Virchows Archiv (2025) 487:203–208
https://doi.org/10.1007/s00428-024-03988-1

BRIEF REPORT

Using artificial intelligence to prioritize pathology samples: report of a test drive

Iván Rienda[1] · João Vale[2,3] · João Pinto[2] · António Polónia[2,4] · Catarina Eloy[2,5]

Received: 8 July 2024 / Revised: 14 October 2024 / Accepted: 20 November 2024 / Published online: 3 December 2024

Abstract

The digital transformation of pathology, through automation and computational tools, addresses current challenges in the field. This study evaluates Paige Pan Cancer, a novel artificial intelligence tool based on the Virchow foundation model, designed to flag invasive cancer in haematoxylin and eosin-stained slides from 16 primary tissue types. Using 62 cases from the Ipatimup Pathology Laboratory, we found the tool had a sensitivity of 93.3% and specificity of 87.5% in biopsies, and 94.7% sensitivity and 75.0% specificity in resections. Overall accuracy was 90.3%. Despite some misclassifications, Paige Pan Cancer demonstrates high sensitivity as a multi-organ screening tool in clinical practice.

Keywords computational pathology · cancer diagnosis · artificial intelligence · workflow · efficiency · foundation model

Introduction

The digital transformation of pathology, understood as the automation and standardization of the laboratory workflow leading to the obtention of high-quality whole-slide images (WSIs) for diagnosis, and the subsequent application of computational tools are called to become a major aid in overcoming the challenges that pathology faces [1]. Computational pathology can aid to decrease workload, provide precise quantitative information, or even assist pathologists in diagnosis [2].

An important barrier to implement computational pathology in clinical use is the lack of enough curated clinical data and manual annotations needed for training-supervised artificial intelligence (AI) models robust to be generalizable in the clinical arena [3]. Additionally, the scarcity of good-quality training data leads to task-specific AI models, limiting their trans-topographic use [4]. Self-, semi-, or unsupervised training techniques may avoid the need of task-specific annotated data, allowing the development of foundation models, which are trained on huge amounts of mainly unannotated data resulting in a wide range of downstream tasks [4].

Different efforts have been made recently to bring foundation model-based tools to pathology, with a successful multitask performance [5, 6]. A novel tool, Paige Pan Cancer, based on the Virchow foundation model, has been developed to flag cases suspicious for invasive cancer [7]. This triage tool is designed to enable the detection and identification of foci suspicious for malignancy in hematoxylin–eosin (H&E)-stained slides from a biopsy material across 16 primary tissue types, including the bladder, bone, brain, breast, cervix, colon, endometrium, liver, lung, lymph node, pancreas, peritoneum, prostate, skin, stomach, and the upper gastrointestinal tract. Paige Pan Cancer generates a binary result (either non-suspicious or suspicious for cancer) at the case and slide level and pinpoints areas of interest with a high likelihood of containing invasive cancer in suspicious slides.

In this study, we externally evaluate the performance of Paige Pan Cancer, at the case level, using a diverse set of cases sourced from the Pathology Laboratory of IPATIMUP.

✉ Catarina Eloy
catarinaeloy@hotmail.com

1 Department of Pathology, Hospital Universitari I Politècnic La Fe, Valencia, Spain

2 Pathology Laboratory, Institute of Molecular Pathology and Immunology of University of Porto (IPATIMUP), Rua Júlio Amaral de Carvalho 45, 4200-135 Porto, Portugal

3 Department of Pathological, Cytological and Thanatological Anatomy, School of Health of Polytechnic Institute of Porto, Porto, Portugal

4 Escola de Medicina E Ciências Biomédicas, Universidade Fernando Pessoa, Porto, Portugal

5 Pathology Department, Medical Faculty of University of Porto, Porto, Portugal

Methods

A retrospective series of 62 cases diagnosed during 2023 were obtained from the digital archive of the fully digital Pathology Laboratory of IPATIMUP. The series was selected through conditional random sampling, ensuring that half of the series corresponded to resection specimens, while the other half consisted of biopsy samples. A single representative WSI (generated using the Pannoramic 1000® 3DHISTECH Ltd, Budapest, Hungary, at×20 objective with a digital resolution of 41x (0.25 µm/pixel)) per case was selected and anonymized, and clinicopathological data was retrieved and recorded in a single database.

The series was divided into two cohorts: 31 resection samples (RS) and 31 biopsy samples (BS), including those with incisional or excisional purposes. The biopsy samples (BS) cohort comprised 30 patients (14 males and 16 females) with ages between 41 and 89 years old (average 62.0 years old) and 13 tissue types, including the bladder (4 cases), brain (1), breast (3), colon (2), head and neck (4), liver (2), lung (3), lymph node (3), pancreas (2), prostate (2), skin (1), stomach (2), and uterus (2). In this cohort, 48.4% (15/31) of cases were diagnosed as malignant. This cohort included 2 out of 4 bladder cases (6.5%, 2/31) of non-invasive urothelial carcinomas (pTa), and they were considered malignant as defined by the WHO Classification of Tumours (5th edition). This cohort did not include tumors of uncertain malignant potential/borderline lesions.

The resection samples (RS) cohort consisted of 31 patients (13 males and 18 females) with ages between 20 and 82 years old (average 60.1 years old) and 13 tissue types, with cases including the bladder (2), breast (1), cervix (1), colon (2), lung (2), lymph node (3), ovary (2), pancreas (2), prostate (2), skin (3), stomach (2), head and neck (5), and uterus (4). Cancer diagnosis was reported in 61.3% (19/31). A case of non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) was not considered under the malignant group, representing a 3.2% (1/31) of the RS cohort. Diagnosis details of the cases are listed in Table 1.

This dataset comprising both biopsy and resection specimens from multiple organs was uploaded to the Paige FullFocus image management system and analyzed using the Paige Pan Cancer tool (v1.1 BETA). The dataset comprised specimens both within the intended use of the AI tool (biopsy material of the 16 topographies listed above) but also specimens outside the intended use (all resections, as well as different topographies), with the idea to stress the predictive diagnostic potential of this tool.

Results

The average time spent uploading each WSI was 4.017 min.

In the BS cohort, 28 out of 31 cases (90.3%) were accurately classified, yielding a sensitivity of 93.3% and a specificity of 87.5%. The misclassifications included 2 false-positive cases and 1 false-negative case (Table 2).

For the RS cohort, 27 out of 31 cases (87.1%) were correctly classified, resulting in a sensitivity of 94.7% and a specificity of 75.0%. Within this group, there were 3 false-positive cases and 1 false-negative case (Table 2).

The overall accuracy across both cohorts was 90.3% (56 out of 62 cases), with an overall sensitivity of 94.3% and an overall specificity of 85.2%.

In case 33, a radical prostatectomy sample removed after radiotherapy was flagged by Paige Pan Cancer as suspicious. The original diagnosis was negative for cancer but, after the AI output and subsequent immunohistochemistry (IHC), a 1-mm focus of carcinoma was detected. This case 33 was considered a correct interpretation by the AI.

Within the cases correctly classified by Paige Pan Cancer, the focus of interest (FOI) was incorrectly located in 3.1% (1/32) of cases (case 33, Fig. 1), and comprised in the tumor but in a marginal fashion in 46.9% (15/32) of cases.

Discussion

Paige Pan Cancer detects slides suspicious for malignancy with a good overall performance in H&E-stained WSIs, opening the door to a clinically applicable multi-organ screening tool. Some AI-assisted tasks, as biomarker prediction, have been suggested to not depend on the tissue sample size [8]; in the case of the suspicious case detection with Paige Pan Cancer, the software seems to be slightly better in the BS cohort of our study, where the tissue areas were generally smaller than in the RS cohort. This observation goes in line with the current intended use of Paige Pan Cancer which is recommended to be performed in biopsy specimens and not in resections. Nevertheless, the performance is very good also in resection specimens, as demonstrated in our series. These results need to be confirmed with the application of the model over a larger case series.

When exposed to topographies outside of the intended use and borderline lesions or malignant, considered non-invasive neoplasms, Paige Pan Cancer had, as expected, somewhat a lower performance (i.e., ovarian cases were correctly classified, while thyroid cases and papillary urothelial neoplasms were highly represented within the misclassified samples).

An important observation in our study was that Paige Pan Cancer was misclassified as suspicious for malignancy two WSI from the RS cohort with fixation defects. The variations in fixation and other preanalytical variables need to be observed in further evaluations of the model. In the difficult trade-off between sensitivity and specificity, apparently, Paige Pan Cancer has been trained with a sensitivity bias and

Table 1 Diagnostic details

Case number	Gender	Age	Topography	Cohort	Diagnosis
1	F	45	Breast	BS	No evidence of malignancy
2	F	50	Breast	BS	Invasive breast carcinoma of no special type
3	F	41	Breast	BS	Benign fibroepithelial lesion
4	M	77	Prostate	BS	No evidence of malignancy
5	M	71	Prostate	BS	Acinar adenocarcinoma
6	F	74	Lung	BS	No evidence of malignancy
7	M	57	Lung	BS	Metastatic adenocarcinoma
8	M	66	Lung	BS	NSCLC
9	F	44	Colon	BS	Chronic active inflammation
10	F	79	Colon	BS	Adenocarcinoma
11	F	44	Skin	BS	Compatible with systemic sclerosis
12	F	50	Lymph node	BS	No evidence of malignancy
13	F	54	Lymph node	BS	Lymphoma B
14	M	62	Lymph node	BS	Lymph node metastasis
15	M	89	Bladder	BS	No evidence of malignancy
16	M	46	Bladder	BS	Inverted urothelial papilloma
17	M	64	Bladder	BS	Papillary urothelial carcinoma, HG
18	F	74	Bladder	BS	Papillary urothelial carcinoma, LG
19	F	46	Uterus	BS	No evidence of malignancy
20	F	49	Uterus	BS	No evidence of malignancy
21	M	68	Pancreas	BS	No evidence of malignancy
22	F	76	Pancreas	BS	Adenocarcinoma
23	M	44	Head and neck	BS	Nasopharyngeal lymphoid hyperplasia
24	M	44	Head and neck	BS	Larynx polyp
25	M	84	Head and neck	BS	Pharynx SCC
26	F	78	Head and neck	BS	Larynx SCC
27	M	54	Liver	BS	Steatotic liver
28	F	57	Liver	BS	Metastatic carcinoma
29	M	81	Stomach	BS	No evidence of malignancy
30	M	76	Stomach	BS	Adenocarcinoma
31	F	77	Brain	BS	Glioblastoma, *IDH* wild type
32	F	77	Breast	RS	Invasive breast carcinoma of no special type
33	M	61	Prostate	RS	Acinar adenocarcinoma
34	M	57	Prostate	RS	Acinar adenocarcinoma
35	F	20	Lung	RS	Pulmonary emphysema
36	M	64	Lung	RS	SCC
37	F	57	Colon	RS	Ileocolic anastomotic inflammation
38	M	72	Colon	RS	Adenocarcinoma
39	F	64	Skin	RS	No evidence of malignancy
40	F	78	Skin	RS	BCC
41	F	65	Skin	RS	Melanoma
42	M	49	Lymph node	RS	Follicular hyperplasia
43	M	82	Lymph node	RS	Lymph node metastasis
44	F	58	Lymph node	RS	Lymphoma B
45	M	72	Bladder	RS	No evidence of malignancy
46	M	70	Bladder	RS	Invasive urothelial carcinoma, HG
47	F	46	Uterus	RS	No evidence of malignancy
48	F	81	Uterus	RS	Endometrioid carcinoma
49	F	53	Uterus	RS	Carcinosarcoma
50	F	77	Uterus	RS	Serous carcinoma
51	M	57	Pancreas	RS	Tumor-free surgical margin

Table 1 (continued)

Case number	Gender	Age	Topography	Cohort	Diagnosis
52	M	67	Pancreas	RS	Ductal adenocarcinoma
53	F	45	Thyroid	RS	No evidence of malignancy
54	M	47	Thyroid	RS	Papillary thyroid carcinoma
55	M	72	Thyroid	RS	Follicular carcinoma
56	M	54	Thyroid	RS	Poorly differentiated thyroid carcinoma
57	F	65	Thyroid	RS	NIFTP
58	F	57	Ovary	RS	No valuable alterations
59	F	28	Ovary	RS	Small cell carcinoma, hypercalcemic type
60	F	48	Stomach	RS	Chronic gastritis
61	F	81	Stomach	RS	Tubular adenocarcinoma
62	F	39	Cervix	RS	No evidence of malignancy

Biopsy samples cohort: cases 1–31; resection samples cohort: cases 32–62

F female, *M* male, *RS* resection samples cohort, *BS* biopsy samples cohort, *NSCLC* non–small cell lung cancer, *SCC* squamous cell carcinoma, *BCC* basal cell carcinoma, *HG* high grade, *LG* low grade, NIFTP non-invasive follicular thyroid neoplasm with papillary-like nuclear features

Table 2 Misclassified cases

Cohort	Case number	Error type	Tissue type	Procedure	Diagnosis	Comment
RS	37	FP	Colon	Ileocolic resection	Ileocolic anastomosis inflammation	Defect of fixation, lymphoid tissue marked as FOI
RS	47	FP	Uterus	Total hysterectomy	No evidence of malignancy	Defect of fixation, unfixed endometrial tissue marked as FOI
RS	57	FP	H&N (thyroid)	Total thyroidectomy	NIFTP	Non-invasive encapsulated tumor with extremely low malignant potential
RS	56	FN	H&N (thyroid)	Total thyroidectomy	Poorly differentiated thyroid carcinoma	
BS	16	FP	Bladder	TURBT	Inverted urothelial papilloma	Non-invasive papillary urothelial neoplasm with benign behavior
BS	23	FP	H&N (nasopharynx)	Curettage	Nasopharyngeal lymphoid hyperplasia	Big biopsy specimen
BS	17	FN	Bladder	TURBT	Papillary urothelial carcinoma, high grade	Non-invasive papillary urothelial neoplasm (pTa) with malignant behavior

RS resection samples cohort, *BS* biopsy samples cohort, *FP* false positive, *FN* false negative, *H&N* head and neck, *TURBT* transurethral resection of bladder tumor, *FOI* focus of interest, *NIFTP* non-invasive follicular thyroid neoplasm with papillary-like nuclear features

associated greater expectation of false positives. However, these false positives should be apparent to the pathologist.

Like any other diagnostic aid, the intended use of these AI tools is to assist the pathologist in their routine workflow, and not to produce definitive diagnoses, so the combined diagnostic system of the pathologist plus AI is greater than either element alone. Different AI-based tools have shown that their application saves costs and time in task-specific settings without compromising the diagnostic accuracy [9, 10]. In our study, the discovery of carcinoma in a sample initially diagnosed as negative suggests that such diagnostic aids could enhance sensitivity compared to pathologists working independently. However, this observation requires further validation in future studies, which could also compare the costs and time savings for pathologists using Paige Pan Cancer versus not using it in a real-world setting. In this particular case, the discovery a small group of glands (less than 1 mm) with cancer features gave rise to an amendment report and did not affect the final patient management.

In terms of verification of the results, an important limitation of Paige Pan Cancer is the lack of a defined demarcation tool that highlights the suspicious tissue; instead, the software offers a FOI in suspicious cases, a single point in

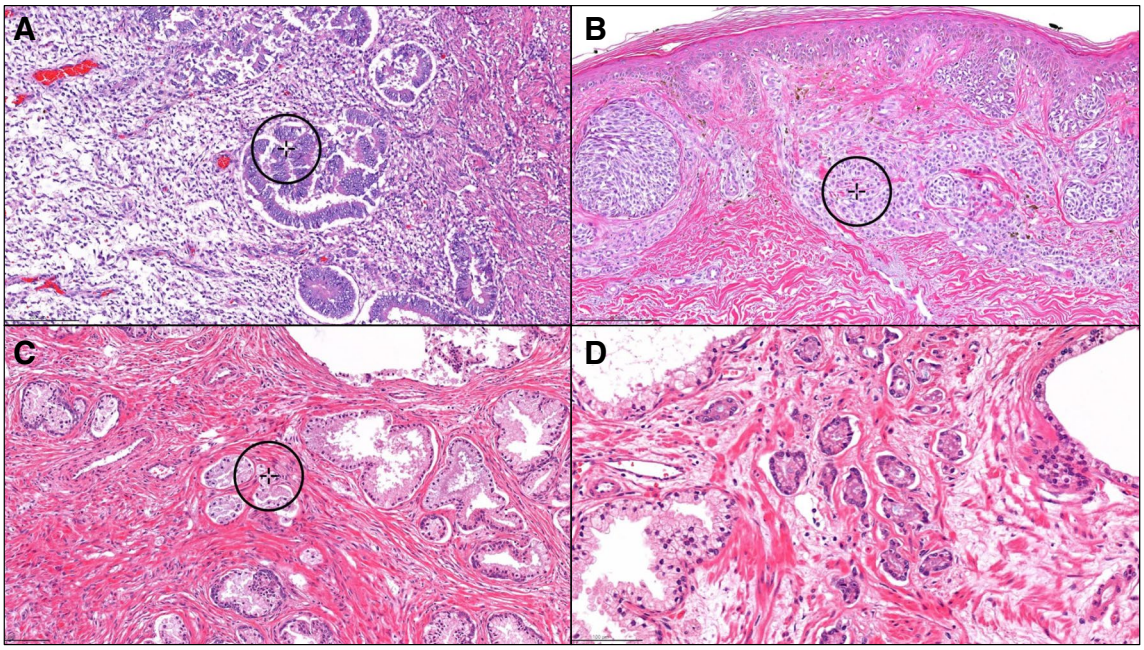

Fig. 1 **A** False positive (case 47), the software pointed out normal endometrial glands with fixation artifact as suspicious for cancer. **B** True positive (case 41), the FOI is correctly located pointing to malignant melanic cells. **C**, **D** True positive with wrong location of the FOI (case 33), the crosshair is located over benign glands (**C**), while acinar adenocarcinoma is visible in a different location of the same WSI (**D**)

which it considers that the image contains invasive cancer. The location of the FOI highlighted correctly invasive carcinoma in most of the suspicious cases, but it failed in a challenging case, in which IHC was required to confirm the malignant diagnosis. Paige Pan Cancer was trained on WSIs obtained from Leica Aperio AT2 and GT450 scanners, while our dataset was generated with the 3DHISTECH's Pannoramic 1000 scanner. It is known that analytical variations such as differences in WSI scanning can affect the performance of AI algorithms [11]. However, a strong topographic correlation has been achieved between different scanners after AI model application [12]. The FOI location fail may obey to a topographic misalignment caused by the differences between scanners used in the training of the model and in our study cases, but the absence of a more robust explainability tool does not facilitate the understanding of why this challenging case was correctly classified.

According to current expert perspectives, AI models are expected to have profound impact on pathology over the next decade. However, there are significant challenges to overcome, such as the digital transformation of pathology laboratories, the proper integration of algorithms into laboratory information systems, ensuring that models meet strict regulatory and ethical requirements, and making their implementation economically efficient [13].

In summary, our work shows that this beta version of Paige Pan Cancer has high sensitivity in the detection of cancer in different topographies, at an acceptable false-positive rate for a screening tool, even in specimens outside its intended use. It generalizes well to data different from the training dataset. Some limitations include its narrow capability to show tissue regions deemed suspicious at the slide level. Even though Paige Pan Cancer is also a task-specific model, its multi-organ character illustrates the generalizability of foundation model-derived AI tools and their use in clinical practice, which potentially could multiply the efficiencies in comparison with single-organ task-specific models.

Acknowledgements The authors thank Paige AI for allowing this experiment in the FullFocus platform with the Paige Pan Cancer/Virchow software (v1.1 BETA).

Data availability Data generated or analyzed during this study are partially included in this published article and may be provided after reasonable request.

Declarations

Ethics approval and consent to participate Formal ethical approval was not required for this study, as it did not involve any patient data collection or impact on patient care.

Conflict of interest CE consulted for Mindpeak, MSD, Biocartis, Diapath, Sakura, and Leica during the last 2 years. AP consulted for Indica Labs for the last 2 years. The use of the Paige Pan Cancer was warranted by Paige AI.

References

1. Eloy C (2023) Postponing evolution: why are we choosing to ignore the need for a digital transformation in pathology? Virchows Arch. https://doi.org/10.1007/s00428-023-03714-3
2. Aeffner F et al (2019) Introduction to digital image analysis in whole-slide imaging: a white paper from the Digital Pathology Association. J Pathol Inform 10:9
3. Abels E et al (2019) Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. J Pathol 249:286–294
4. Waqas A et al (2023) Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models. Lab Invest 103:100255
5. Chen RJ et al (2024) Towards a general-purpose foundation model for computational pathology. Nat Med 30:850–862
6. Lu MY et al (2024) A visual-language foundation model for computational pathology. Nat Med 30:863–874
7. Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Liu S, Severson K, Zimmermann E, Hall J, Tenenholtz N, Fusi N, Mathieu P, Van Eck A, Lee D, Viret J, Robert E, Wang YK, Kunz JD, Lee MCH, Bernhard J, Godrich RA, Oakley G, Millar E, Hanna M, Retamero J, Moye WA, Yousfi R, Kanan C, Klimstra D, Rothrock B, Fuchs TJ. Virchow: a million-slide digital pathology foundation modelXiv:2309.07778v5 [eess.IV]
8. Arslan S et al (2024) A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. Commun Med 4:48
9. Kacew AJ et al (2021) Artificial intelligence can cut costs while maintaining accuracy in colorectal cancer genotyping. Front Oncol 11:630953
10. Eloy C et al (2023) Artificial intelligence-assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Arch 482:595–604
11. Duenweg SR et al (2023) Whole slide imaging (WSI) scanner differences influence optical and computed properties of digitized prostate cancer histology. J Pathol Inform 14:100321
12. Vazzano J et al (2023) Evaluation of a computer-aided detection software for prostate cancer prediction: excellent diagnostic accuracy independent of preanalytical factors. Lab Invest 103:100257
13. Berbís MA et al (2023) Computational pathology in 2030: a Delphi study forecasting the role of AI in pathology within the next decade. EBioMedicine 88:104427