Virchows Archiv
https://doi.org/10.1007/s00428-026-04594-z

ORIGINAL ARTICLE

An artificial intelligence solution for evaluation of prostate needle core biopsy specimens: impact in a non-specialist setting

Gerald Niedobitek[1,4] · Gernot Schmitz[1] · Manuel Fella[1] · André Oliveira[1] · John Theunissen[2] · Manuela Vecsler[2] · Jens Köllermann[3]

Received: 22 January 2026 / Revised: 5 March 2026 / Accepted: 19 May 2026
© The Author(s) 2026

Abstract

The diagnosis of prostate cancer rests on the histopathological evaluation of prostate needle core biopsy specimens (NCBS) supplemented by immunohistochemistry as required. Because of the well-known interobserver variability in diagnosis and grading of prostate cancer, we have studied the capability of a commercially available artificial intelligence (AI) solution to aid in the diagnosis of PCa in a non-expert setting. For this, 2828 H&E stained NCBS slides from 249 cases were digitised to produce whole slide images (WSI) and subjected to second-read analysis using the Ibex prostate AI solution. Ground truth was established by a combination of primary pathologists' diagnoses supplemented by immunohistochemistry and external expert revision. For cancer detection, the AI solution achieved a false negative rate of 0.2%, significantly lower than that of the reporting pathologists (1.5%). False positive rates were similar for reporting pathologists (2.3%) and prostate AI solution (3.1%). A Gleason pattern 4 alert was raised by the AI solution in 2 of 29 cases (7%) of cases originally diagnosed as Gleason score 3+3, both of which were upheld after expert review. Because of the therapeutic consequences, the triggering of a "higher than 3+3 Gleason score" alert represents a useful safety feature. In conclusion, our findings support the growing evidence that by enhancing diagnostic accuracy AI-based diagnostic solutions can play a pivotal role in alleviating the increasing workload faced by pathologists due to the rising global incidence of prostate cancer.

Introduction

Prostate cancer (PCa) is the most common cancer in males worldwide and the number of newly diagnosed cases is projected to double between 2020 and 2040 [1, 2]. The diagnosis of PCa rests on the histopathological assessment of prostate needle core biopsy specimens (NCBS). Usually at least 12 random samples are taken which may be supplemented by targeted NCBS from lesions identified by multi parametric MRI, leading to a significant workload for pathologists [1, 3].

Further treatment decisions rest on the correct identification of prostate carcinoma, determination of tumour extent in the NCBS and accurate grading [4–6]. Because of the well-known interobserver variability in diagnosis and grading of prostate cancer, standard revision by a second pathologist would be desirable to minimise errors with impact on therapy decisions [7, 8]. It is common practice in many pathology departments that a new diagnosis of PCa is confirmed by a second pathologist while negative results are not usually subject to review. However, this is not subject to standardised recommendations. In everyday practice, high workload in combination with the widespread lack of histopathologists will likely result in confirmation of malignancy by a second pathologist based on review of a representative slide only in most cases [9]. Large scale second review by external expert pathologists is not common procedure and is unlikely to be feasible given the above considerations.

DedicationThis work is dedicated to the memory of John Theunissen.

✉ Gerald Niedobitek
niedobitekg@gmail.com

1 Institute for Pathology, Sana Klinikum Lichtenberg, Berlin, Germany

2 Ibex Medical Analytics, Tel Aviv, Israel

3 Dr. Senckenberg Institute of Pathology, University Hospital Frankfurt, Goethe University Frankfurt, Frankfurt/M, Germany

4 Institut für Pathologie, Klinikum Ernst von Bergmann, Charlottenstr. 72, 14467 Potsdam, Germany

Published online: 03 June 2026

The advent of digital pathology aided by high throughput scanners has facilitated the introduction of artificial intelligence (AI) solutions to support histopathological diagnosis [10–15]. Several studies have reported AI-based systems for the diagnosis and grading of prostate cancer and suggest that such algorithms can accurately detect and grade prostate cancer in NCBS [15–18]. A direct review of multiple AI algorithms by expert uropathologists from the United States and Europe, has shown a high agreement between AI solutions and expert pathologists [19].

The aim of this study was to analyse the capability of a commercially available AI solution to aid in the diagnosis of PCa in a non-expert setting. Specifically, we wanted to address the questions if an AI-based solution could increase the accuracy of PCa diagnosis, reduce the number of false negative results, and improve the grading of PCa.

Materials and methods

Study cohort

For this retrospective study, NCBS from 249 patients comprising a total of 2828 NCBS were retrieved from the archives of the Institute of Pathology, Sana Klinikum Lichtenberg, Berlin. For each NCBS, one H&E slide containing 5 sections, was generated during routine diagnostic work-up. All cases were collected in a three-year period from 2018 to 2021 representing the entire prostate NCBS workload during this time period. The specimens were obtained from exclusively male patients with an age range from 50 to 93 (average of 71) years. Among these, 167 cases (67.1%) were originally diagnosed as PCa, while 82 cases (32.9%) were reported as benign. For 58 cases, immunohistochemistry was utilised as described below to obtain the primary diagnosis. Of the 2828 NCBS slides, 1878 slides (66.4%) were diagnosed as benign and 950 slides (33.6%) as PCa.

Study design

All slides were scanned using a Philips Ultra Fast Scanner (UFS, Philips, Hamburg, Germany) at 40x magnification (resolution of 0.25 µm/pixel) and up-loaded onto the Ibex Prostate AI platform (Ibex Medical Analytics) via the Philips Image Management System. Once up-loaded, the original diagnoses were provided on a case-by-case basis as benign vs. malignant. For all cases with a PCa diagnosis, the summary Gleason score was also provided. The AI solution analysed H&E-stained WSIs only to produce AI results on slide level that were then compared to the provided diagnoses for benign vs. malignant and Gleason score, specifically Gleason 6 versus Gleason 7 and above.

This resulted in three scenarios: no alerts, technical alerts, or clinical alerts. No alert indicated that the Ibex Prostate AI solution agreed with the assessment of the reporting pathologist. Technical alerts related to out of focus images not suitable for AI analysis. This alert was raised when more than 20% of any slide was out of focus and resulted in exclusion of this NCBS slides from the case analysis; the case as a whole remained in the analysis. There were two types of clinical alerts: 1). A "cancer detected" alert was given in all cases where the original diagnosis was benign, and the AI application identified areas with a high cancer likelihood (score above 95%); 2). A "Gleason Score above 3 + 3" alert was raised for all cases with an original diagnosis of prostate cancer, Gleason score 3 + 3 in which AI analysis indicated areas of higher Gleason patterns resulting in high likelihood for a higher Gleason score. The algorithmic score that triggered an alert (i.e., alert threshold) when compared to the initial diagnoses was set to correspond to a specificity of 95% for cancer on slide level from previous cohorts [11, 14]. Thus alerts should be raised for the 5% most suspicious slides.

All clinical alerts were resolved by external genitourinary expert review (JK) of the H&E slides and if necessary, by carrying out additional immunohistochemical stains. For this, double labelling immunohistochemistry was carried out using monoclonal antibodies 34βE12 (Agilent Technologies, Waldbronn Germany), specific for high molecular weight cytokeratins, and 13H4 (Zytomed Systems, Berlin, Germany), directed against alpha-methylacyl-CoA-racemase (AMACR) on the Ventana Benchmark Ultra platform (Roche, Mannheim, Germany).

In a second round of analysis, examination at a single NCBS slide level was carried out by separating the raw cancer probability data generated by the Ibex Prostate AI solution into three groups: "high cancer likelihood" (AI cancer score$\geq$0.95), "low cancer likelihood" (AI cancer score$\leq$0.8), and "suspicious" (AI cancer score$>$0.8 to $<$0.95). Subsequently, the original diagnoses rendered on each NCBS were grouped as "benign" or "PCa" and compared to the Ibex Prostate AI solution results.

Results

First, all cases were analysed for alerts raised by the prostate AI cancer algorithm. 47 NCBS slides (1.7%) pertaining to 36 cases were excluded from the analysis because of an "out of focus" technical alert. These technical alerts accounted for at most 4 NCBS per case. No case was excluded from AI examination because of out of focus alerts.

"Cancer detected" alerts were raised in 13 of 82 cases that were originally designated as "benign" (16%), indicating

carcinoma may have been missed by the reporting pathologists (Table 1; Fig. 1A, B). Examination of these cases by an expert prostate pathologist (JK) supplemented by immunohistochemistry as required revealed cancer in five cases (6.1%) and one case with atypical small acinar proliferation (ASAP, 1.2%). In seven cases (8.5%), the original diagnosis as "benign" was maintained (Table 1). Atrophy and tissue contaminants were the most common causes of false positive cancer alerts. Of the 5 additional cancer cases identified by the Ibex Prostate AI solution and upheld following expert review, 4 displayed single small foci of Gleason score 3 + 3 PCa in one NCBS each. In one case, a small cribriform growth was identified which displayed an intact basal cell layer in immunohistochemistry and thus was judged to probably represent intraductal carcinoma. Adjacent to this, a very small focus of Gleason pattern 4 invasive cancer was detected.

Additionally, the Ibex Prostate AI solution raised two "Gleason higher than 3 + 3" alerts among 29 PCa cases originally diagnosed with Gleason score 3 + 3 PCa (7%, Fig. 1C, D). After expert review both cases were revised to Gleason 3 + 4, although in one case, the proportion of Gleason pattern 4 barely reached 5%.

Next, we analysed the results at the single NCBS level. Pathologists' diagnoses and Ibex Prostate AI solution results matched in 2601 out of 2781 (93.5%) NCBS, with 1712 samples diagnosed as benign rendering a low cancer probability score (92%) and 889 samples diagnosed as cancer yielding high cancer probability scores (96%, Table 2). In 180 (6%) NCBS, the results were mismatched, consisting of 39 slides originally diagnosed as PCa and 141 slides originally reported as benign (Table 2). Of the 39 NCBS with an original diagnosis of PCa, the Ibex Prostate AI solution scores indicated a low cancer probability in 22 NCBS and an intermediate cancer probability score in 17 NCBS (Table 2). In one case, this would have changed the overall diagnosis on case basis from PCa to benign. Following expert review, the original diagnosis of PCa was retained in this case.

Conversely, of 141 NCBS samples originally reported as benign by the reporting pathologist, cancer probability scores were high in 68 samples and intermediate in 73 NCBS (Table 2). These mismatches between AI and pathologists' diagnoses did not raise alerts as they pertained to cases that were already diagnosed with cancer in at least one other NCBS; thus, an overall diagnosis of PCa had already been entered as reporting pathologists' diagnosis.

Next, the 180 NCBS with mismatches between reporting pathologists' diagnosis and Ibex prostate AI solution assessment were examined in more detail. For a summary see Table 3. Slides of four NCBS were not available for expert review (Table 3).

Of 22 mismatches with a reporting pathologists' diagnosis of PCa and low cancer probability scores by Ibex prostate AI solution assessment, expert pathologist review confirmed the AI assessment for 14 NCBS (14/22, 64%; Table 3). Seven of these 22 NCBS were considered to contain PCa or ASAP lesions.

In 68 NCBS, a discrepant reporting pathologists' assessment as benign and high cancer probability scores as recorded by the AI solution was observed. PCa was confirmed in 26 of these NCBS (38%, Table 3) following expert review.

Compared to the "ground truth" as established by expert review supplemented by immunohistochemistry, 40 of 1.853 NCBS diagnosed as benign by pathologists were not confirmed, resulting in a false negative rate of 2.2%. Considering only definite diagnoses of cancer as established by expert review, the false negative rate of pathologists was 1.5% (27/1.853). Of 928 NCBS considered malignant by the reporting pathologist, 28 were not confirmed by expert review, resulting in a false positive rate of 3%. Considering only cases judged to be benign or containing PIN lesions by expert review as discrepant, the false positive rate of reporting pathologists was 2.3% (21/928).

For the AI solution, only cases with a low cancer probability score were considered as benign. Of 1.734 NCBS with a low cancer probability score, one was not available for expert review. Of the remaining NCBS, expert review considered 7/1.733 NCBS to harbour either ASAP lesions or cancer, resulting in a false negative rate of 0.4%. Considering only definite cancer diagnoses, the false negative rate for the AI solution was 0.2% (3/1.733). In 957 NCBS, the AI solution rendered high cancer probability scores. Expert review did not confirm this for 42 NCBS resulting in a false positive rate of 4.3%. Considering only NCBS diagnosed

Table 1 Summary of clinically relevant alerts on a case basis

Case	No. of alerts	AI comment	Resolution after expert review
1	2	Cancer detected	Cancer
2	3	Cancer detected	Benign / Tissue contaminants
3	1	Cancer detected	ASAP
4	1	Cancer detected	Tissue contaminants
5	2	Cancer detected	Benign
6	1	Cancer detected	Benign
7	1	Cancer detected	Cancer
8	1	Cancer detected	Atrophy
9	1	Cancer detected	Atrophy
10	1	Cancer detected	Cancer
11	1	Cancer detected	Cancer
12	1	Cancer detected	Cancer
13	1	Cancer detected	Benign
14	1	Higher Gleason Score detected	Gleason 3 + 4 ISUP 2
15	1	Higher Gleason Score detected	Gleason 3 + 4 (GP4 barely 5%)

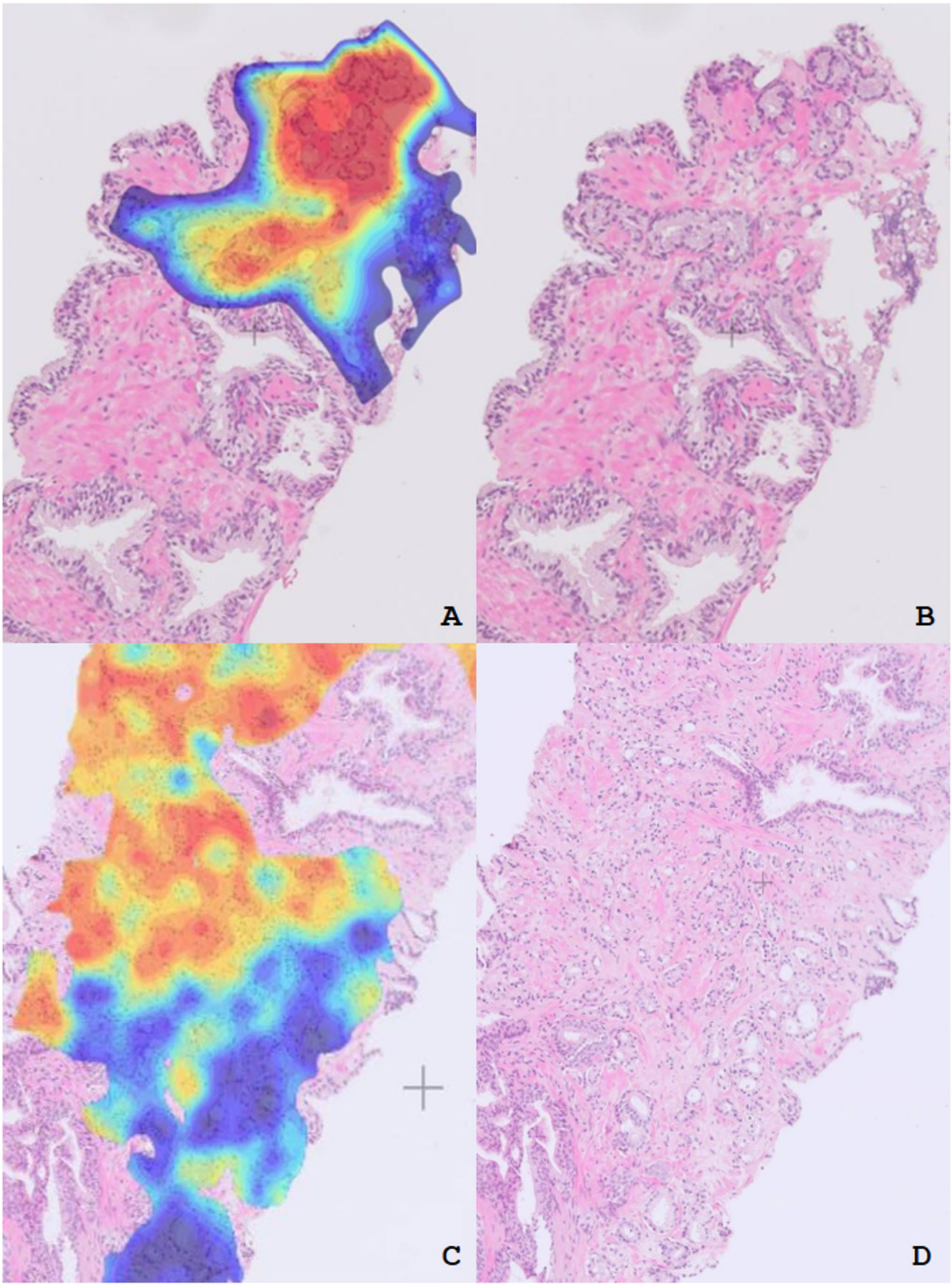

A
B
C
D

Fig. 1 **A** Heatmap indicates a region of high cancer probability (red/orange) as identified by the prostate AI solution. **B** The corresponding H&E stained images reveals a small focus of Gleason score 3+3 prostatic adenocarcinoma missed by the reporting pathologist. **C** Heatmap indicates an area with a high probability (red/orange) of a higher Gleason score than the originally reported 3+3 pattern. **D** The corresponding HE stained image shows an area with small fused glands corresponding to Gleason pattern 4

as benign or PIN as "false", the false positive rate of the AI solution was 3.6% (34/957 NCBS). For 4 NCBS, high cancer probability scores were due to the presence of tissue contaminants on the slides. Excluding these 4 slides from the analysis, the false positive rate of the AI solution was 3.1% (30/953 NCBS).

We also computed the AUC to measure the accuracy of the Prostate AI scores. Our findings demonstrated that the AI algorithm achieved high accuracy for cancer detection, with an AUC of 0.996 (95%CI: 0.994; 0.998) for distinguishing between benign cases and those with PCa. The sensitivity and specificity of the prostate AI solution were 96.3% (95% CI: 94.9%, 97.4%) and 98.9% (95% CI: 98.2%, 99.3%), respectively. Further, the PPV and NPV were 97.8% (95% CI: 96.8%, 98.7%) and 98.1% (95% CI: 97.5%, 98.7%), respectively. Finally, the accuracy for high Gleason score (Gleason 3+3 vs. Gleason 7 and above) was 0.948 (95% CI: 0.922, 0.965), demonstrating high accuracy.

Discussion

Prostate carcinoma has accounted for approximately 1.4 million newly diagnosed cases worldwide in 2020 [2] . Particularly high rates have been recorded amongst others, in western Europe [2]. Notably, it has been predicted that prostate carcinoma case numbers will more than double until 2040 to around 2.9 million new cases per year globally leading to an additional workload for already strained histopathology services [2, 10]. This situation is confounded by the need for second opinions particularly in cancer diagnosis [8]. It is now common practice to subject newly diagnosed cancer cases to intradepartmental review, meaning that each case has to be seen by at least two pathologists. In addition, in prostate cancer, the well-known variations in assessment of Gleason grade between pathologists together with the important implications of assigning cases to low grade (Gleason score 3 + 3 or 3 + 4) as opposed to high grade (Gleason score 4 + 3 and above) groups may trigger external expert review in many cases [7, 20]. It is generally anticipated that AI solutions will help to alleviate the histopathology workload thus at least partially offsetting the observed reduction in the numbers of practicing pathologists [9, 10, 13, 21].

We were therefore interested to test the impact of a commercially available prostate AI solution on the histopathological assessment of prostate NCBS in a non-specialist setting. Several prostate cancer AI solutions have been reported, and it has generally been concluded that their diagnostic performance equals that of pathologists, both with regards to diagnosis of carcinoma as well as to differentiating between low and high Gleason grades [14, 15, 18, 19, 21, 22]. Specifically, in our study an AI prostate solution was applied in a second review setting, i.e., following primary evaluation of cases by pathologists. Ground truth was established by a combination of primary pathologists' diagnoses supplemented by immunohistochemistry and external expert revision.

At case level, 13 missed cancer alerts were raised in 82 cases diagnosed as benign by the reporting pathologists (16%). Following expert review, the diagnosis was changed from benign to malignant in 5 cases (6.1%). Atrophy and tissue contaminants were the most likely causes of false positive cancer alerts generated by the AI solution. Thus, use of the second review prostate AI application contributed to the detection of a small number of PCa cases, missed by the reporting pathologists. In most cases, single small foci of Gleason pattern 3 PCa were identified, but in one case a clinically relevant focus of Gleason pattern 4 PCa with adjacent intraductal spread was identified.

In addition, the prostate AI solution identified 2 of 29 case originally reported a Gleason 3+3 with Gleason pattern 4 areas (7%). Both alerts were upheld following expert review, although the Gleason 4 pattern was only very small (5%) in one case.

The prostate AI solution as used in this study raised only two types of alerts at case level, "cancer detected" alerts and "Gleason higher than 3+3" alerts. This implies that, conversely, cases diagnosed as carcinoma by the reporting pathologist would not generate an alert, even if the cancer probability score of the prostate AI solution was low. For this reason, we next carried out an in-depth analysis of the data at the single NCBS level.

At the single slide level, 27 of 1.853 slides originally diagnosed as benign by the reporting pathologists were considered to harbour foci of prostatic adenocarcinoma following expert review resulting in a pathologists' false negative rate of 1.5%. By contrast the prostate AI solution recorded low cancer probability values for 1.733 NCBS. Following expert review, 3 were judged to contain foci of adenocarcinoma. Thus, the AI false negative rate was 0.2%, significantly lower than that of the reporting pathologists. ($p<0.0001$, Fishers exact test).

Furthermore, at the single slide level the reporting pathologists diagnosed 928 NCBS as malignant. Following expert review, the true discrepancies resulted in a pathologists' false positive rate of 2.3% (21/928). By comparison, the prostate AI solution recorded high cancer probability

Table 2 Summary of pathologists' diagnoses and Ibex prostate AI solution results on needle core biopsy sample level

Pathologists' diagnosis	AI diagnosis		
	Low probability	Intermediate probability	High probability
Benign	1712	73	68
Malignant	22	17	889

Table 3 Summary of discrepancies between pathologists' diagnoses and Ibex prostate AI solution results on needle core biopsy sample level

Pathologists' diagnosis	AI diagnosis		Expert revision	
Ca (39 NCBS)	Intermediate cancer probability (probability score>0.8 to <0.95)	17	Benign	5 (29%)
			ASAP	3 (18%)
			PIN	2 (12%)
			Cancer	7 (41%)
	Low cancer probability (probability score≤0.8)	22	Benign	14 (64%)
			ASAP	4 (18%)
			Cancer	3 (14%)
			Not available for revision	1 (4%)
Benign (141 NCBS)	High cancer probability (probability score≥0.95)	68	Cancer	26 (38%)
			ASAP	7 (10%)
			PIN	1 (1.5%)
			Atypical ductal lesion	1 (1.5%)
			Benign	29 (43%)
			Tissue contaminants	4 (6%)
	Intermediate cancer probability (probability score>0.8 to <0.95)	73	Cancer	1 (1%)
			ASAP	5 (7%)
			PIN	2 (3%)
			Benign	62 (85%)
			Not available for revision	3 (4%)

scores for 957 NCBS, with a true false positive rate of 3.1%, which is not significantly different compared to the pathologists' ($p=0.2599$, Fisher's exact test).

Analysis of Gleason grade was possible only at the case level. Prostate AI solution identified Gleason pattern 4 areas in 2 of 29 cases originally reported a Gleason 3 + 3 (7%). Both alerts were upheld following expert review. In both cases, only a small proportion of Gleason pattern 4 cases were present (< 10%), which has no influence on differential therapeutic considerations [23]. In addition to the simple malignancy check (benign vs. malignant), this additional "Gleason higher than 3 + 3" alert represents a further, clinically highly relevant safety feature. The risk of undergrading high grade carcinomas by the primary investigator is considered to be very low [8]. After an expert review of a series of 1483 prostate biopsies, including 351 cases with an initial Gleason core of 3 + 3, Taylor et al. reported no cases of upgrading to Gleason score > 4 + 3 [8]. However, around 18% of cases were upgraded to Gleason score 3 + 4. In the meantime, the therapeutic options for these patients have become increasingly differentiated [2]. Depending on other clinical parameters, the range extends from active surveillance strategies to local, organ-preserving therapies and up to non-organ-preserving therapies. The triggering of the "Gleason higher than 3 + 3" alert induces a particularly critical second assessment of these specimens, which must then at the latest also be the reason for a statement regarding the presence of particularly therapy-relevant Gleason pattern 4 types, such as cribriform, ductal or intraductal patterns. In the two cases in our own study with AI-induced upgrading, this did not lead to a higher clinical risk classification with regard to grading.

In summary, applied in a second read setting a commercially available prostate AI solution has led to the identification of a small number of prostate cancer cases not detected by the reporting pathologists and, in two cases, to an upgrading of the Gleason score from 3 + 3 to 3 + 4. The prostate AI solution used in this study demonstrated an AUC of 0.996 for distinguishing between benign and cancer. Correspondingly sensitivity, specificity, as well as PPV and NPV were also very high. The prostate AI solution demonstrated diagnostic performance that closely matched, and in some instances even surpassed, that of experienced histopathologists; notably, it achieved a false negative rate of just 0.2%, which was significantly lower than that of the reporting pathologists, highlighting its potential to enhance diagnostic accuracy in prostate cancer detection. These observations are fully in line with previous studies analysing this as well as other prostate AI tools [11, 13–15, 17–19, 22, 24]. Thus, using this AI solution, increases the diagnostic safety of pathologists. Furthermore, our results suggest that the AI solution may reduce the need for internal review of newly diagnosed prostate cancer cases, potentially streamlining pathology workflows and improving efficiency. It has to be pointed out, however, that this is a single centre study. If these results are transferable to other institutions remains to be examined.

As a result of this study, we have integrated the prostate AI solution in our daily workflow as in a first read setting, i.e., the reporting pathologists are initially aware of AI results prior to analysing the cases. We would, however, advocate having both settings available in departments involved in the training of junior pathologists. We would see a risk that using the first read setting may affect the proficiency of junior pathologists. However, the impact of using the prostate AI solution in first or second read setting on the training of junior pathologists has yet to be examined.

In conclusion, our findings support the growing evidence that AI-based diagnostic solutions can play a pivotal role in alleviating the increasing workload faced by pathologists

due to the rising global incidence of prostate cancer. By enhancing diagnostic accuracy and efficiency, particularly when integrated into primary diagnostic workflows, such tools have the potential to sustainably support pathology services in meeting current and future healthcare demands.

Acknowledgements We would like to thank Ira Krasnitsky, MA and Dana Mevorach, MSc, Ibex Medical Analytics for their assistance with data retrieval and analyses.

Author contribution **GN**, **JK**: conception and design of the work, acquisition, analysis and interpretation of data, drafting the MS and revising it critically for important intellectual content and scientific integrity. **GS**, **MF**, **AO**, **JF**, **MV**: acquisition, analysis and interpretation of data, drafting the MS and revising it critically for important intellectual content and scientific integrity.All authors have read and approved the final manuscript.

Data availability The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Declaration

Samples were used in accordance with national ethical principles for the retrospective use of tissue samples. Specifically, all tissue samples were obtained for diagnostic purposes and were analysed in an anonymised manner.

Conflict of interest The authors declare no competing financial or non-financial interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Berenguer CV, Pereira F, Câmara JS, Pereira JAM (2023) Underlying features of prostate cancer—statistics, risk factors, and emerging methods for its diagnosis. Curr Oncol 30:2300–2321. https://doi.org/10.3390/curroncol30020178
2. James ND, Tannock I, N'Dow J et al (2024) The Lancet Commission on prostate cancer: planning for the surge in cases. Lancet 403:1683–1722. https://doi.org/10.1016/S0140-6736(24)00651-2
3. Mottet N, Bellmunt J, Bolla M et al (2017) EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 71:618–629. https://doi.org/10.1016/j.eururo.2016.08.003
4. Epstein JI (2010) An update of the Gleason grading system. J Urol. https://doi.org/10.1016/j.juro.2009.10.046
5. Belisario YG, van Leenders GJLH (2025) Current grading of prostate cancer. Pathologie (Heidelb) 46:10–15. https://doi.org/10.1007/s00292-025-01450-w
6. Epstein JI, Amin MB, Fine SW et al (2020) The 2019 Genitourinary Pathology Society (GUPS) white paper on contemporary grading of prostate cancer. Arch Pathol Lab Med 145:461–493. https://doi.org/10.5858/arpa.2020-0015-RA
7. Melia J, Moseley R, Ball RY et al (2006) A UK-based investigation of inter- and intra-observer reproducibility of Gleason grading of prostatic biopsies. Histopathology 48:644–654. https://doi.org/10.1111/j.1365-2559.2006.02393.x
8. Taylor C, Puzyrenko A, Iczkowski KA (2022) Trends in disagreement with outside genitourinary pathology diagnoses at an academic center. Pathol Res Pract 236:153997. https://doi.org/10.1016/j.prp.2022.153997
9. Walsh E, Orsi NM (2024) The current troubled state of the global pathology workforce: a concise review. Diagn Pathol 19:163. https://doi.org/10.1186/s13000-024-01590-2
10. Jurmeister P, Müller K-R, Klauschen F (2022) Artificial intelligence: a solution for the lack of pathologists? Pathologe 43:218–221. https://doi.org/10.1007/s00292-022-01071-7
11. Pantanowitz L, Quiroga-Garza GM, Bien L et al (2020) An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. Lancet Digit Health 2:e407–e416. https://doi.org/10.1016/S2589-7500(20)30159-X
12. Fatima G, Alhmadi H, Ali Mahdi A et al (2024) Transforming diagnostics: a comprehensive review of advances in digital pathology. Cureus 16:e71890. https://doi.org/10.7759/cureus.71890
13. van Diest PJ, Flach RN, van Dooijeweert C et al (2024) Pros and cons of artificial intelligence implementation in diagnostic pathology. Histopathology 84:924–934. https://doi.org/10.1111/his.15153
14. Lami K, Yoon H-S, Parwani AV et al (2024) Validation of prostate and breast cancer detection artificial intelligence algorithms for accurate histopathological diagnosis and grading: a retrospective study with a Japanese cohort. Pathology 56:633–642. https://doi.org/10.1016/j.pathol.2024.02.009
15. Eloy C, Marques A, Pinto J et al (2023) Artificial intelligence–assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Arch 482:595–604. https://doi.org/10.1007/s00428-023-03518-5
16. Arvaniti E, Fricker KS, Moret M et al (2018) Automated Gleason grading of prostate cancer tissue microarrays via deep learning. Sci Rep 8:12054. https://doi.org/10.1038/s41598-018-30535-1
17. Egevad L, Swanberg D, Delahunt B et al (2020) Identification of areas of grading difficulties in prostate cancer and comparison with artificial intelligence assisted grading. Virchows Arch 477:777–786. https://doi.org/10.1007/s00428-020-02858-w
18. Liu S, Jamal M, Smotherman C et al (2025) The potential use of artificial intelligence in reviewing atypical small acinar proliferation in prostate core biopsy. Virchows Arch. https://doi.org/10.1007/s00428-025-04136-z
19. Bulten W, Kartasalo K, Chen P-HC et al (2022) Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat Med 28:154–163. https://doi.org/10.1038/s41591-021-01620-2
20. Stark JR, Perner S, Stampfer MJ et al (2009) Gleason score and lethal prostate cancer: does 3+4=4+3? J Clin Oncol 27:3459–3464. https://doi.org/10.1200/JCO.2008.20.4669
21. Rienda I, Vale J, Pinto J et al (2025) Using artificial intelligence to prioritize pathology samples: report of a test drive. Virchows Arch 487:203–208. https://doi.org/10.1007/s00428-024-03988-1

22. Marletta S, Eccher A, Martelli FM et al (2024) Artificial intelligence-based algorithms for the diagnosis of prostate cancer: a systematic review. Am J Clin Pathol 161:526–534. https://doi.org/10.1093/ajcp/aqad182
23. Morlacco A, Cheville JC, Rangel LJ et al (2017) Adverse disease features in Gleason score 3+4 favorable intermediate-risk prostate cancer: implications for active surveillance. Eur Urol 72:442–447. https://doi.org/10.1016/j.eururo.2016.08.043
24. Kartasalo K, Ström P, Ruusuvuori P et al (2022) Detection of perineural invasion in prostate needle biopsies with deep neural networks. Virchows Arch 481:73–82. https://doi.org/10.1007/s00428-022-03326-3

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