Artificial intelligence-based pathological model for pan-cancer lymph node metastasis detection: a multicentre diagnostic study with retrospective and prospective validation

Shaoxu Wu, Guibin Hong*, Yun Wang*, Hong Zeng*, Zhen Lin*, Jie Yang*, Jianning Chen*, Hongtao Chen*, Yan Huang, Liangli Hong, Yuanyuan Wang, Weineng Feng, Yunjie Zeng, Huasheng Huang, Chunhui Wang, Chen Xing, Gehong Dong, Xing Hua, Changhai Qi, Aijun Liu, Jihua Wu, Xu Chen, Changhao Chen, Wenlong Zhong, Ye Xie, Mengyi Zhu, Chengxiao Liao, Haoxuan Wang, Hongkun Yang, Fan Jiang, Runnan Shen, Shuting Huang, Rui Chen, Tianxin Lin*

Summary

Background Accurate detection of lymph node metastasis is crucial for precise tumour staging and treatment planning. Conventional pathological examination can overlook lymph node micrometastasis, resulting in underdiagnosis and suboptimal clinical outcomes. This study aimed to develop a pan-cancer artificial intelligence diagnostic model (PanCAM) for detecting lymph node metastasis across cancer types.

Methods In this multicentre diagnostic study, patients who had undergone tumour resection and lymph node dissection from 17 hospitals in China were included. The entire dataset included nine common and 24 rare cancers. Histological slides of resected lymph nodes were collected and scanned to generate whole slide images (WSIs). PanCAM was developed by use of supervised learning and incremental learning strategies and trained and internally validated retrospectively on WSIs from nine common cancers at Sun Yat-sen Memorial Hospital of Sun Yat-sen University (Guangzhou, China). Its generalisability was retrospectively validated by use of WSIs from 15 external hospitals in China and the publicly available CAMELYON16 dataset from the Netherlands, representing both common and rare cancers. Prospective validation was done in a multicentre study (NCT06517979) across nine hospitals in China. The primary outcome was diagnostic sensitivity for detecting lymph node metastasis. In a secondary analysis, we compared performance between PanCAM and pathologists.

Findings Between Jan 1, 2013, and Nov 30, 2024, 9256 patients from 17 hospitals in China were included in the study (4735 [51·2%] men; 4521 [48·8%] women; median age 60 years [IQR 50–69]; 3486 [37·7%] with lymph node metastasis). The dataset comprised 1303 patients in the training set, 558 in the internal validation set, 6006 in the external validation sets, and 1389 in the prospective validation sets, totalling 69 502 images and 153 985 lymph nodes. The CAMELYON16 dataset consisted of 399 images. In the retrospective validation, the diagnostic sensitivity of PanCAM for detecting lymph node metastasis ranged from 0·97 (95% CI 0·92–0·99) to 1·00 (0·98–1·00) across 16 hospitals, and was 0·96 (0·92–0·99) on the CAMELYON16 dataset. In the prospective validation, PanCAM exhibited sensitivity ranging from 0·93 (0·78–0·99) to 1·00 (0·98–1·00) across nine hospitals. Despite being trained solely on images from common cancers, PanCAM achieved a sensitivity of 0·98 (0·95–1·00) for rare cancers in both retrospective and prospective validations. At the patient level, PanCAM identified 120 additional patients with lymph node metastasis who were missed by pathologists in the retrospective validation and 21 additional cases in the prospective validation.

Interpretation PanCAM provided a generalisable solution for detecting lymph node metastasis across cancer types. With high sensitivity and robust performance, the model could assist pathologists in diagnosing lymph node metastasis, improving diagnostic accuracy, supporting treatment decision making, and ultimately enhancing patient outcomes.

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

Introduction

Cancer is a leading contributor to the global disease burden and poses a serious threat to human health. Both incidence and mortality rates of cancer continue to rise annually, with 20 million new cases and 9·7 million cancer-related deaths in 2022.[1] Among patients with cancer, lymph node metastasis is a key prognostic factor associated with lower overall and recurrence-free survival.[2–4] Several studies have shown

Lancet Digit Health 2026; 8: 100961

Published Online March 5, 2026 https://doi.org/10.1016/j.landig.2025.100961

For the Chinese translation of the abstract see Online for appendix 1

*Co-first authors, contributed equally

Department of Urology, Guangdong Provincial Clinical Research Centre for Urological Diseases, 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 (S Wu MD, G Hong MD, Yun Wang MD, X Chen MD, C Chen MD, W Zhong MD, Y Xie MD, M Zhu MD, C Liao MD, H Wang MD, H Yang MD, F Jiang MD, R Shen MD, Prof T Lin PhD); **Department of Pathology, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou, China** (H Zeng MD); **CellsVision Medical Technology Services, Guangzhou, China** (Z Lin ME, S Huang BD, R Chen BD); **Department of Pathology, Zhujiang Hospital of Southern Medical University, Guangzhou, China** (J Yang MD); **Department of Pathology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China** (J Chen MD); **Department of Clinical Laboratory** (H Chen MD) **and Department of Urology** (Prof T Lin), **The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China; Department of Pathology, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, China** (Y Huang MD); **Department of**

Pathology, The First Affiliated Hospital of Shantou University Medical College, Shantou, China (L Hong MD); **Department of Pathology, Shantou Central Hospital, Shantou, China** (Yua Wang MD); **Department of Pulmonary Oncology, The First People's Hospital of Foshan, Foshan, China** (Prof W Feng MD); **Department of Pathology, Shen-Shan Central Hospital, Shanwei, China** (Y Zeng MD); **Department of Urology, Houjie Hospital of Dongguan, Dongguan, China** (H Huang MD); **Department of Urology, Yan'an Hospital of Kunming, Kunming, China** (C Wang MD); **Department of Urology, The Fifth Affiliated Hospital of Xinjiang Medical University, Xinjiang, China** (C Xing MD); **Department of Pathology, Beijing Tiantan Hospital of Capital Medical University, Beijing, China** (G Dong MD); **Department of Pathology, Guangzhou Red Cross Hospital, Guangzhou, China** (Prof X Hua MD); **Department of Pathology, Aerospace Centre Hospital, Beijing, China** (C Qi MD); **Department of Pathology, The Seventh Medical Centre of the Chinese People's Liberation Army General Hospital, Beijing, China** (Prof A Liu MD); **Department of Pathology, The Ninth Medical Centre of the Chinese People's Liberation Army General Hospital, Beijing, China** (J Wu MD)

Correspondence to: Prof Tianxin Lin, Department of Urology, Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou 510120, China
lintx@mail.sysu.edu.cn

Research in context

Evidence before this study

We searched PubMed on Jan 31, 2025, without language or date restrictions, using the terms ("pathology" OR "whole slide image") AND ("lymph node metastasis") AND ("artificial intelligence" OR "deep learning"). Among the 436 search results, 21 relevant original articles that applied artificial intelligence to detect lymph node metastasis in whole slide images of cancers were identified. One study proposed a pan-origin lymph node metastasis detection system, incorporating data from three hospitals and covering 52 organs. However, this study included only 3800 images and reported a suboptimal specificity of 72·2%. The remaining 20 studies focused on single or a small number of cancer types—most commonly breast, gastric, colorectal, bladder, or prostate cancer—with finite generalisability across different cancer types. Moreover, most studies showed suboptimal sensitivity and specificity, lacked validation in large-scale multicentre cohorts, and did not include prospective evaluation.

Added value of this study

In this multicentre study, we developed a pan-cancer artificial intelligence diagnostic model (PanCAM) for detecting lymph node metastasis on whole slide images, which showed reliable performance, particularly in identifying lymph node micrometastasis. To our knowledge, this study used the largest multicentre dataset to date in this field, consisting of 9256 patients and 69 502 images representing 33 cancer types across 17 hospitals, and did the first multicentre prospective validation for lymph node metastasis detection by use of artificial intelligence. The model, trained on histological images from nine common cancers, generalised well to rare cancers. PanCAM showed robust diagnostic performance in both retrospective and prospective validation and exhibited favourable generalisability across hospitals, cancer types, and scanner types. Notably, PanCAM identified 141 patients with lymph node metastasis who were missed by pathologists, potentially reducing the missed diagnosis rate by ten (1%) of 697 to eight (23%) of 35 across hospitals.

Implications of all the available evidence

Artificial intelligence-based models can complement conventional pathological diagnosis by improving the sensitivity of lymph node metastasis detection, thereby reducing missed diagnoses of nodal micrometastasis and enabling more precise treatment. PanCAM showed robust performance across cancer types and hospitals, indicating its potential for broad clinical adoption, particularly in settings with poor medical resources.

that postoperative adjuvant therapy and intensified surveillance in patients with lymph node metastasis can significantly improve survival outcomes.[5–7] These findings underscore the increasing demand for accurate detection of lymph node metastasis.

Histological examination of lymph nodes remains the gold standard for metastasis detection. However, the heavy workload in clinical pathology, lack of time for detailed review, and inherent challenges in identifying micrometastasis (tumour <2 mm in diameter) can result in missed lesions during routine pathological practice.[8] Such oversight of micrometastasis might lead to underestimation of nodal staging and compromise treatment decisions. Therefore, more precise and efficient diagnostic tools are urgently needed to enhance diagnostic accuracy.

Artificial intelligence (AI) has shown potential in clinical pathology.[9–11] AI-based models can extract crucial features from digitised pathology images and have already been employed in early cancer detection, staging, treatment response evaluation, and prognostic prediction.[12–16] In the context of lymph node metastasis, several AI-based diagnostic models have been developed that match or even surpass expert-level performance, improve diagnostic efficiency, and reduce reliance on immunohistochemistry.[8,17–19] Retrospective analyses by use of AI models have increased the detection of positive cases by 8–22%.[17,20,21]

Nevertheless, most existing AI models for lymph node metastasis detection are restricted to one or a few cancer types, limiting their clinical utility. For rare cancers, the scarcity of available data further hampers the development of accurate and robust AI models.[22] Leveraging large-scale pathological image datasets encompassing diverse cancer types might enable AI models to learn more generalisable features, thereby enhancing diagnostic performance across a broad spectrum of malignancies.

We hypothesised that an AI model could automatically detect lymph node metastasis across cancer types by integrating well established algorithms with advanced computing power and big data, thereby assisting pathologists. In this study, a pan-cancer artificial intelligence diagnostic model (PanCAM) was developed to detect lymph node metastasis, by use of the largest multicentre dataset reported to date, comprising 9256 patients and 69 502 whole slide images (WSIs) from 33 cancer types across 17 hospitals.

Methods

Study design and participants

This multicentre diagnostic study involved a retrospective study across 16 hospitals in China to develop PanCAM and evaluate its generalisability, and a prospective validation at nine hospitals in China to assess its clinical applicability. The publicly available CAMELYON16 dataset from the Netherlands was incorporated as an international validation cohort. For model development, patients who had undergone tumour resection and lymph node dissection across China were included, without restrictions on age, sex, histological type, or receipt of neoadjuvant therapy (race and ethnicity data were not collected). The study covered nine common cancers (lung, breast, colorectal,

prostate, gastric, thyroid, cervical, bladder, and oesophageal cancer) and 24 rare cancers (appendix 2 pp 8, 10). Participating hospitals were selected from both tertiary and grass-roots hospitals to represent health-care disparities. The workflow of this study is illustrated in figure 1.

This study was approved by the medical ethics committee of Sun Yat-sen Memorial Hospital of Sun Yat-sen University (SYSMH, approval number: SYSKY–2024–513–01), and the research protocol has also been reviewed and filed by the ethics committees of all sub-centres. Patient information was de-identified for privacy protection. Each patient had provided written informed consent for the use of clinical residual samples on admission, and the institutional ethics committee waived additional informed consent requirements for both retrospective and prospective validation given the non-interventional nature of the study and use of de-identified data. This study followed STARD guidelines and was conducted in accordance with the Helsinki Declaration, and the prospective validation was registered with ClinicalTrials.gov, NCT06517979.

Data and image acquisition

Baseline characteristics, including age, sex, cancer type, histological type, and neoadjuvant therapy status, were extracted from the electronic medical records systems of the participating hospitals. The data were queried and entered into a standardised electronic data collection form by the participating investigators at each hospital. To ensure data quality, a random sample of the entered data was cross-verified by authors TL, SW, GH, and YunW against the original records. Pathological slides stained with haematoxylin and eosin from lymphadenectomy specimens were retrieved from pathology archives and scanned at 20× magnification to generate WSIs. The slide scanners included SQS-600P (Shenzhen, China), KF-PRO-400 (Ningbo, China), and Pannoramic SCAN II (Budapest, Hungary). Severely faded images were excluded. CAMELYON16 images were obtained by use of Pannoramic 250 Flash II (Budapest, Hungary) and NanoZoomer-XR C12000 (Hamamatsu, Japan).[8] Details of scanner types used at each hospital are shown in appendix 2 (p 9).

Model development

Two senior pathologists (JY and JC, each with >15 years of experience) independently reviewed all WSIs across all datasets and provided diagnoses, with immunohistochemistry employed for definitive diagnosis in cases of disagreement (appendix 2 p 6). The consensus diagnosis from the two pathologists, combined with immunohistochemistry results, was considered the reference standard. WSIs were categorised as positive (including micrometastasis [tumour cell cluster diameter >0 mm to ≤2 mm] and macrometastasis [diameter >2 mm]) or negative. In the retrospective study, PanCAM was trained and internally validated on WSIs from patients with nine common cancers at SYSMH (Guangzhou, China). Patients were stratified by both lymph node metastasis status (positive *vs* negative) and cancer type, and were then randomly assigned to training and internal validation sets in a 7:3 ratio using a computer-generated random number sequence. For positive images from nine common cancers in the training set, representative tumour regions were selected by two senior pathologists (JY and JC) for pixel-level annotation. To expedite annotation a model-assisted strategy was employed, in which an initial model generated potential tumour regions for manual review. False positives and false negatives were manually corrected to ensure accuracy. These verified annotations were integrated into the training set for iterative model improvement (appendix 2 pp 4, 24). For challenging cases, annotation labels were finalised through discussion between the same two pathologists, with immunohistochemistry used when consensus was not achieved (appendix 2 p 6).

See Online for appendix 2

PanCAM was developed by use of the DeepLabv3+ segmentation framework with RegNet-Y40 as the encoder. The model was trained through incremental learning, starting with annotated WSIs from patients with prostate cancer and progressively incorporating eight other common cancers (appendix 2 pp 4, 24). A comprehensive data sampling strategy combining key patch selection, hard sample integration, and fixed-ratio random sampling was employed to optimise model training with heterogeneous data. To enhance robustness, spatial and colour-based image augmentation techniques were applied during training. Image classification was done by use of a pixel threshold method, which involved two variables: the pixel threshold, in which pixels above the threshold were classified as positive, and the number of positive pixels threshold, in which images exceeding this count were classified as positive. Optimal pixel threshold and number of positive pixels threshold values were established by maximising the F2 score, which weights recall twice as much as precision, in the internal validation set. The F2 score is calculated using the following formula: $F2\ score=(5\times precision\times recall)/(4\times precision+recall)$.

The visualisation of PanCAM development is presented in figure 2, with detailed methods provided in appendix 2 (pp 4–6).

In addition to the WSI-level evaluation, we also assessed the PanCAM's diagnostic performance at the patient level. A patient was considered positive for lymph node metastasis if any of their WSIs was predicted as positive by PanCAM. PanCAM was externally validated retrospectively on WSIs from patients with 29 common and rare cancers at 15 hospitals and on WSIs from patients with breast cancer in the CAMELYON16 dataset (appendix 2 p 8). For the CAMELYON16 dataset, the performance of PanCAM was compared with the best-reported results (state-of-the-art) from previously published studies on the same dataset (appendix 2 p 17).

Prospective validation of the model

PanCAM parameters were frozen following model development. To further evaluate the clinical applicability of PanCAM in a rigorous manner that mimics real-world

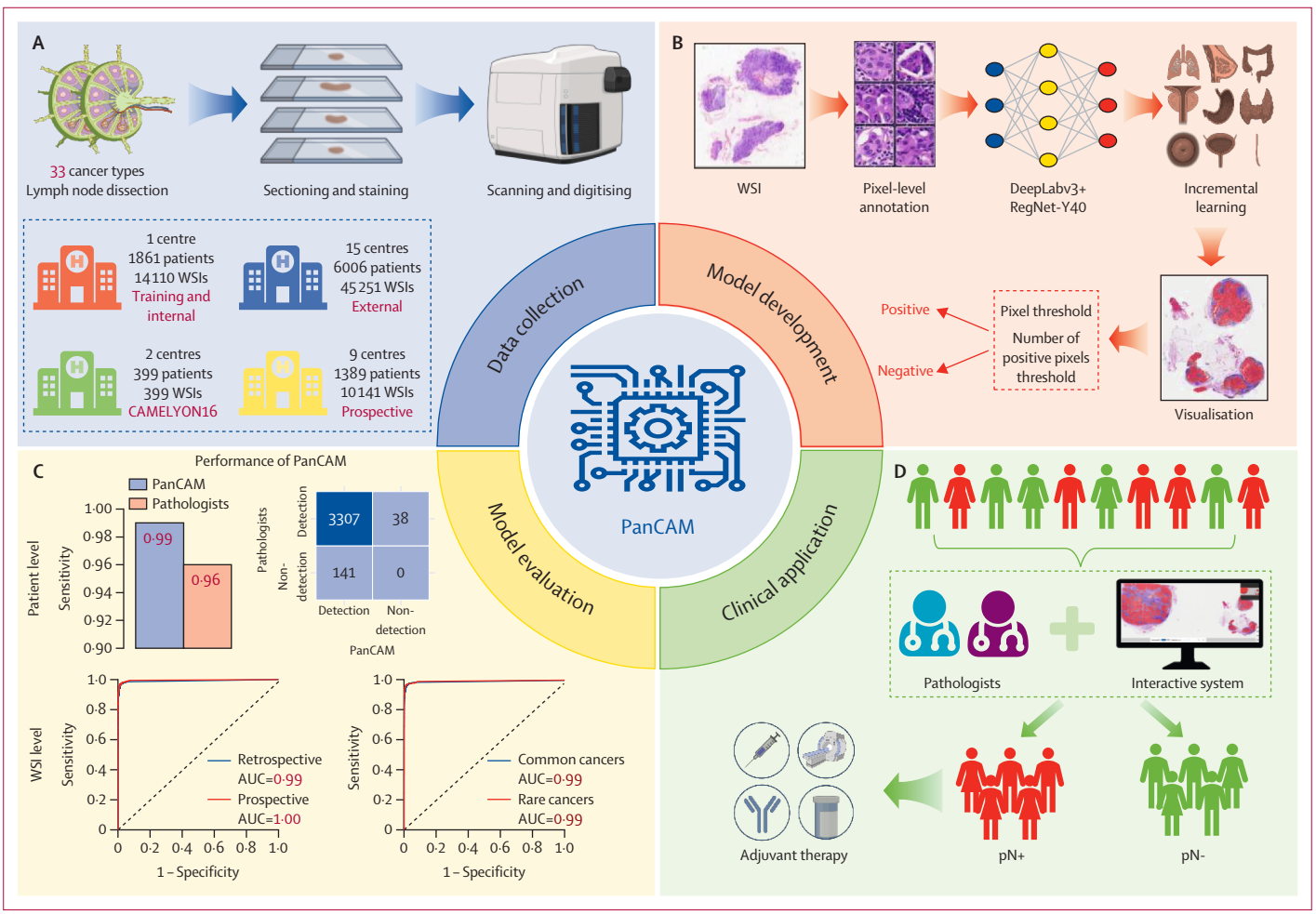


Figure 1: **Workflow of the study**

The development of PanCAM comprised four stages. (A) The data collection phase involved obtaining lymph node sections from 33 cancer types across multiple medical centres, followed by scanning and digitising to generate WSIs. WSIs were also obtained from the publicly available CAMELYON16 dataset from the Netherlands. (B) PanCAM was developed by use of pixel-level annotation of WSIs, on the basis of a model architecture integrating DeepLabv3+ and RegNet-Y40 for segmentation and classification. An incremental learning strategy was employed to optimise model performance. Image classification was done by use of a pixel threshold method, relying on two key variables: pixel threshold and number of positive pixels threshold. (C) Model performance was evaluated at both the patient and WSI levels. (D) PanCAM was integrated into clinical settings in a prospective multicentre study to assist pathological diagnosis and support adjuvant therapy decision making. Some items were created by use of BioRender.com. PanCAM=pan-cancer artificial intelligence diagnostic model. AUC=area under the curve. pN−=pathological lymph node metastasis negative. pN+=pathological lymph node metastasis positive. WSI=whole slide image.

deployment, we designed and conducted a prospective validation across nine hospitals (appendix 2 p 10). Lymph node slides from enrolled eligible patients at nine participating hospitals were prospectively collected and digitised into WSIs. The data collection was planned before the index test (PanCAM diagnosis) and the reference standard were done. Clinical pathologists examined lymph node slides under standard pathological procedures using microscopy, while PanCAM concurrently analysed the WSIs to generate its diagnostic output. Both clinical pathologists and PanCAM were masked to each other's diagnostic results. The WSIs were subsequently reviewed by two senior pathologists (JY and JC) to establish the reference standard. In cases of discordance between clinical pathologists and PanCAM, senior pathologists provided arbitration for research analysis, with immunohistochemistry employed as needed to confirm the diagnosis (appendix 2 p 6).

Benchmarking against cancer-specific models

To comprehensively evaluate the diagnostic performance of PanCAM, a comparative analysis was done against other cancer-specific models. Specifically, two previously validated models were selected: bladder cancer-specific model (BCa model),[17] trained on bladder cancer, and prostate cancer-specific model (PCa model),[21] trained on prostate cancer. PanCAM and the cancer-specific models were trained by use of an identical dataset of bladder and prostate cancer samples. A dataset from three hospitals in China was used to evaluate the models' diagnostic performance (appendix 2 p 11).

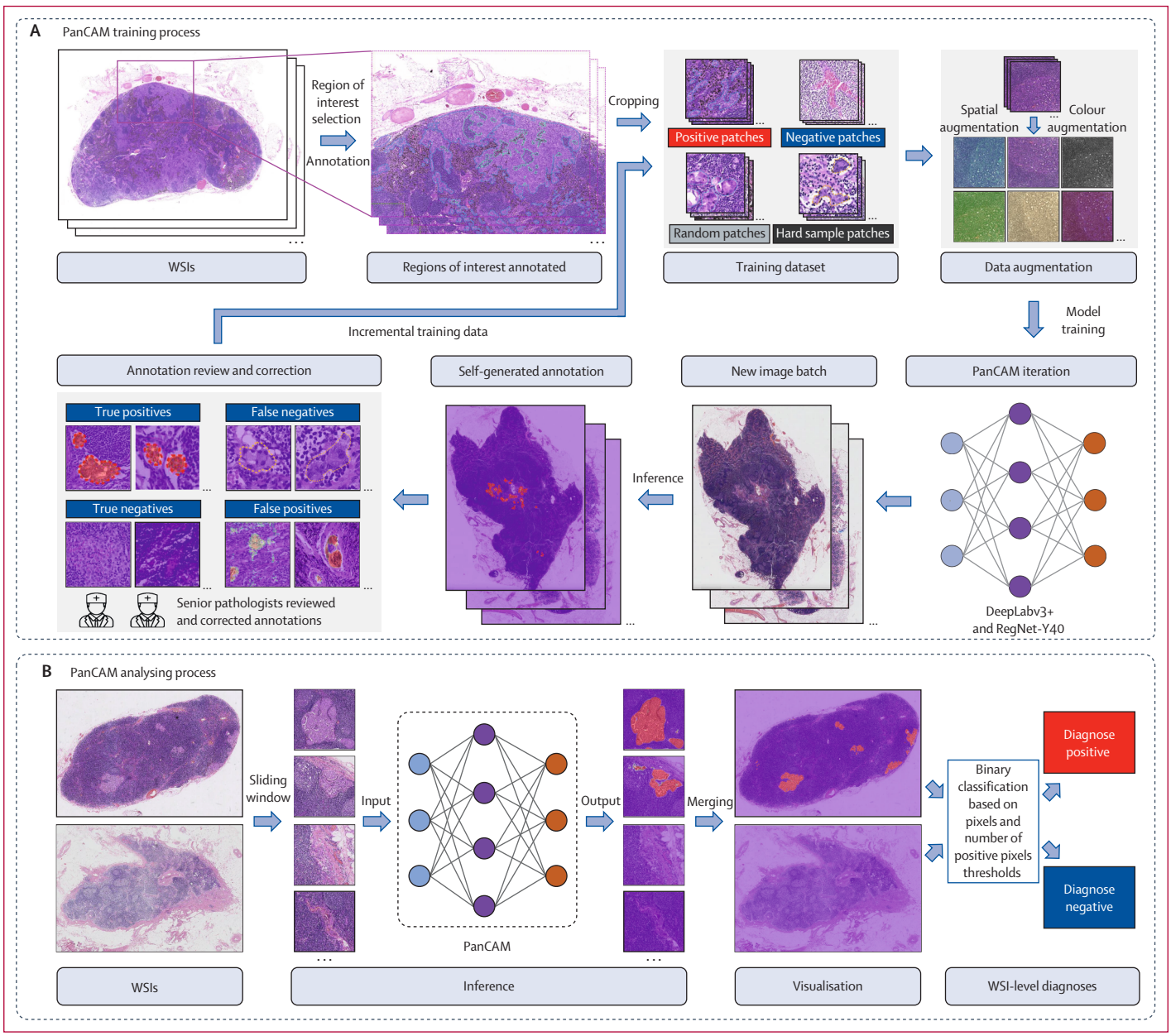


***Figure 2*: PanCAM development**

(A) PanCAM training process. Pathologists manually selected regions of interest on WSIs for pixel-level annotation. These regions of interest were then cropped into square patches. A comprehensive training dataset was constructed by incorporating positive, negative, hard sample, and random patches. Following data augmentation, the patches were used to train PanCAM, which was built on the DeepLabv3+ and RegNet-Y40 architecture. Subsequently, PanCAM analysed a new batch of WSIs and generated preliminary annotations to expedite the annotation process, followed by pathologists' review and correction. These verified annotations were integrated into the training dataset for model iteration. More details are shown in appendix 2 (pp 4–6, 24). (B) PanCAM analysing process. WSIs were processed by use of a sliding window approach to generate patches, which were then input into PanCAM. The model produced segmentation heatmaps for visualisation and pixel-level prediction probabilities for classification. WSI-level diagnoses were determined on the basis of pixel threshold and number of positive pixels threshold, with details provided in appendix 2 (p 6). PanCAM=pan-cancer artificial intelligence diagnostic model. WSI=whole slide image.

Outcomes

The primary outcome was the diagnostic sensitivity of PanCAM for detecting lymph node metastasis. Secondary outcomes included specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV; calculated from the 2×2 contingency table), as well as area under the receiver operating characteristic curve (AUC); a comparison of sensitivity between PanCAM and pathologists; the performance of PanCAM in subgroups of diverse histological types, neoadjuvant therapy, and tumour micrometastasis and macrometastasis; and performance differences between PanCAM and cancer-specific models.

	Training set (n=1303)	Internal validation set (n=558)	External validation sets (n=6006)	Prospective validation sets (n=1389)
Number of hospitals	1	1	15	9
Number of whole slide images	9953	4157	45 251	10 141
Number of lymph nodes	22 522	9298	98 407	23 758
Median age, years	60 (50–68)	63 (55–69)	60 (50–69)	59 (49–68)
Sex				
Male	688 (52·8%)	376 (67·4%)	3031 (50·5%)	640 (46·1%)
Female	615 (47·2%)	182 (32·6%)	2975 (49·5%)	749 (53·9%)
Lymph node status				
With metastasis	511 (39·2%)	186 (33·3%)	2277 (37·9%)	512 (36·9%)
Without metastasis	792 (60·8%)	372 (66·7%)	3729 (62·1%)	877 (63·1%)
Histological type				
Adenocarcinoma	1091 (83·7%)	331 (59·3%)	4711 (78·4%)	1208 (87·0%)
Squamous cell carcinoma	181 (13·9%)	49 (8·8%)	612 (10·2%)	83 (6·0%)
Adenosquamous carcinoma	9 (0·7%)	3 (0·5%)	12 (0·2%)	4 (0·3%)
Others*	22 (1·7%)	175 (31·4%)	671 (11·2%)	94 (6·8%)
Receipt of neoadjuvant therapy				
Yes	284 (21·8%)	140 (25·1%)	883 (14·7%)	242 (17·4%)
No	1019 (78·2%)	418 (74·9%)	5123 (85·3%)	1147 (82·6%)

Data are n, median (IQR), or n (%). *Other histological types include urothelial carcinoma, melanoma, lymphoepithelioma, serous carcinoma, small-cell neuroendocrine carcinoma, large-cell neuroendocrine carcinoma, adenoid cystic carcinoma, undifferentiated carcinoma, sarcomatoid carcinoma, carcinosarcoma, clear-cell carcinoma, medullary carcinoma, endometrioid carcinoma, basal-cell carcinoma, rhabdomyosarcoma, pilomatrix carcinoma, chromophobe carcinoma, hepatocellular carcinoma, embryonal carcinoma, carcinoid, fibrosarcoma, and metaplastic carcinoma.

Table 1: **Baseline characteristics of included patients**

Statistical analysis

The sample size for the retrospective validation was determined on the basis of sample availability, whereas that for the prospective validation was calculated by power analysis with PASS software (version 15.0.5, appendix 2 p 7). Statistical analyses were done by use of R software (version 4.4.2). AUCs were calculated by use of the pROC package (version 1.18.5). AUC comparison was done by use of the DeLong test. 95% CIs for sensitivity, specificity, accuracy, PPV, and NPV were calculated by use of the Clopper–Pearson method. Diagnostic performance comparisons between PanCAM and pathologists, as well as across different models, were done by use of the χ^2 test. All analyses were prespecified. A two-sided p value of less than 0·05 was considered significant.

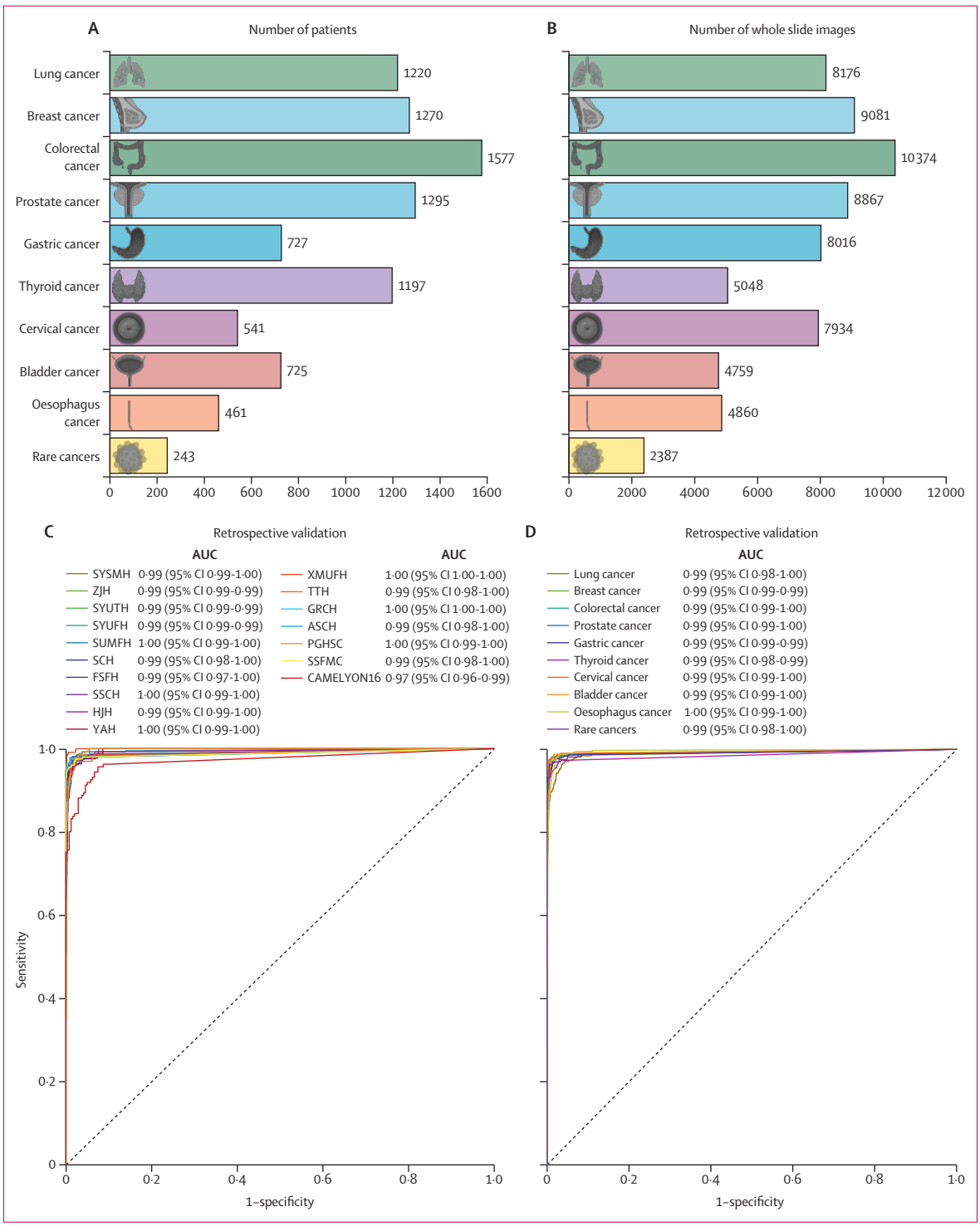
Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

In the retrospective study, 7867 patients (4095 [52·1%] men, 3772 [47·9%] women; median age 60 years [IQR 50–69]; 2974 [37·8%] with lymph node metastasis) were included from 16 hospitals between Jan 1, 2013, and June 30, 2024. Severe fading led to the exclusion of 62 images. Ultimately, 59 361 WSIs and 130 227 lymph nodes were included in the analysis. Among the 7867 patients, 1303 (9953 WSIs and 22 522 lymph nodes) from SYSMH were assigned to the training set, 558 (4157 WSIs and 9298 lymph nodes) from SYSMH to the internal validation set, and 6006 (45 251 WSIs and 98 407 lymph nodes) from the remaining 15 hospitals to external validation sets. The CAMELYON16 dataset (399 WSIs) was used as an international validation set. In the prospective validation phase, 1389 patients (640 [46·1%] men, 749 [53·9%] women; median age 59 years [IQR 49–68]; 512 [36·9%] with lymph node metastasis) were included from nine hospitals between Sept 1, 2024, and Nov 30, 2024, yielding 10 141 WSIs and 23 758 lymph nodes (table 1; figure 3A, B; appendix 2 pp 12, 25).

In the retrospective validation across 16 hospitals, PanCAM showed sensitivity ranging from 0·97 (95% CI 0·92–0·99) to 1·00 (0·98–1·00), AUC from 0·99 (0·97–1·00) to 1·00 (1·00–1·00), and specificity from 0·88 (0·84–0·91) to 0·98 (0·96–0·99). Accuracy ranged between 0·89 (95% CI 0·85–0·92) and 0·98 (0·97–0·99) and PPV between 0·49 (0·39–0·60) and 0·93 (0·89–0·96), and NPV exceeded 0·99 (0·97–1·00) at all hospitals. For the CAMELYON16 dataset, PanCAM had a sensitivity of 0·96 (95% CI 0·92–0·99), an AUC of 0·97 (0·96–0·99), and a specificity of 0·91 (0·87–0·94), with comparative analysis against state-of-the-art performance provided in appendix 2 (p 17). Additionally, PanCAM showed sensitivity ranging from 0·96 (0·95–0·97) to 1·00 (0·98–1·00) and specificity from 0·90 (0·89–0·91) to 0·98 (0·97–0·98) across nine common cancers; for rare cancers, sensitivity was 0·98 (0·95–1·00) and specificity was 0·92 (0·90–0·94; figure 3C, D; table 2; appendix 2 pp 13, 15, 26–27).



(Figure 3 continues on next page)

In the prospective validation across nine hospitals, PanCAM showed sensitivity ranging from 0·93 (95% CI 0·78–0·99) to 1·00 (0·98–1·00), AUC from 0·96 (0·91–1·00) to 1·00 (1·00–1·00), and specificity from 0·89 (0·87–0·91) to 0·98 (0·97–0·98). For common cancers, PanCAM had a sensitivity ranging from 0·96 (0·78–1·00) to 1·00 (0·89–1·00) and a specificity ranging from 0·85 (0·75–0·92) to 0·97 (0·94–0·98); for rare cancers, sensitivity was 0·98

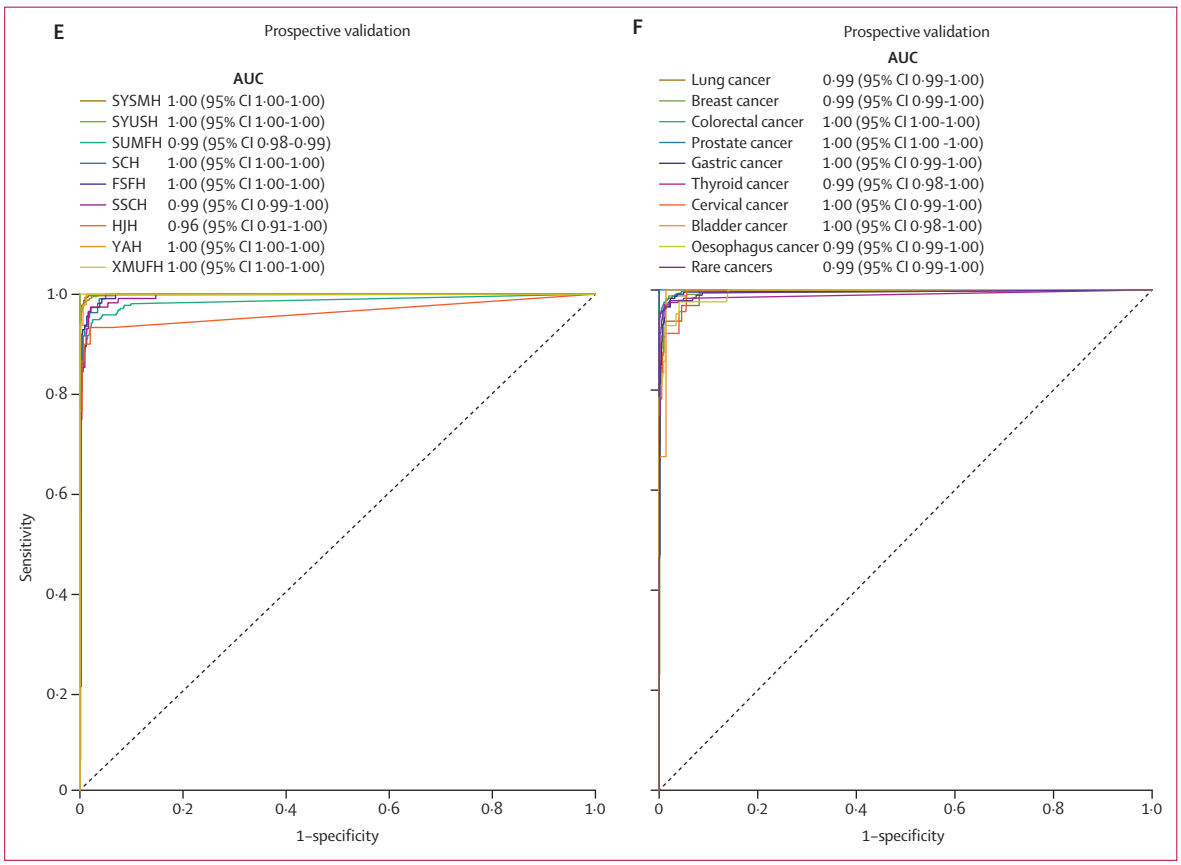


Figure 3: **Performance of the PanCAM across hospitals and cancer types**
(A, B) Number of patients and whole slide images included across nine common and 24 rare cancer types from 17 hospitals. (C, D) ROC curves showing the performance of PanCAM in detecting lymph node metastasis across hospitals and cancer types in the retrospective validation. (E, F) ROC curves showing the performance of PanCAM in detecting lymph node metastasis across hospitals and cancer types in the prospective validation. Details on rare cancer types and the full names of participating hospitals in China are provided in appendix 2 (pp 8, 10). CAMELYON16 is a publicly available dataset in the Netherlands, used for external validation. Some items were created by use of BioRender.com. AUC=area under the receiver operating characteristic curve. PanCAM=pan-cancer artificial intelligence diagnostic model. ROC=receiver operating characteristic.

(0·95–1·00) and specificity was 0·93 (0·91–0·94; figure 3E, F; table 2; appendix 2 pp 14, 16, 28–29). In the retrospective and prospective validation set of 7953 patients, PanCAM generated 2364 false-positive classifications (0·3 per patient), each requiring a median additional review time of 13 s (IQR 10–16).

For patient-level diagnosis, PanCAM exhibited sensitivity ranging across hospitals from 0·96 (95% CI 0·86–1·00) to 1·00 (0·94–1·00) in the retrospective validation and 0·99 (0·94–1·00) to 1·00 (0·97–1·00) in the prospective validation. Pathologists had sensitivities ranging across hospitals from 0·77 (0·60–0·90) to 1·00 (0·93–1·00) in the retrospective validation and 0·87 (0·60–0·98) to 1·00 (0·78–1·00) in the prospective validation (appendix 2 pp 18–19). Across the entire validation set, PanCAM exhibited higher overall sensitivity than pathologists (0·99 [0·99–0·99] *vs* 0·96 [0·95–0·97], p<0·0001), whereas pathologists showed superior specificity (1·00 [1·00–1·00] *vs* 0·82 [0·81–0·83], p<0·0001; figure 4A–C; appendix 2 pp 20–21). PanCAM identified 120 patients who were positive but missed by pathologists, of whom 114 were micrometastasis cases, and reduced the missed diagnosis rate by 1–23% (from ten of 697 to eight of 35) across hospitals in the retrospective validation. In the non-interventional prospective validation, PanCAM identified 21 additional patients who were positive but overlooked by pathologists, all of whom were micrometastasis cases, potentially reducing the missed diagnosis rate by 2–13% (from one of 44 to two of 15) across hospitals (appendix 2 pp 18–19, 30).

PanCAM showed robust performance across diverse histological types, with AUCs of 0·99 (95% CI 0·99–0·99) for adenocarcinoma, 1·00 (1·00–1·00) for squamous cell carcinoma, 1·00 (1·00–1·00) for adenosquamous carcinoma, and 0·99 (0·98–0·99) for other histological types (appendix 2 p 31). Among patients who had received neoadjuvant therapy, PanCAM had an AUC of 0·99 (0·99–0·99; appendix 2 p 32). In subgroups of positive WSIs, PanCAM had a sensitivity of 0·95 (95% CI 0·95–0·96) for tumour micrometastasis and 1·00 (1·00–1·00) for tumour macrometastasis (appendix 2 p 22).

	Sensitivity (95% CI)	AUC (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)
Retrospective validation set						
Lung cancer	0·98 (0·97–0·99)	0·99 (0·98–1·00)	0·90 (0·89–0·91)	0·91 (0·90–0·91)	0·40 (0·37–0·43)	1·00 (1·00–1·00)
Breast cancer	0·98 (0·98–0·99)	0·99 (0·99–0·99)	0·96 (0·96–0·97)	0·97 (0·96–0·97)	0·90 (0·89–0·92)	0·99 (0·99–1·00)
Colorectal cancer	0·99 (0·98–0·99)	0·99 (0·99–1·00)	0·96 (0·95–0·97)	0·97 (0·96–0·97)	0·85 (0·83–0·87)	1·00 (1·00–1·00)
Prostate cancer	0·98 (0·97–0·99)	0·99 (0·99–1·00)	0·97 (0·97–0·97)	0·97 (0·97–0·98)	0·81 (0·78–0·84)	1·00 (1·00–1·00)
Gastric cancer	0·98 (0·97–0·99)	0·99 (0·99–0·99)	0·96 (0·96–0·97)	0·97 (0·96–0·97)	0·90 (0·88–0·91)	0·99 (0·99–1·00)
Thyroid cancer	0·96 (0·95–0·97)	0·99 (0·98–0·99)	0·98 (0·97–0·98)	0·97 (0·96–0·98)	0·96 (0·95–0·97)	0·97 (0·97–0·98)
Cervical cancer	0·99 (0·97–1·00)	0·99 (0·99–1·00)	0·98 (0·97–0·98)	0·98 (0·97–0·98)	0·78 (0·74–0·81)	1·00 (1·00–1·00)
Bladder cancer	0·99 (0·97–1·00)	0·99 (0·99–1·00)	0·96 (0·95–0·97)	0·96 (0·96–0·97)	0·66 (0·62–0·70)	1·00 (1·00–1·00)
Oesophagus cancer	1·00 (0·98–1·00)	1·00 (0·99–1·00)	0·92 (0·91–0·93)	0·93 (0·92–0·94)	0·65 (0·61–0·69)	1·00 (1·00–1·00)
Rare cancers	0·98 (0·95–1·00)	0·99 (0·98–1·00)	0·92 (0·90–0·94)	0·93 (0·92–0·95)	0·70 (0·64–0·76)	1·00 (0·99–1·00)
Prospective validation set						
Lung cancer	1·00 (0·89–1·00)	0·99 (0·99–1·00)	0·90 (0·88–0·92)	0·90 (0·88–0·92)	0·28 (0·20–0·37)	1·00 (1·00–1·00)
Breast cancer	0·99 (0·97–1·00)	0·99 (0·99–1·00)	0·95 (0·93–0·96)	0·96 (0·95–0·97)	0·88 (0·84–0·91)	1·00 (0·99–1·00)
Colorectal cancer	0·99 (0·98–1·00)	1·00 (1·00–1·00)	0·97 (0·96–0·97)	0·97 (0·97–0·98)	0·85 (0·82–0·87)	1·00 (1·00–1·00)
Prostate cancer	1·00 (0·69–1·00)	1·00 (1·00–1·00)	0·95 (0·88–0·98)	0·95 (0·89–0·98)	0·67 (0·38–0·88)	1·00 (0·96–1·00)
Gastric cancer	0·99 (0·97–1·00)	1·00 (0·99–1·00)	0·95 (0·93–0·97)	0·96 (0·95–0·97)	0·88 (0·83–0·91)	1·00 (0·99–1·00)
Thyroid cancer	0·97 (0·95–0·99)	0·99 (0·98–1·00)	0·97 (0·94–0·98)	0·97 (0·95–0·98)	0·94 (0·91–0·97)	0·98 (0·97–0·99)
Cervical cancer	0·96 (0·78–1·00)	1·00 (0·99–1·00)	0·95 (0·93–0·97)	0·95 (0·93–0·97)	0·42 (0·29–0·57)	1·00 (0·99–1·00)
Bladder cancer	1·00 (0·29–1·00)	1·00 (0·98–1·00)	0·85 (0·75–0·92)	0·86 (0·76–0·93)	0·21 (0·05–0·51)	1·00 (0·94–1·00)
Oesophagus cancer	0·98 (0·87–1·00)	0·99 (0·99–1·00)	0·90 (0·86–0·93)	0·91 (0·88–0·94)	0·55 (0·43–0·67)	1·00 (0·98–1·00)
Rare cancers	0·98 (0·95–1·00)	0·99 (0·99–1·00)	0·93 (0·91–0·94)	0·94 (0·92–0·95)	0·69 (0·63–0·74)	1·00 (0·99–1·00)

AUC=area under the receiver operating characteristic curve. PanCAM=pan-cancer artificial intelligence diagnostic model.

Table 2: **Performance of PanCAM in detecting lymph node metastasis across cancer types in the retrospective and prospective validation sets**

In the comparison of PanCAM with cancer-specific models using the bladder cancer dataset, PanCAM showed a sensitivity of 0·99 (95% CI 0·97–1·00), an AUC of 0·99 (0·99–1·00), a specificity of 0·98 (0·97–0·98), accuracy of 0·98 (0·97–0·98), a PPV of 0·81 (0·76–0·85), and an NPV of 1·00 (1·00–1·00). By comparison, BCa model showed a sensitivity of 0·98 (0·96–0·99), an AUC of 0·99 (0·99–1·00), a specificity of 0·94 (0·93–0·95), accuracy of 0·94 (0·94–0·95), a PPV of 0·62 (0·57–0·66), and an NPV of 1·00 (1·00–1·00). Similar findings of PanCAM were observed in the prostate cancer dataset compared with PCa model (figure 4D, E, appendix 2 p 23).

The interactive interface of PanCAM is shown in appendix 2 (p 33). Visualisation showed strong concordance between model-identified and pathologist-annotated tumour regions (appendix 2 p 34). Analysis of false diagnoses (appendix 2 pp 35–36) revealed that most false negatives resulted from the number of positive pixels setting threshold misclassifying subthreshold metastases, although heatmaps still highlighted these tumour regions. Complete misses occurred when tumour metastases were obscured by lymph node tracers or were small-cell neuroendocrine carcinomas. False positives primarily involved proliferating histiocytes, high endothelial venules, and germinal centres.

Discussion

In this study we developed PanCAM, an AI-driven pathological diagnostic model capable of detecting lymph node metastasis in nine common and 24 rare cancers and validated it across 17 hospitals in China. PanCAM showed robust performance in both retrospective and prospective validation, particularly in detecting lymph node micrometastasis.

Pathological examination remains the gold standard for detecting lymph node metastasis, but it is a labour-intensive and time-consuming process. Pathologists might overlook tumour micrometastasis during routine slide review. To address this limitation, several studies have applied AI-based models to analyse WSIs from various cancer types, such as bladder, gastric, prostate, breast, and colorectal cancer, and reported improvements in diagnostic accuracy in detecting tumour metastasis.[17,19,21,23,24] However, most of these models were developed for specific cancer types and not validated in large-scale multicentre cohorts. One study[25] proposed a pan-origin lymph node metastasis detection system by use of data from 52 organs across three hospitals, but it included only 3800 images, achieved a specificity of 72·2%, and was not prospectively validated. In the present study, we assembled the largest multicentre dataset to date in this field, including 9256 patients and 69 502 WSIs from 33 cancer types across 17 hospitals, comprising 153 985 lymph nodes. PanCAM achieved robust diagnostic performance in both retrospective and prospective validation, with overall sensitivity exceeding 0·98 and specificity surpassing 0·95. When compared with two representative cancer-specific models, PanCAM showed similar or better diagnostic performance, underscoring its robustness across cancer types. The PPV was low

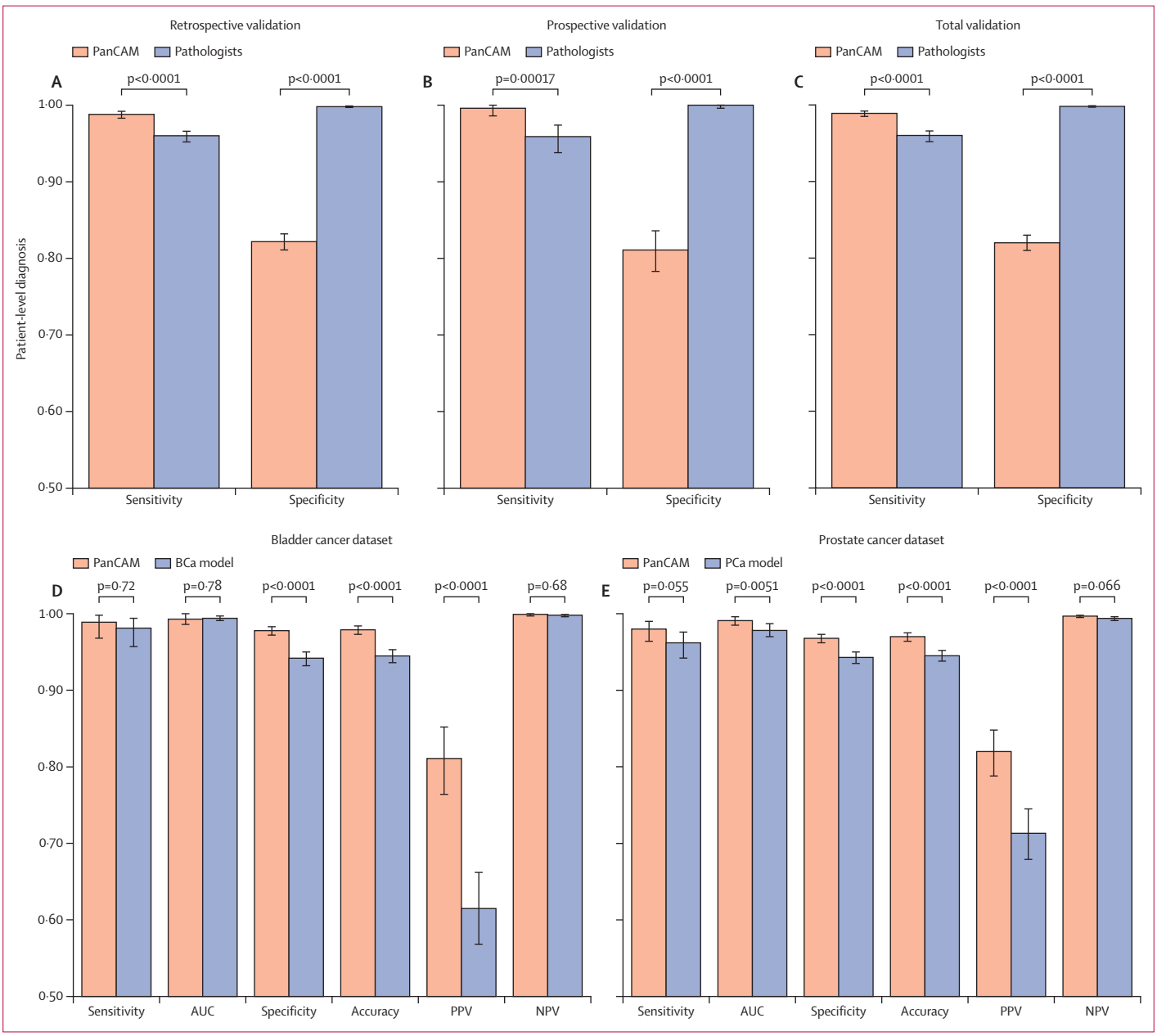


Figure 4: **Performance of PanCAM at the patient level and in cancer-specific datasets**
(A–C) Patient-level diagnostic sensitivity and specificity of PanCAM and pathologists for lymph node metastasis detection in the retrospective and prospective validation. (D, E) Performance comparison of PanCAM and cancer-specific models in the bladder and prostate cancer datasets. PanCAM=pan-cancer artificial intelligence diagnostic model. BCa model=bladder cancer-specific model. AUC=area under the receiver operating characteristic curve. PCa model=prostate cancer-specific model. NPV=negative predictive value. PPV=positive predictive value.

for certain cancer types (eg, lung and bladder cancer) in the current study, which is primarily attributable to the low proportion of positive images relative to negatives. Also, the sensitivity of PanCAM for small-cell neuroendocrine carcinoma remained suboptimal, probably due to its challenging morphology and the inadequate number of training samples (only two negative cases). Future model refinement should include more samples of this subtype to improve detection accuracy.

Foundation models[26–28] are increasingly showing potential for multitask analysis in pathology through self-supervised learning, which generates pseudo-labels from inherent structures of data to reduce annotation costs. However, for high-precision and fine-grained tasks such as tumour metastasis detection and segmentation, supervised learning approaches remain more effective. This study used pixel-level annotation to precisely localise lymph node metastasis, particularly micrometastasis, despite the higher annotation burden. To mitigate this burden, an annotation acceleration strategy was implemented to enhance pathologist efficiency.

Current AI models remain fundamentally data-dependent, requiring extensive high-quality training data

for optimal performance. In oncology, cancer prevalence follows a long-tail distribution, with rare malignancies posing particular challenges for data acquisition, even in multicentre collaborations. PanCAM, initially trained on nine common cancers, achieved an overall diagnostic sensitivity of 0·98 and specificity of 0·93 for 24 rare cancers in validation sets, matching its performance on common cancers. This shows favourable generalisability while highlighting the model's potential to bridge the rare disease gap in computational pathology.

To verify the integration of PanCAM with existing clinical workflows, this study did a multicentre prospective validation, aiming to assess the performance stability of PanCAM in real-world settings. This phase focused on evaluating the auxiliary value of PanCAM in improving clinical practice, with the participants being PanCAM and pathologists rather than patients. Clinical decisions were made by pathologists and physicians, rather than PanCAM. All the data used for analysis had been de-identified, ensuring no compromise to patient interests or privacy. Additionally, all patients had signed a general informed consent form at admission, authorising the use of their remaining clinical samples (such as blood, urine, stool, and surgically resected tissues, etc) for teaching or medical research. Given the non-interventional nature of the study, the use of de-identified data, and the fact that patients were not the direct participants, the institutional ethics committee waived the need for informed consent, in line with previously reported studies.[18,29–35]

The generalisability of AI models is paramount for clinical application. PanCAM, initially developed at SYSMH, showed consistent diagnostic performance across multiple external validation hospitals encompassing both tertiary and grass-roots hospitals. Its cross-institutional validity was further confirmed on the CAMELYON16 dataset. Most notably, PanCAM showed good performance in the multicentre, pan-cancer prospective validation, further supporting the reliability of PanCAM. Despite inherent variations in slide scanner types (resolution, colour, and contrast), PanCAM maintained robust and stable performance across multiple scanner brands. Interpretability remains a key consideration for clinical implementation. PanCAM could concurrently segment tumour metastases and generate prediction probabilities. Benefiting from the supervised learning strategy employed during training, the model exhibited good segmentation performance, with strong concordance between model-identified and pathologist-annotated tumour metastasis regions. The model performance, coupled with transparent decision visualisation, confirms its clinical applicability while avoiding black-box classification limitations.

For patient-level diagnosis, PanCAM identified 141 additional patients who were positive, potentially reducing the missed diagnosis rate by 1–23% across hospitals. Most of the missed cases were lymph node micrometastases, the accurate detection of which could facilitate more precise staging and treatment. Centre-specific analysis revealed that the missed diagnosis rate was notably higher in four municipal hospitals and one provincial hospital. This finding indicates that PanCAM could provide substantial benefits in municipal hospitals, which typically have fewer resources than provincial hospitals. In the provincial hospital, further analysis attributed the higher rate to high surgical volumes and consequent pathologist workload. Increasing sensitivity is typically associated with an increased risk of false positives, but the false positives in the current study were primarily composed of proliferating histiocytes, high endothelial venules, and germinal centres, all of which are typically identifiable during routine pathological review. As reported in the Results section, PanCAM generated 2364 false-positive classifications in the validation set, each requiring a median additional review time of 13 s (IQR 10–16). Although this modestly increases pathologists' workload, the trade-off would likely be considered acceptable given the clinically significant value of reducing missed diagnoses. In real-world applications, dynamic threshold adjustment across hospitals could help minimise misclassification.

Another factor contributing to missed diagnoses is the low number of lymph node slides examined in routine practice. Owing to the high workloads, pathologists typically prepare only one or two sections per resected lymph node. Given the random distribution of micrometastasis, this approach might miss tumour metastasis. Previous studies[36,37] have shown that serial sectioning improves the lymph node metastasis detection rate but at the cost of substantially increasing the manual review burden. PanCAM's high-throughput automated analysis enables it to continuously process large volumes of images, making serial sectioning more feasible in clinical practice.

Our study has some limitations. First, although PanCAM showed robust performance in Chinese cohorts and the CAMELYON16 dataset (the Netherlands), additional validation across diverse countries, populations, and scanner types is needed to confirm generalisability. Second, PanCAM was developed by use of a well established architecture to ensure clinical practicality. Future work should explore newer, lightweight architectures for potential performance improvement. Third, PanCAM was developed by use of paraffin-embedded sections and needs further evaluation for intraoperative frozen section applications. Fourth, detailed ethnicity data are unavailable in the involved hospitals, as it is not mandatory for patient registration, leaving the performance of PanCAM in different ethnic subgroups unknown.

In conclusion, PanCAM shows potential as a generalisable and interpretable diagnostic tool for lymph node metastasis detection, particularly nodal micrometastasis. Its robust performance across cancer types and medical centres

highlights its value as a powerful complement to conventional pathological workflow.

Contributors

TL, SW, GH, YunW, HZ, JY, JC, and HC conceived and designed the study. TL, SW, HZ, and HC coordinated multicentre data acquisition. HZ, JY, JC, HC, YH, LH, YuaW, WF, YZ, HH, CW, CX, GD, XH, CQ, AL, JW, XC, CC, WZ, YX, MZ, CL, HW, HY, FJ, and RS contributed to data acquisition. HZ and HC did the pathological image quality control. JY and JC provided pathological review, annotation, and arbitration. ZL, YunW, SH, and RC did the artificial intelligence analysis. SW, GH, YunW, HZ, and ZL did the data analysis and interpretation. SW, GH, and YunW drafted and revised the manuscript. TL, SW, GH, and YunW accessed and verified the data reported in the study. TL and SW supervised the study. All authors had full access to all the data reported in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

All authors declare no competing interests.

Data sharing

A subset of anonymised pathological images has been uploaded to Zenodo (https://doi.org/10.5281/zenodo.15448153). To protect patient privacy, other pathological images and related patient information from the in-house collection are not publicly accessible. These data might be made available on reasonable request to the corresponding author via email. A signed data access agreement will be required before release. The related codes are available at https://github.com/cellsvision/PanCAM.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (U21A20383, 92459303, and 82341018), the National Science and Technology Major Project (2024ZD0525700), the Science and Technology Projects in Guangzhou (2024A03J1190), the Guangdong Provincial Clinical Research Centre for Urological Diseases (2020B1111170006), and the Science and Technology Planning Project of Guangdong Province (2023B1212060013). The authors thank Kehong Zhang from Ivy Medical Editing and Phei Er Saw from Sun Yat-Sen University for their assistance in editing the manuscript.

References

1 Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; **74:** 229–63.
2 Ravi P, Xie W, Buyse M, et al. Refining risk stratification of high-risk and locoregional prostate cancer: a pooled analysis of randomized trials. *Eur Urol* 2025; **87:** 217–24.
3 Ji H, Hu C, Yang X, et al. Lymph node metastasis in cancer progression: molecular mechanisms, clinical significance and therapeutic interventions. *Signal Transduct Target Ther* 2023; **8:** 367.
4 Chang JM, Leung JWT, Moy L, Ha SM, Moon WK. Axillary nodal evaluation in breast cancer: state of the art. *Radiology* 2020; **295:** 500–15.
5 Del Mastro L, De Placido S, Bruzzi P, et al, and the Gruppo Italiano Mammella (GIM) investigators. Fluorouracil and dose-dense chemotherapy in adjuvant treatment of patients with early-stage breast cancer: an open-label, 2 × 2 factorial, randomised phase 3 trial. *Lancet* 2015; **385:** 1863–72.
6 Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684. *J Clin Oncol* 2023; **41:** 425–35.
7 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:** 890–96.
8 Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al, and the CAMELYON16 Consortium. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318:** 2199–210.
9 Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019; **20:** e253–61.
10 van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med* 2021; **27:** 775–84.
11 Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer* 2022; **3:** 1026–38.
12 Skrede O-J, De Raedt S, Kleppe A, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020; **395:** 350–60.
13 Zhao Y, Xiong S, Ren Q, et al. Deep learning using histological images for gene mutation prediction in lung cancer: a multicentre retrospective study. *Lancet Oncol* 2025; **26:** 136–46.
14 Zeng Q, Klein C, Caruso S, et al, and the HCC-AI study group. Artificial intelligence-based pathology as a biomarker of sensitivity to atezolizumab-bevacizumab in patients with hepatocellular carcinoma: a multicentre retrospective study. *Lancet Oncol* 2023; **24:** 1411–22.
15 Yamashita R, Long J, Longacre T, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol* 2021; **22:** 132–41.
16 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:** 222–32.
17 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:** 360–70.
18 van Dooijeweert C, Flach RN, Ter Hoeve ND, et al. Clinical implementation of artificial-intelligence-assisted detection of breast cancer metastases in sentinel lymph nodes: the CONFIDENT-B single-center, non-randomized clinical trial. *Nat Cancer* 2024; **5:** 1195–205.
19 Huang S-C, Chen C-C, Lan J, et al. Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. *Nat Commun* 2022; **13:** 3347.
20 Vestjens JHMJ, Pepels MJ, de Boer M, et al. Relevant impact of central pathology review on nodal classification in individual breast cancer patients. *Ann Oncol* 2012; **23:** 2561–66.
21 Wu S, Wang Y, Hong G, et al. An artificial intelligence model for detecting pathological lymph node metastasis in prostate cancer using whole slide images: a retrospective, multicentre, diagnostic study. *EClinicalMedicine* 2024; **71:** 102580.
22 Stenzinger A, Alber M, Allgäuer M, et al. Artificial intelligence and pathology: from principles to practice and future applications in histomorphology and molecular profiling. *Semin Cancer Biol* 2022; **84:** 129–43.
23 Challa B, Tahir M, Hu Y, et al. Artificial intelligence-aided diagnosis of breast cancer lymph node metastasis on histologic slides in a digital workflow. *Mod Pathol* 2023; **36:** 100216.
24 Chuang W-Y, Chen C-C, Yu W-H, et al. Identification of nodal micrometastasis in colorectal cancer using deep learning on annotation-free whole-slide images. *Mod Pathol* 2021; **34:** 1901–11.
25 Pan Y, Dai H, Wang S, et al. Clinically applicable pan-origin cancer detection for lymph nodes via artificial intelligence-based pathology. *Pathobiology* 2024; **91:** 345–58.
26 Xu H, Usuyama N, Bagga J, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature* 2024; **630:** 181–88.
27 Vorontsov E, Bozkurt A, Casson A, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nat Med* 2024; **30:** 2924–35.
28 Chen RJ, Ding T, Lu MY, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med* 2024; **30:** 850–62.
29 Feng L, Liu Z, Li C, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health* 2022; **4:** e8–17.
30 Adams R, Henry KE, Sridharan A, et al. Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nat Med* 2022; **28:** 1455–60.
31 Liu X, Liu H, Yang G, et al. A generalist medical language model for disease diagnosis assistance. *Nat Med* 2025; **31:** 932–42.

32 Bouzid Z, Sejdic E, Martin-Gill C, et al. Electrocardiogram-based machine learning for risk stratification of patients with suspected acute coronary syndrome. *Eur Heart J* 2025; **46:** 943–54.

33 Kehayias CE, Bontempi D, Quirk S, et al. A prospectively deployed deep learning-enabled automated quality assurance tool for oncological palliative spine radiation therapy. *Lancet Digit Health* 2025; **7:** e13–22.

34 Dembrower K, Crippa A, Colón E, Eklund M, Strand F, and the ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health* 2023; **5:** e703–11.

35 Eisemann N, Bunk S, Mukama T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nat Med* 2025; **31:** 917–24.

36 Haglund F, Garvin S, Ihre-Lundgren C, et al. Detailed lymph node sectioning of papillary thyroid carcinoma specimen increases the number of pN1a patients. *Endocr Pathol* 2016; **27:** 346–51.

37 Ballester M, Dubernard G, Lécuru F, et al. Detection rate and diagnostic accuracy of sentinel-node biopsy in early stage endometrial cancer: a prospective multicentre study (SENTI-ENDO). *Lancet Oncol* 2011; **12:** 469–76.