26%)
Neoadjuvant hormonal therapy							
Yes	149 (44%)	66 (49%)	84 (26%)	42 (26%)	90 (48%)	44 (41%)	26 (49%)
No	189 (56%)	70 (51%)	233 (74%)	117 (74%)	97 (52%)	63 (59%)	27 (51%)

Data are median (interquartile range) or n (%). The training dataset, internal validation dataset and additional validation dataset were from Sun Yat-sen Memorial Hospital of Sun Yat-sen University [SYSMH]. The external validation dataset 1 was from The Third Affiliated Hospital of Sun Yat-sen University [SYUTH]. The external validation dataset 2 was from Nanfang Hospital of Southern Medical University [NFH]. The external validation dataset 3 was from Zhujiang Hospital of Southern Medical University [ZJH]. The external validation dataset 4 was from The First Affiliated Hospital of Chongqing Medical University [CMUFH]. ISUP = International Society of Urological Pathology. [a]Pathological complete response (pCR) or majority remission after neoadjuvant hormonal therapy (NHT). [b]pN1mi: the maximum diameters of all metastatic cancer lesions in lymph nodes are less than 2 mm; pN1: at least one metastatic cancer lesion with a maximum diameter greater than 2 mm. [c]pCR or massive histological reaction to NHT. For the above two situations, pathological report did not provide Gleason score.

Table 1: **Baseline characteristics of included patients with PCa.**

$P < 0.0001$; for senior 2, from 617s to 450s, –27.0%, $W = 1976$, $P = 0.00031$; Fig. 3D).

Further, ProCaLNMD identified 28 patients with nodal micrometastasis with a missed diagnosed in five medical centres (Appendix p 9). Specifically, in five centres, ProCaLNMD identified 43 positive slides that were missed by the pathological report, 13 of which belonged to seven pN + patients correctly diagnosed by pathological reports (pathologists missed a few of these patients' several positive slides). The remaining 30 slides belonged to 28 "LNM-negative" patients who were previously falsely diagnosed by the pathological reports. The pathological reports of each centre missed 8 of 109 (7.34%, 95% CI 3.22–14.0) pN + patients in SYSMH, 6 of 93 (6.45%, 2.40–13.5) in SYUTH, 5 of 48 (10.4%, 3.47–22.7) in NFH, 8 of 53 (15.1%, 6.75–27.6) in ZJH, and 1 of 28 (3.57%, 0.0904–18.3) in CMUFH. Correspondingly, ProCaLNMD failed to identify 2 of 109 (1.83%, 95% CI 0.223–6.47) pN + patients in SYSMH, 1 of 93 (1.08%, 0.0272–5.85) in SYUTH, and 1 of 53 (1.89%, 0.0478–10.1) in ZJH. No pN + patients were missed by ProCaLNMD in NFH and CMUFH. Overall, 8.46% (28/331, 5.69–12.0) pN + patients were missed by the pathological report, while ProCaLNMD missed 1.21% (4/331, 0.330–3.07), with a significant difference ($P < 0.0001$, Appendix p 9). The clinical characteristics and follow-up details of the 28 missed pN + patients (all were pN1mi) from five centres are shown in the Appendix p 5. Ten patients (10/28, 35.7%, 18.6–55.9) experienced recurrence within 5 years after surgery.

Discussion

We developed an AI-based model, called ProCaLNMD, for automatically detecting of metastatic PCa clusters in lymph nodes. We trained and validated this model using 12,587 WSIs from 1823 patients across five large medical centres and evaluated its usefulness for incorporating into the clinical workflow. The results showed that ProCaLNMD demonstrated high sensitivity and specificity in multiple external centres, outperformed the pathological reports performance in sensitivity, and showed no difference to senior pathologists. In addition,

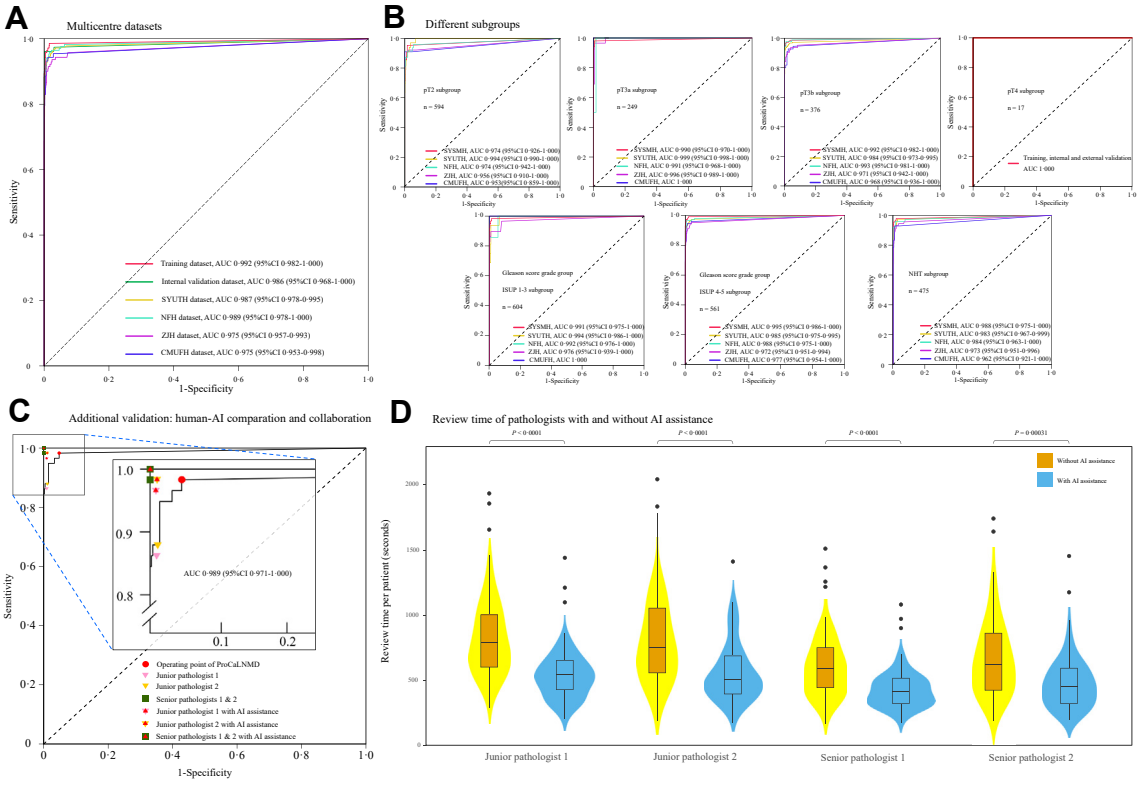

Fig. 3: **Diagnostic performance of ProCaLNMD and pathologists in different datasets.** (A) ROC curves of ProCaLNMD in training, internal and external validation datasets. (B) ROC curves of ProCaLNMD in different subgroups of training, internal and external validation datasets. (C) Diagnostic performance of ProCaLNMD and pathologists of different seniorities (with and without AI assistance) in the additional validation. Given that the performances of senior pathologists 1 and 2 were consistent both in AI-assisted mode and non-AI-assisted mode, their diagnostic performances were represented by the same points in the ROC curve. (D) The average reviewing time per patient of the pathologists with and without AI assistance in the additional validation. Training dataset and internal validation dataset were from Sun Yat-sen Memorial Hospital of Sun Yat-sen University [SYSMH]. SYUTH = the Third Affiliated Hospital of Sun Yat-sen University. NFH = Nanfang Hospital of Southern Medical University. ZJH = Zhujiang Hospital of Southern Medical University. CMUFH = the First Affiliated Hospital of Chongqing Medical University. ROC curve = receiver operating characteristic curve. AUC = area under the ROC curve. CI = confidence interval. NHT = neoadjuvant hormonal therapy. ISUP = International Society of Urological Pathology. ProCaLNMD = prostate cancer lymph node metastasis detector.

ProCaLNMD could assist pathologists, regardless of their seniority, by improving their diagnostic performance and substantially reducing their slide reviewing time. Moreover, not only for PCa LNM, ProCaLNMD exhibited satisfactory ability in detecting bladder cancer LNM. To our knowledge, this is the first study to develop an AI-based model for the automated pathological evaluation of PCa LNM, and it also includes the largest numbers of cases in the application of AI for evaluating the pN stage of cancer.

Nodal micrometastases are common in PCa and are easily missed during pathological assessment. ProCaLNMD exhibited excellent sensitivity in detecting nodal micrometastases in PCa, surpassing that of pathologists who issued the routine pathological reports. This was demonstrated in identifying 8.5% pN + patients (all pN1mi) who were previously missed in the pathological assessment. Within 5 years after surgery, 36% of these neglected patients experienced recurrence, which was higher than that previously reported for pN0 patients (19%).[4] ProCaLNMD achieved a significantly higher sensitivity compared with that of routine pathological reports. Therefore, it can be used in clinical applications for pathological LNM evaluation to significantly reduce the missed-diagnosis rate by approximately 7.3%, and ultimately improve the prognosis of patients with PCa.

The robustness of ProCaLNMD was demonstrated through its performance in various subgroups, cross-cancer and multicentre datasets. The model showed satisfactory diagnostic capabilities in patients with PCa in different pT stages, Gleason scores, pN + cases with foam-like subtype, and those who underwent NHT before surgery, and even patients with bladder cancer.

	TP	FN	TN	FP	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95%CI)	NPV (95%CI)
Multicentre validation of ProCaLNMD								
Training dataset	139	2	1780	44	0.986 (0.950–0.998)	0.976 (0.968–0.982)	0.760 (0.691–0.820)	0.999 (0.996–1.000)
Internal validation dataset	75	2	987	31	0.974 (0.909–0.997)	0.970 (0.957–0.979)	0.708 (0.611–0.792)	0.998 (0.993–1.000)
External validation dataset 1	312	7	1778	131	0.978 (0.955–0.991)	0.931 (0.919–0.942)	0.704 (0.659–0.746)	0.996 (0.992–0.998)
External validation dataset 2	167	3	575	49	0.982 (0.949–0.996)	0.921 (0.898–0.941)	0.773 (0.711–0.827)	0.995 (0.985–0.999)
External validation dataset 3	135	6	724	57	0.957 (0.910–0.984)	0.927 (0.906–0.944)	0.703 (0.633–0.767)	0.992 (0.982–0.997)
External validation dataset 4	84	4	505	23	0.955 (0.888–0.987)	0.956 (0.935–0.972)	0.785 (0.700–0.859)	0.992 (0.980–0.998)
All external validation datasets	698	20	3582	260	0.972 (0.957–0.983)	0.932 (0.924–0.940)	0.729 (0.699–0.757)	0.994 (0.991–0.997)
Additional validation								
ProCaLNMD	57	1	457	23	0.983 (0.908–1.000)	0.952 (0.929–0.969)	0.713 (0.600–0.808)	0.998 (0.988–1.000)
Junior pathologist 1	50	8	476	4	0.862 (0.746–0.939)	0.992 (0.979–0.998)	0.926 (0.821–0.979)	0.983 (0.968–0.993)
Junior pathologist 2	51	7	475	5	0.879 (0.767–0.950)	0.990 (0.976–0.997)	0.911 (0.804–0.970)	0.985 (0.970–0.994)
Senior pathologist 1	57	1	480	0	0.983 (0.908–1.000)	1	1	0.998 (0.988–1.000)
Senior pathologist 2	57	1	480	0	0.983 (0.908–1.000)	1	1	0.998 (0.988–1.000)
Junior pathologist 1 with ProCaLNMD assistance	56	2	476	4	0.966 (0.881–0.996)	0.992 (0.979–0.998)	0.933 (0.838–0.982)	0.996 (0.985–1.000)
Junior pathologist 2 with ProCaLNMD assistance	57	1	475	5	0.983 (0.908–1.000)	0.990 (0.976–0.997)	0.919 (0.822–0.973)	0.998 (0.988–1.000)
Senior pathologist 1 with ProCaLNMD assistance	58	0	480	0	1	1	1	1
Senior pathologist 2 with ProCaLNMD assistance	58	0	480	0	1	1	1	1

The training, internal and additional validation dataset were from Sun Yat-sen Memorial Hospital of Sun Yat-sen University [SYSMH]. The external validation dataset 1 was from the Third Affiliated Hospital of Sun Yat-sen University [SYUTH]. The external validation dataset 2 was from Nanfang Hospital of Southern Medical University [NFH]. The external validation dataset 3 was from Zhujiang Hospital of Southern Medical University [ZJH]. The external validation dataset 4 was from the First Affiliated Hospital of Chongqing Medical University [CMUFH]. TP = true positive, FN = false negative, TN = true negative, FP = false positive, PPV = positive predicative value, NPV = negative predicative value, CI = confidence interval, ProCaLNMD = prostate cancer lymph node metastasis detector.

Table 2: **The diagnostic performance of ProCaLNMD and pathologists in different datasets.**

Moreover, it performed well in multiple external centres, achieving high sensitivity and specificity, which is challenging for AI models aimed at detecting pathological LNM. Several studies[16–23] have reported the development of AI models for LNM detection on histopathological slide images. However, few[20,23] of these previous studies achieved sensitivities exceeding 0.95 under the premise of specificities no less than 0.90 in independent external validation datasets. In this study, ProCaLNMD was validated in four independent external datasets containing 4560 WSIs (10,405 lymph nodes) from 770 patients, with sensitivity > 0.954 and specificity > 0.920 in each dataset. The favourable generalizability of ProCaLNMD may be attributed to the following factors. First, the training dataset were collected from a 7-year consecutive cohort from SYSMH, a high-volume tertiary hospital in China, covering diverse representations of men with PCa of Chinese population. Second, we adopted supervised learning strategy with pixel-level detailed manual annotation to optimize diagnostic performance. Third, ProCaLNMD was trained using the "human-in-the-loop"[28] principle. Specifically, a dynamic training scheme, which allowed senior pathologists to get real-time insights into complex samples that confuse the AI model and offer timely expert-knowledge guidance, was applied during training phase. In short, the consistent diagnostic performance of ProCaLNMD was satisfactorily demonstrated, indicating its reliability and general applicability.

The potential utility of integrating ProCaLNMD into the clinical workflow was further validated. ProCaLNMD assisted pathologists, regardless of their seniority and especially for juniors, in identifying previously missed positive slides, resulting in an improvement in sensitivity ranging from 1.7% to 10.3%. In addition, with the assistance of ProCaLNMD, pathologists spent significantly less time (27–34%) reviewing the slides compared with that without assistance. With the help of ProCaLNMD's time-effective capabilities, serial sectioning, which can reduce the risk of missed-diagnosis of nodal micrometastasis,[29] is expected to become a routine practice. Currently, owing to the relative lack of pathologists and heavy workload of pathological assessment, only one to two sections of each formalin-fixed, paraffin-embedded lymph node are routinely used for H&E-staining and pathological evaluation, increasing the risk of missing metastatic cancer in the rest of the lymph nodes.[29] However, the labour time required for serial sectioning significantly increases.[10] The application of ProCaLNMD is promising in resolving this contradiction, reducing the risk of overlooking micrometastasis, and improving survival outcomes in patients with PCa.

In the real-world clinical practice, ProCaLNMD would serve as a useful tool to help pathologists quickly notice suspicious areas of metastatic cancer, and assist in substantially reducing the slide reviewing time. Given the high sensitivity and relatively low PPV of ProCaLNMD, pathologists would need to review all positive alarms identified by ProCaLNMD to identify true positive from false positive. Further, we analysed the false

alarms generated by ProCaLNMD and found that they can be classified into four categories: vascular walls, small stains or foreign objects, high endothelial venules in the lymph nodes, and normal cells containing loose cytoplasm in the lymph sinuses (Appendix p 13). Pathologists, regardless of their seniority, readily recognised these false alarms, and little burden was caused for pathologists accordingly.

Our study has some limitations. First, all slides used in our study were collected and analysed retrospectively, which may lead to a degree of selection bias. Still, the favourable performance of ProCaLNMD in multicentre consecutive cohorts (3–10-year) indicate that the bias is to some extent not prominent. Second, given that some subgroups only had a small number of cases, there is still room for improvement in the present study sample size. Therefore, further prospective validation involving more centres and larger sample size is required. Third, while we used high-resolution WSIs produced by whole-slide imaging scanners to develop ProCaLNMD, digital microscopes are more commonly used in resource-limited regions of developing countries. Hence, the development of AI models based on affordable devices is to be explored. Fourth, while ProCaLNMD was trained and validated in multiple large Chinese cohorts, its applicability to other regional populations remains to be tested.

In summary, ProCaLNMD can decrease the number of missed diagnoses in pN + patients with PCa by approximately 7.3% compared with that of routine pathological reports. Furthermore, it can enhance the accuracy and efficiency of pathological evaluations of PCa LNM. The effectiveness and generalisability of ProCaLNMD have been confirmed via multicentre external validation, making it a valuable tool in clinical settings.

Contributors

TL, SW, YW, GH, YL, and ZL conceived and designed the study. RS, AX, PW, MX, XL, PR, QY, ZF, and QH procured the data. HZ, RS, FJ, YX, CL, XH and RC screened the patients and images. HZ curated the pathological examinations. ZL trained and developed the AI model. SW, YW, GH, and ZL performed data analysis and interpretation and wrote the original draft of the manuscript. TL, SW, YW, and GH accessed and verified the data. TL supervised the study.

Data sharing statement

To protect patient privacy, pathological slide images and other patient-related data are not publicly accessible. However, all data are available upon reasonable request emailed to the corresponding author.

Declaration of interests

All authors declare no competing interests.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant numbers U21A20383, 81825016, 82341018, and 82003151), the National Key Research and Development Programme of China (2018YFA0902803), the Science and Technology Planning Project of Guangdong Province (2023B1212060013, 2018B010109006), the Guangdong Provincial Clinical Research Centre for Urological Diseases (2020B1111170006), and the Science and Technology Projects in Guangzhou (2024A03J1190). We thank Prof. Phei-Er Saw for providing assistance in editing this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102580.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
2 Lestingi JFP, Guglielmetti GB, Trinh QD, et al. Extended versus limited pelvic lymph node dissection during radical prostatectomy for intermediate- and high-risk prostate cancer: early oncological outcomes from a randomized phase 3 trial. *Eur Urol*. 2021;79(5):595–604.
3 Zuur LG, de Barros HA, van der Mijn KJC, et al. Treating primary node-positive prostate cancer: a scoping review of available treatment options. *Cancers*. 2023;15(11):2962.
4 Maxeiner A, Grevendieck A, Pross T, et al. Lymphatic micrometastases predict biochemical recurrence in patients undergoing radical prostatectomy and pelvic lymph node dissection for prostate cancer. *Aktuelle Urol*. 2019;50(6):612–618.
5 Perera M, Lebdai S, Tin AL, et al. Oncologic outcomes of patients with lymph node invasion at prostatectomy and post-prostatectomy biochemical persistence. *Urol Oncol*. 2023;41(2):105.e19–105.e23.
6 Touijer KA, Karnes RJ, Passoni N, et al. Survival outcomes of men with lymph node-positive prostate cancer after radical prostatectomy: a comparative analysis of different postoperative management strategies. *Eur Urol*. 2018;73(6):890–896.
7 Pagliarulo V, Hawes D, Brands FH, et al. Detection of occult lymph node metastases in locally advanced node-negative prostate cancer. *J Clin Oncol*. 2006;24(18):2735–2742.
8 Wilson ML, Fleming KA, Kuti MA, Looi LM, Lago N, Ru K. Access to pathology and laboratory medicine services: a crucial gap. *Lancet*. 2018;391(10133):1927–1938.
9 Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol*. 2020;21(2):222–232.
10 Schilling D, Hennenlotter J, Gakis G, et al. Prospective assessment of histological serial sectioning of pelvic lymph nodes in prostate cancer: a cost-benefit analysis. *BJU Int*. 2012;110(6 Pt B):E166–E171.
11 Wang X, Zhang J, Han B. Neoadjuvant hormonal therapy for prostate cancer: morphologic features and predictive parameters of therapy response. *Adv Anat Pathol*. 2022;29(4):252–258.
12 Hudson J, Cao D, Vollmer R, Kibel AS, Grewal S, Humphrey PA. Foamy gland adenocarcinoma of the prostate: incidence, Gleason grade, and early clinical outcome. *Hum Pathol*. 2012;43(7):974–979.
13 Madakshira MG. Foamy gland variant of prostate adenocarcinoma. *Med J Armed Forces India*. 2023;79(2):241–243.
14 Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20(5):e253–e261.
15 Kers J, Bülow RD, Klinkhammer BM, et al. Deep learning-based classification of kidney transplant pathology: a retrospective, multicentre, proof-of-concept study. *Lancet Digit Health*. 2022;4(1):e18–e26.
16 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199–2210.
17 Hu Y, Su F, Dong K, et al. Deep learning system for lymph node quantification and metastatic cancer identification from whole-slide pathology images. *Gastric Cancer*. 2021;24(4):868–877.
18 Wang X, Chen Y, Gao Y, et al. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. *Nat Commun*. 2021;12(1):1637.
19 Huang SC, Chen CC, Lan J, et al. Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. *Nat Commun*. 2022;13(1):3347.
20 Wu S, Hong G, Xu A, et al. Artificial intelligence-based model for lymph node metastases detection on whole slide images in bladder cancer: a retrospective, multicentre, diagnostic study. *Lancet Oncol*. 2023;24(4):360–370.

21 Chuang WY, Chen CC, Yu WH, et al. Identification of nodal micrometastasis in colorectal cancer using deep learning on annotation-free whole-slide images. *Mod Pathol*. 2021;34(10):1901–1911.
22 Jansen P, Baguer DO, Duschner N, et al. Deep learning detection of melanoma metastases in lymph nodes. *Eur J Cancer*. 2023;188:161–170.
23 Khan A, Brouwer N, Blank A, et al. Computer-assisted diagnosis of lymph node metastases in colorectal cancers using transfer learning with an ensemble model. *Mod Pathol*. 2023;36(5):100118.
24 Chen LC, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. *arXiv*. 2017. publish online December 5. 1706.05587 (preprint).
25 Xu J, Pan Y, Pan X, Hoi S, Yi Z, Xu Z. RegNet: self-regulated network for image classification. *IEEE Trans Neural Netw Learn Syst*. 2023;34(11):9562–9567.
26 Liu LY, Jiang HM, He PC, et al. On the variance of the adaptive learning rate and beyond. *arXiv*. 2021. publish online October 26. 1908.03265 (preprint).
27 Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell*. 2020; 42(2):318–327.
28 Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal A. Human-in-the-loop machine learning: a state of the art. *Artif Intell Rev*. 2023;56(4):3005–3054.
29 Schiavina R, Capizzi E, Borghesi M, et al. Nodal occult metastases in intermediate- and high-risk prostate cancer patients detected using serial section, immunohistochemistry, and real-time reverse transcriptase polymerase chain reaction: prospective evaluation with matched-pair analysis. *Clin Genitourin Cancer*. 2015;13(2):e55–e64.