Journal of Pathology Informatics 15 (2024) 100378

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Validation and three years of clinical experience in using an artificial intelligence algorithm as a second read system for prostate cancer diagnosis—real-world experience

Juan Carlos Santa-Rosario *, Erik A. Gustafson, Dario E. Sanabria Bellassai, Phillip E. Gustafson, Mariano de Socarraz

CorePlus Servicios Clínicos y Patológicos; Plazoleta la Cerámica, Suite 2-6 Ave. Sánchez Vilella, Esq, PR-190, Carolina, PR 00983, USA

ARTICLE INFO

Keywords:
Prostate cancer
Artificial intelligence
Gleason grading
Digital pathology
AI impact™

ABSTRACT

Background: Prostate cancer ranks as the most frequently diagnosed cancer in men in the USA, with significant mortality rates. Early detection is pivotal for optimal patient outcomes, providing increased treatment options and potentially less invasive interventions. There remain significant challenges in prostate cancer histopathology, including the potential for missed diagnoses due to pathologist variability and subjective interpretations.

Methods: To address these challenges, this study investigates the ability of artificial intelligence (AI) to enhance diagnostic accuracy. The Galen™ Prostate AI algorithm was validated on a cohort of Puerto Rican men to demonstrate its efficacy in cancer detection and Gleason grading. Subsequently, the AI algorithm was integrated into routine clinical practice during a 3-year period at a CLIA certified precision pathology laboratory.

Results: The Galen™ Prostate AI algorithm showed a 96.7% (95% CI 95.6–97.8) specificity and a 96.6% (95% CI 93.3–98.8) sensitivity for prostate cancer detection and 82.1% specificity (95% CI 73.9–88.5) and 81.1% sensitivity (95% CI 73.7–87.2) for distinction of Gleason Grade Group 1 from Grade Group 2+. The subsequent AI integration into routine clinical use examined prostate cancer diagnoses on >122,000 slides and 9200 cases over 3 years and had an overall AI Impact ™ factor of 1.8%.

Conclusions: The potential of AI to be a powerful, reliable, and effective diagnostic tool for pathologists is highlighted, while the AI Impact™ in a real-world setting demonstrates the ability of AI to standardize prostate cancer diagnosis at a high level of performance across pathologists.

Background

Early detection of prostate cancer holds paramount importance due to its potential impact on patient outcomes and overall well-being and offers a range of benefits, including increased treatment options and the possibility of less invasive interventions. With early detection, localized treatments such as surgery or radiation therapy become viable, aiming to prevent the progression of the disease to advanced stages. The long-term survival rates are generally more favorable for individuals with prostate cancer detected early, highlighting the significance of timely intervention.[1]

Traditionally, a prostate cancer diagnosis is performed through histopathological analysis using light microscopy. However, there is potential variability among pathologists' expertise in interpreting benign vs cancer and Gleason growth patterns. Due to the subjective nature of the pathologist evaluation, previous studies have reported low interobserver reproducibility in Gleason grading among urologic[2] and general pathologists.[3] This can lead to under- and over-grading of prostate cancer which can potentially impact patient care.[4]

Artificial intelligence (AI) has substantially advanced diagnostic accuracy in prostate biopsy histopathology by revolutionizing the analysis of pathology images. AI algorithms excel in processing large datasets with unparalleled speed and precision, offering a comprehensive examination of biopsy samples. These algorithms are particularly adept at recognizing intricate patterns including between benign and cancer as well as different Gleason grades, providing pathologists with invaluable support for more accurate and consistent diagnoses.[5–7]

In a previous study, Pantanowitz et al. developed, validated, and deployed an AI-based algorithm with an area under the ROC (receiver operating characteristic) curve (AUC) of 0.997 (95% CI: 0.995–0.998) and of 0.991 (0.979–1.00) for prostate cancer detection in internal and external validation sets, respectively. Additionally, this algorithm was able to distinguish between low grade (Gleason score 6 or ASAP) and high grade (Gleason score 7–10) tumors (0.941; 95% CI: 0.905–0.977).[8] This algorithm is the basis for the Galen™ Prostate AI-based solution from Ibex Medical Analytics.

* Corresponding author.
E-mail address: juan.santa@corepluspr.com (J.C. Santa-Rosario).

http://dx.doi.org/10.1016/j.jpi.2024.100378
Received 22 March 2024; Received in revised form 23 April 2024; Accepted 23 April 2024
Available online 30 April 2024

Although the use of AI algorithms is extremely beneficial for pathologists and patients, they are susceptible to bias during their development and validation.[9,10] This can lead to poor performance of the algorithm when employed on cohorts that are different from the ones used for its development. Therefore, independent validation of the algorithms for prostate cancer grading should be performed across different patient populations, pathology labs, digital pathology scanner providers, and reference standards derived from intercontinental panels of uropathologists.[7]

In this study, the Galen™ Prostate AI algorithm was validated for cancer detection using prostate core needle biopsies (PCNBs) from Puerto Rican men at CorePlus Servicios Clínicos y Patológicos (CorePlus), an independent laboratory in Puerto Rico. This AI algorithm was then implemented into routine clinical use. Results show a similar high diagnostic accuracy for cancer in this Puerto Rican population as in the original populations vs benign with an AUC of 0.994 (95% CI: 0.988–0.997). There is an additional demonstrated utility for Gleason grading distinguishing between GG1 and GG2 + with an AUC of 0.901 (95% CI 0.858–0.934). Lastly, the algorithm was implemented as a quality control second read system for 3 years in our pathology laboratory. Here, we introduce the term "AI Impact™" to quantify the effect AI had on the diagnostic routine, including between pathologists. We show an overall AI Impact™ of 1.8% for the entire 3-year period.

The validation of the Galen™ Prostate AI as a second read among a diverse cohort of Puerto Rican men in this study is of significant interest. According to the American Cancer Society, in 2023, approximately 288,300 new prostate cancer cases were expected to be diagnosed in the USA. Prostate cancer will continue to be the most diagnosed cancer in men, accounting for 14% of all new cancer cases diagnosed.[11] In terms of mortality, around 34,700 prostate cancer-related deaths are estimated to occur in the USA. This makes prostate cancer the second leading cause of cancer-related mortality in men in the USA; however, in Puerto Rican men, it is the first cause of cancer-related death.[12]

The study by Chinea et al. showed that non-Hispanic Blacks (NHB) have the highest prostate cancer-specific mortality (PCSM) in the USA, followed by Hispanics and non-Hispanic Whites (NHW). However, when the group of Hispanics was further divided into subgroups (including Mexicans, Cubans, Puerto Ricans, South or Central Americans, and Dominicans), Puerto Rican men had a significantly higher PCSM than even NHB.[13] Due to the high prostate cancer specific mortality in Puerto Rican men, an improvement in prostate cancer screening procedures for this population is needed.

Methods

Sample selection

For the validation study, 101 formalin-fixed paraffin embedded PCNB cases randomly selected from 2020, were de-identified and the 1279 associated slides were fully digitized as whole-slide images (WSIs). Benign and malignant tissue from PCNB was stained using the traditional H&E staining method. Images were obtained by scanning using the P-250 PANNORAMIC digital slide scanners (3DHISTECH, Ltd.; Budapest, Hungary) at 40× magnification with a resolution of 0.24–0.25 μm/pixel. Each slide has one prostate core divided into three levels (parts) requiring the algorithm to run on each level per slide.

Galen™ prostate algorithm

The prostate algorithm, Galen™ Prostate AI, was obtained from Ibex Medical Analytics in Israel. This algorithm was based on a multilayered convolutional neural network as described by Pantanowitz et al.[8] The algorithm was tested and validated internally by Ibex prior to the study.

Ground truth

Ground truth was established from the original diagnosis of digitized slides by four US board-certified pathologists with a combined 60 years' experience in general pathology. The diagnosis was made in a manner consistent with the International Society of Urological Pathology and College of American Pathologists guidelines[14] with no time constraint.

Algorithm testing

Algorithm accuracy to distinguish benign and malignant tissue was evaluated using—AI-generated alerts compared to an independent pathologist reviewer, using the original cutoff setting for the algorithm. The pathologist evaluation was considered the ground-truth ascertainment set. Alerts were raised when encountering discrepancies between the automated analysis and the pathologist's diagnosis, prompting a second pathologist review. Two types of alerts can be raised: (a) slides from benign cases that received a high cancer score; and (b) slides from Grade Group 1 cancer cases that received Grade Group 2 and above score. However, in this study, the focus was to discriminate between cancer and benign tissues.

Statistical analysis

The performance of the Galen™ Prostate AI algorithm as a classifier test for cancer status was assessed through the AUC of a ROC curve. Also, a contingency table was built to visualize the algorithm performance and determine accuracy, clinical sensitivity and specificity, and negative- and positive-predictive values.

Algorithm applied in routine clinical use

The Galen™ Prostate AI algorithm has been implemented in clinical use at CorePlus from 2020 to 2023 as a second-read system. In this capacity, the algorithm reviews all slides, functioning as a 100% quality control tool. Slides were digitized by the P1000 digital slide scanners (3DHISTECH, Ltd.; Budapest, Hungary) at 40× magnification with a resolution of 0.24–0.25 μm/pixel then analyzed by the Galen™ Prostate AI algorithm, followed by manual review by a pathologist in a routine workflow. Alerts raised by Galen™ Prostate AI are reviewed by the assigned pathologist with occasional intradepartmental consultation and resolved before case sign-out.

Results

The patient characteristics of the validation cohort are shown in Table 1 and are representative of the Hispanic population undergoing prostate cancer biopsies in Puerto Rico for age, PSA values and Gleason score in which all grade groups (GGs) are represented.[15]

The study compared the cancer diagnosis aided by the Galen™ Prostate AI system to that of CorePlus pathologist established ground truth. To this end, the same cut-off was used to raise cancer alerts as was originally validated for the AI algorithm. This cut-off is somewhat arbitrary, in the sense that it is configured to strike a balance between the number of alerts raised by the system (the specificity) and the probability of catching a missed cancer (the sensitivity). A lower cut-off will result in more alerts (the vast majority of which will be false positives), but a higher chance of detecting cancer that had been misdiagnosed.

A total of 1279 slides were analyzed in the validation. Out of 1022 slides that were diagnosed as benign by CorePlus in this validation set, 986 slides (96.5%) received a cancer score below the cut-off, i.e., they did not raise an alert by Galen™ Prostate AI. Of the 230 cancer slides, 221 (96.1%) passed the cut-off, i.e., a cancer alert was raised by Galen™ Prostate AI. The AUC for cancer detection in the validation set was 0.994 (95% CI 0.988–0.997), with a 96.7% (95% CI 95.6–97.8) specificity and a 96.6% (95% CI 93.3–98.8) sensitivity. Importantly, the Galen™ Prostate AI had a NPV of 99.2% (95% CI 98.4–99.6) indicating the high certainty of negative calls by the algorithm (Fig. 1A and B).

During the validation, ground-truth ascertainment for cancer vs benign diagnosis resulted in algorithm alerts for 17 slides across 12 cases. The alerts are on the slide level and are also aggregated on the case level. If a slide is called cancer by the algorithm, but benign by the ground truth, an

Table 1
Distribution of patient age and diagnosis in the study.

Number of cases	Number of H&E slides	Age (years) distribution (%)					
		30–39	40–49	50–59	60–69	70–79	80–89
101	1279	1	4	16	43	30	7
Number of cases		PSA score–cases					
		N/A	0–4	5–9	10–19		> 20
101		4	38	41	9		9
Diagnosis	Benign	ASAP	Cancer				
Slides	81.6%	2.1%	18.3%				
			GG-1	GG-2	GG-3	GG-4	GG-5
			9.5%	4.7%	2.3%	0.3%	1.3%
Cases	47.3%	2.1%	52.7%				
			GG-1	GG-2	GG-3	GG-4	GG-5
			19.3%	22.3%	3.1%	4.3%	1.2%

GG = Gleason Grade Group.

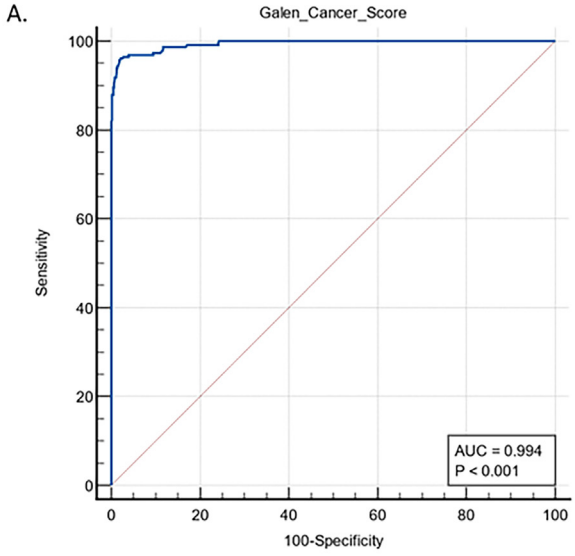
alert will only occur if the case diagnosis was benign by the ground truth. Following pathologist review, all 17 slide alerts were resolved and shown to be triggered by glands (normal, atrophic, and seminal vesicle) as well as inflammation (Table 2), features that may mimic malignant glands.

Likewise, as to the cancer diagnosis, the Gleason grading of the Galen™ Prostate AI system was compared to that of the CorePlus pathologists. For this grading analysis, two sets of slides were used which included low grade (GG1 and ASAP) and high grade (GG2+). ASAP was included in the low grade as it is difficult to distinguish it from GG1 cancer and results are often subjective. Using the same cut-off as validated in the original algorithm, out of 146 slides that were diagnosed by CorePlus as low grade, 113 slides (77.4%) received a score below the GG2+ cut-off, i.e., also scored as low grade by Galen™ Prostate AI. Out of 111 high-grade slides, 89 slides (80.2%) received a score above the cut-off and scored by Galen™ Prostate AI as high-grade as well. The AUC for Gleason grading in the validation set was 0.901 (95% CI 0.858–0.934) with an 82.1% specificity (95% CI 73.9–88.5) and a 81.1% sensitivity (95% CI 73.7–87.2) (Fig. 2A and B).

During the validation, ground-truth ascertainment for GG1 vs GG2 and above resulted in algorithm alerts for 18 GG1 slides across 8 cases. Following pathologist review, three slides from two cases were confirmed as GG2 (Table 3).

Following validation, the Galen™ Prostate AI algorithm was incorporated into the pathology lab as a second-read system for prostate cancer. The term "second read" indicates that following a primary review by the pathologist, the algorithm is applied to all the WSIs of the case. Cancer alerts and Gleason alerts are generated at the slide level and aggregated at the case level. When the Galen AI finds a suspicious focus, it alerts the pathologist to the slide containing the area of interest. Fig. 3 shows the workflow for the second-read implementation of this algorithm in our laboratory. The way alerts are visualized is by presenting heatmaps over areas of interest. As an example, shown in Fig. 4 is an algorithm generated heatmap indicating high areas of probability of cancer in a prostate biopsy. Areas of red indicate a high likelihood of cancer.

Table 4 shows the overall results for the Galen™ Prostate AI algorithm used for 3 years as a second-read system at the case level. Between June 2020 and May 2023, 9267 cases (122,441 slides) were processed by the second-read system at CorePlus. Cancer alerts were raised for 5.7% (3239/57119) of the slides or 39.5% (1733/4385) of the cases diagnosed by the pathologist as benign. After review by the pathologist, 0.3% (155/57119) of the benign slides and 2.9% (128/4385) of benign cases required a change of diagnosis. A cancer AI Impact™ score of 1.5% was computed from the ratio of total revised benign cases to the total number of cases ×



B.

Galen AI	Ground Truth: Cancer	Ground Truth: Benign	
Cancer	224	32	256
Benign	8	987	995
	232	1019	1251

		95% CI
Sensitivity	96.6%	93.3% to 98.5%
Specificity	96.7%	95.6% to 97.8%
Positive Predictive Value	87.5%	83.3% to 90.8%
Negative Predictive Value	99.2%	98.4% to 99.6%
Accuracy	96.8%	95.7% to 97.7%

Fig. 1. Sensitivity and specificity analysis of the Galen™ Prostate AI algorithm for cancer detection in samples from Puerto Rican men. (A) ROC curve analysis to evaluate the accuracy of the algorithm to identify cancer vs. benign tissues. The blue curve represents the data transformation and balance between sensitivity and specificity. (B) Contingency table for comparison of Galen™ Prostate AI prostate cancer diagnosis to ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Cancer alerts in validation study—slides from benign cases with the highest cancer scores.

Slide	Cancer score	Review comments
0176 J1	98.3	Atrophy
0176 N1	96.2	Atrophy
0405 G1	95.4	Inflammation
0406 C1	98.3	Atrophy
0407 I1	96.6	Adenosis
0576 F1	97.4	Inflammation
0576 L1	99.0	Ganglion cells
0663 A1	98.3	Seminal vesicle glands
0663 C1	97.8	Seminal vesicle glands
0900 D1	95.2	Inflammation
0903 B1	95.2	Normal glands
1170 C1	95.2	Atrophy
1449 L1	96.4	Normal glands
1645 H1	95.2	Atrophy
1645 K1	98.3	Atrophy
1645 M1	97.4	Atrophy
1858 I1	98.6	Seminal vesicle glands

100. In addition, Gleason alerts were raised during the validation on 2.1% (563/26189) of the slides or 19.2% (380/1984) of the cases diagnosed by the pathologist as GG1. After review by the pathologist, only 0.11% (28/26189) of the GG1 slides and 1.21% (24/1984) of the GG1 cases were altered due to the Gleason alerts (Table 5). A Gleason AI Impact™ score of 0.26% was computed from the ratio of total revised ASAP/GG1 cases to the total number of cases × 100. Therefore, the total overall AI Impact™ from revised benign cases and revised ASAP/GG1 cases was 1.8%.

Finally, an inter-rater reliability analysis between pathologists was performed over the same 3-year period for prostate cancer diagnosis. In Fig. 5, the AI Impact™ per individual pathologist is shown. Individual interpretations can vary which is not unexpected and, in this example, varied from an AI Impact™ of >3.5% for pathologist A and an AI Impact™ of <1% for pathologist B. The use of the Galen™ Prostate AI algorithm in a second-read implementation had the effect of providing alerts for review such that the false-negative rate of the group was reduced by the amount indicated by the AI Impact™.

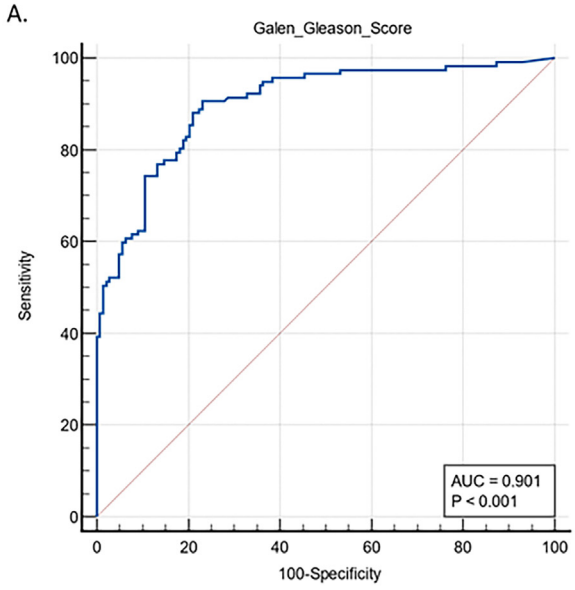
Table 3
Gleason alerts in validation study—slides from GG1 cancer cases with the highest scores for GG2+.

Slide	Gleason score	Resolution	Comments
0177 A1	98.4	Retain Dx	
0843 F1	99.7	Retain Dx	
0843 G1	98.1	Retain Dx	
0898 K1	99.4	Retain Dx	
0912 I1	99.4	Confirmed	Grade Group 2 (3 + 4) with <5% Gleason pattern 4
1861 B1	98.1	Retain Dx	
1983 L1	100	Retain Dx	
1994 A1	99.1	Retain Dx	
1994 H1	99.1	Retain Dx	
1994 L1	99.4	Retain Dx	
1994 M1	100	Retain Dx	
1783 E1	98.4	Confirmed	Grade Group 2 (3 + 4) with <5% Gleason pattern 4
1783 M1	99.1	Confirmed	Grade Group 2 (3 + 4) with <5% Gleason pattern 4
1783 F1	99.4	Retain Dx	
1783 G1	98.4	Retain Dx	
1783 H1	99.1	Retain Dx	
1783 K1	99.4	Retain Dx	
1783 L1	98.1	Retain Dx	

Discussion

This study describes the validation of the Galen™ Prostate AI algorithm on a diverse Puerto Rican cohort and represents, to the best of our knowledge, the first clinical implementation in the Americas using AI as a prostate cancer diagnostic tool routinely on all PCNBs samples. Though there have been many previous studies demonstrating the utility of various AI algorithms for prostate cancer detection, we are not familiar with any published reports of implementation in a real-time clinical application.

Perincheri et al.,[16] in one of the larger previous validation studies assessing AI in prostate cancer, evaluated the performance of an FDA approved algorithm[17] to analyze 1876 WSIs of PCNBs across 118 cases using the original, non-AI subspecialized genitourinary pathologist



B.

	Ground Truth		
Galen AI	GG1/ASAP	GG2+	
GG1	116	21	137
GG2	27	96	123
	143	117	260

		95% CI
Sensitivity	81.1%	73.7% to 87.2%
Specificity	82.1%	73.9% to 88.5%
Positive Predictive Value	84.7%	78.8% to 89.1%
Negative Predictive Value	78.0%	71.5% to 83.5%
Accuracy	81.5%	76.3% to 86.1%

Fig. 2. Sensitivity and specificity analysis of the Galen™ Prostate AI algorithm for Gleason grade in samples from Puerto Rican men. (A) ROC curve analysis to evaluate the accuracy of the algorithm to identify Gleason Grade Group 1 vs. Gleason Grade Group 2 and above. The blue curve represents the data transformation and balance between sensitivity and specificity. (B) Contingency table for comparison of Galen™ Prostate AI prostate Gleason grade to ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

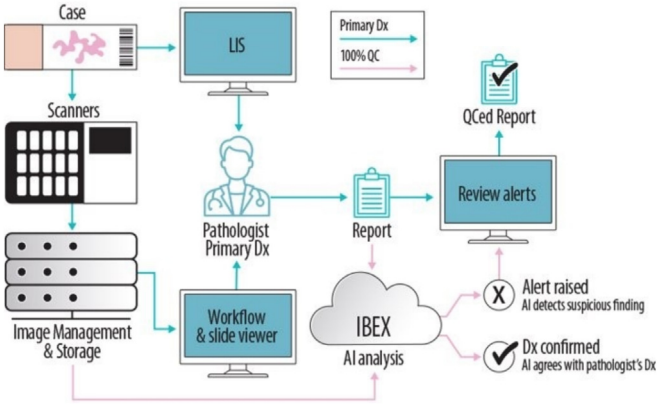


Fig. 3. Implementation of a second-read workflow for prostate cancer detection.

diagnosis as ground truth. They report the AI algorithm had a sensitivity and PPV of >97% and a specificity and NPV of over 99%. Likewise, in recent, but somewhat smaller validation studies the same FDA approved algorithm was validated by comparing multiple pathologists' diagnoses to an established ground truth. Good sensitivity and specificity at the WSI level were reported by da Silva et al.[18] while at the case level, sensitivity and NPV were 100%, but a lower specificity of 78%. The authors contend this performance capability reduces the probability of a false negative (FN) while allowing a tolerable number of patients to be flagged as suspicious who are benign. Raciti et al.[19] report an increased sensitivity of the general pathologists when using the AI algorithm, especially in smaller, lower grade tumors. A very recent report by Eloy et al.[20] demonstrates equivalent performance to the pathologist ground truth using the AI algorithm, but also additional benefits of a statistically significant reduction in IHC orders, requirement for second opinions and reporting time. Lastly, in a separate independently developed algorithm, Kott et al.[21] report the ability of their algorithm to distinguish prostate cancer from benign tissue with a sensitivity and specificity of 93% and 90%, respectively. In Gleason grading, the same algorithm could distinguish benign from Gleason 3, Gleason 4, and Gleason 5 with an 83% sensitivity and a 94% specificity.

In the current study with a cohort of Puerto Rican men, the Galen™ Prostate AI had a sensitivity and specificity of 96.6% and 96.7%, respectively. In validation, Galen AI did not miss any cancer at the case level out of 1279 WSI. At the slide level, Galen™ called nine slides benign that were cancer. Seven of the nine were GG1, the other two were GG2 and GG4. Like the original Pantanowitz et al. report[8] and the report from Kott et al.,[21] the Galen AI algorithm performance on Gleason grading was somewhat lower than benign vs cancer performance with a sensitivity of 77.4% and specificity of 80.2% compared to the pathologist grading. Perhaps this finding is not surprising, since it is being compared to a pathologist grading that this Gleason variability is within 30% of human interobserver variability.[2] Nevertheless, AI should bring more consistency to the grading.

It is clear from this validation and the literature that AI algorithms can reach pathologist level accuracy.[5–7] The Galen™ Prostate AI algorithm was implemented in a second-read capacity at CorePlus driven by high sensitivity (96.1%) and NPV (99.1%) of reporting cancer vs benign.

One of the better ways to reduce errors in pathology is to have a case review, ideally by expert pathologists.[22] However, manual QC review requires extra pathologist workload which in common practice is around 1–10% of cases.[23] An AI second-read system provides the benefit of having a 100% QC of all cases, thereby alerting to any potential error on each case reviewed.

Another benefit of an AI second-read system is to assist pathologists with maintaining accuracy in diagnosis. For example, the true rate of FN results among pathologists is difficult to know, with literature reports estimating between 1 and 10%.[24,25] The subjective nature of the work is one possible reason for this. Supporting this notion, one study suggests that errors correlate more to the individual pathologist and less to their years of experience or workload.[22] AI can compensate for differences between pathologists, bringing the FN rate to a low value consistently. This study clearly indicates that performance can vary significantly between pathologists. However, the use of the Galen™ Prostate AI provided a compensation effect such that the performance of each pathologist and the group improved. This improvement can be described by the "AI Impact™", the percentage of cases alerted and changed before signing out that otherwise would have been either FNs or lower grade cancer had the AI not been in place. In our real-world data, AI Impact™ ranged from 0.75% up to 3.5% per individual pathologist, with a normalized group effect of 1.8%.

Lastly, there are major costs associated with interpretive errors in pathology, with estimates of $21,000 and $70,000 per case at two major institutions.[23] A reduction in diagnostic error will almost certainly lead to substantial savings over time for the patient and healthcare system in addition to improvement of patient care. However, at this early stage of AI implementation into routine practice, another thing to consider is that AI might detect cancers that potentially place the patient at risk for overdiagnosis and harm due to therapeutic interventions.[26] Additional studies are needed to evaluate such implementations' impact on the overall clinical management of patients.

In summary, the benefits of utilizing a proven accurate, sensitive, and specific AI algorithm for assisting pathologists in prostate cancer diagnosis have been shown and discussed here. These include improved diagnostic accuracy resulting from a lower FN rate, normalization of diagnostic

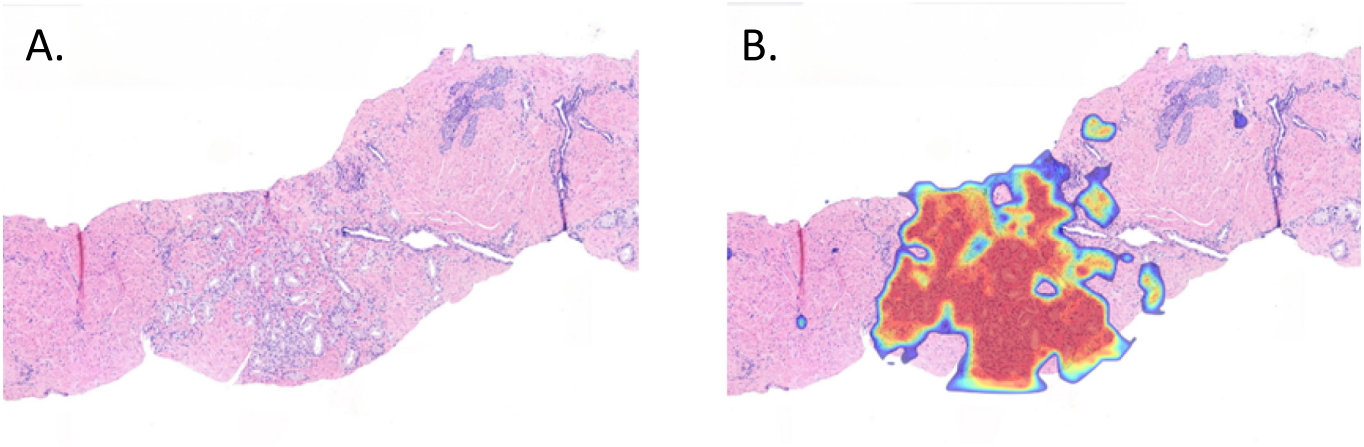


Fig. 4. Example of a second-read review by the Galen™ Prostate AI algorithm in a prostate cancer case. (A) H&E staining shows an area with cancer. (B) Review by the Galen™ Prostate AI algorithm identified with a heatmap over the same area of a high probability of cancer.

Table 4
Results of AI algorithm as second read on cancer diagnosis over a 3-year period.

		Jun'20–May'21	Jun'21–May'22	Jun'22–May'23	Total
			Num. (%)		Num. (%)
Slides	#Slides	40,609 (100)	40,510 (100)	41,322 (100)	122,441 (100)
	#Slides in Benign cases	19,724 (48.6)	18,735 (45.4)	18,660 (45.2)	57,119 (46.7)
	#Benign alerted for cancer	1261 (6.4)	920 (4.9)	1058 (5.7)	3239 (5.7)
	#Revised benign slides	61 (0.3)	57 (0.3)	37 (0.2)	155 (0.3)

		Jun'20–May'21	Jun'21–May'22	Jun'22–May'23	Total
			Num. (%)		Num. (%)
Cases	#Cases	3107 (100)	3089 (100)	3071 (100)	9267 (100)
	#Benign cases	1558 (50.1)	1428 (46.2)	1399 (45.6)	4385 (52.2)
	#Benign alerted for cancer	656 (42.1)	528 (40.0)	549 (39.2)	1733 (39.5)
	#Revised benign cases	56 (3.6)	47 (3.3)	33 (2.4)	136 (3.1)
				AI Impact™	1.5%

Table 5
Results of AI algorithm as second read on Gleason revisions over a 3-year period.

		Jun'20–May'21	Jun'21–May'22	Jun'22–May'23	Total
			Num. (%)		Num. (%)
Slides	#Slides	40,609 (100)	40,510 (100)	41,322 (100)	122,441 (100)
	#Slides in ASAP/GG1 cases	8118 (20.0)	8773 (21.7)	9298 (22.5)	26,189 (21.4)
	#ASAP/GG1 Gleason alert	162 (2.0)	161 (1.8)	240 (2.6)	563 (2.1)
	#Revised GG1 slides	11 (0.14)	9 (0.10)	8 (0.09)	28 (0.11)

		Jun'20–May'21	Jun'21–May'22	Jun'22–May'23	TOTAL
			Num. (%)		Num. (%)
Cases	#Cases	3107 (100)	3089 (100)	3071 (100)	9267 (100)
	#ASAP/GG1 cases	623(20.1)	670 (21.7)	691 (22.5)	1984 (21.4)
	#ASAP/GG1 Gleason alert	107 (17.2)	111 (16.6)	162 (23.4)	380 (19.2)
	#Revised ASAP/GG1 cases	9 (1.44)	7 (1.04)	8 (1.2)	24 (1.21)
				AI impact™	0.26%

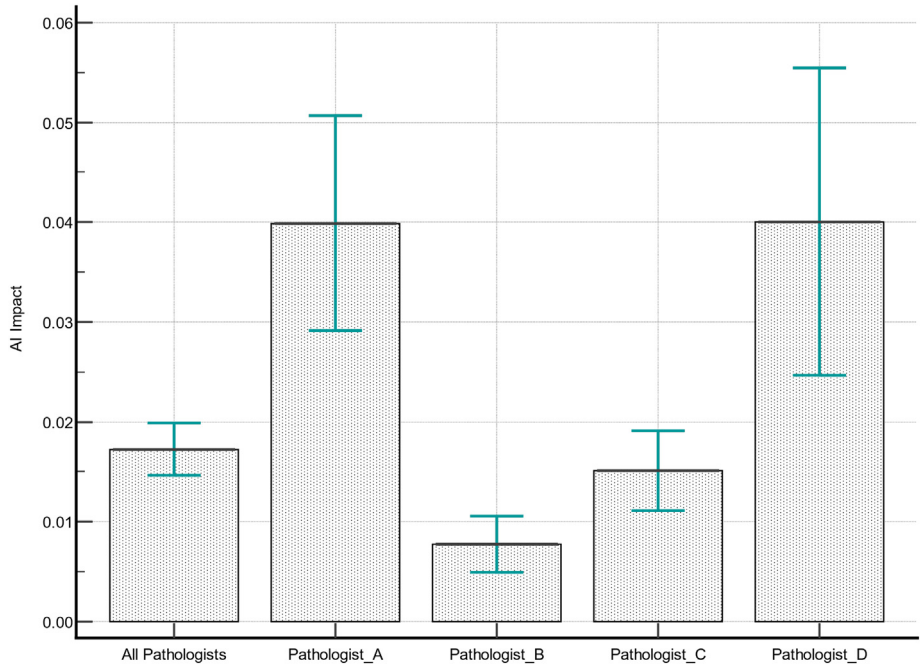


Fig. 5. AI Impact™ of the Galen™ Prostate AI algorithm for prostate cancer detection across pathologists over a 3-year period.

performance between pathologists, Gleason grade standardization, and with the second-read approach, 100% QC of cases at sign-out. Each of these has already benefited patients referred to our facility for years. At the time of this writing, we have since implemented a first-read application using the same AI algorithm described here which gives the additional benefits of reducing the pathologist workload by supporting them with expert level Gleason grading, locating clinically significant findings such as perineural invasion, and performance of mundane tasks such as tumor measurement. Lastly, AI algorithms have been developed that include not just diagnostic assistance, but predictive and prognostic information as well.[27] Such algorithms should help round out a full suite of AI capabilities that will enable precision pathology to improve diagnostic accuracy, benefit patient care and reduce healthcare costs.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Mariano de Socarraz reports a relationship with Ibex Medical Analytics that includes: equity or stocks. Juan C. Santa Rosario reports a relationship with Ibex Medical Analytics that includes: equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Brenner H, Arndt V. Long-term survival rates of patients with prostate cancer in the prostate-specific antigen screening era: population-based estimates for the year 2000 by period analysis. J Clin Oncol 2005;23(3):441–447. https://doi.org/10.1200/JCO.2005.11.148.
2. Allsbrook Jr W, Mangold K, Johnson M, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. Hum Pathol 2001;32(1):74–80.
3. Singh RV, Agashe SR, Gosavi AV, Sulhyan KR. Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists. Indian J Cancer 2011;48(4):488–495. https://doi.org/10.4103/0019-509X.92277.
4. Flach RN, Willemse PPM, Suelmann BBM, et al. Significant inter-and intralaboratory variation in Gleason grading of prostate cancer: a nationwide study of 35,258 patients in the Netherlands. Cancers (Basel) 2021;13(21). https://doi.org/10.3390/cancers13215378.
5. Kartasalo K, Bulten W, Delahunt B, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer in biopsies—current status and next steps. Eur Urol Focus 2021;7(4):687–691. https://doi.org/10.1016/j.euf.2021.07.002.
6. Morozov A, Taratkin M, Bazarkin A, et al. A systematic review and meta-analysis of artificial intelligence diagnostic accuracy in prostate cancer histology identification and grading. Prostate Cancer Prostatic Dis 2023;26:681–692.
7. Bulten W, Kartasalo K, Chen PHC, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat Med 2022;28(1):154–163. https://doi.org/10.1038/s41591-021-01620-2.
8. Pantanowitz L, Quiroga-Garza GM, Bien L, et al. Articles An Artificial Intelligence Algorithm for Prostate Cancer Diagnosis in Whole Slide Images of Core Needle Biopsies: A Blinded Clinical Validation and Deployment Study. www.thelancet.com/; 2020.
9. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019;28(3):231–237. https://doi.org/10.1136/bmjqs-2018-008370.
10. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. Commun Med 2021;1(1). https://doi.org/10.1038/s43856-021-00028-w.
11. Cancer Facts and Figures 2023. American Cancer Society. Published 2023. Accessed January 17, 2024. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html
12. Torres-Cintrón C, Alvarado-Ortiz M, Román-Ruiz Y, Ortiz-Ortiz K, Zavala-Zegarra D, Tortolero-Luna G. Cancer in Puerto Rico, 2014–2018. San Juan, PR: Puerto Rico Central Cancer Registry. Published online 2021.
13. Chinea FM, Patel VN, Kwon D, et al. Ethnic Heterogeneity and Prostate Cancer Mortality in Hispanic/ Latino Men: A Population-Based Study www.impactjournals.com/oncotarget
14. Urinary and Male Genital Tumors 5th ed. WHO Classification of Tumours Editorial Board. 2022.
15. Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: current state and prognostic implications. Diagn Pathol 2016;11(1). https://doi.org/10.1186/s13000-016-0478-2.
16. Perincheri S, Levi AW, Celli R, et al. An independent assessment of an artificial intelligence system for prostate cancer detection shows strong diagnostic accuracy. Mod Pathol 2021;34(8):1588–1595. https://doi.org/10.1038/s41379-021-00794-x.
17. Campanella G, Hanna M, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25:1301–1309.
18. da Silva LM, Pereira EM, Salles PGO, et al. Independent real-world application of a clinical-grade automated prostate cancer detection system. J Pathol 2021;254(2):147–158. https://doi.org/10.1002/path.5662.
19. Raciti P, Sue J, Ceballos R, et al. Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. Mod Pathol 2020;33(10):2058–2066. https://doi.org/10.1038/s41379-020-0551-y.
20. Eloy C, Marques A, Pinto J, et al. Artificial intelligence–assisted cancer diagnosis improves the efficiency of pathologists in prostatic biopsies. Virchows Arch 2023;482(3):595–604. https://doi.org/10.1007/s00428-023-03518-5.
21. Kott O, Linsley D, Amin A, et al. Development of a deep learning algorithm for the histopathologic diagnosis and Gleason grading of prostate cancer biopsies: a pilot study. Eur Urol Focus 2021;7(2):347–351. https://doi.org/10.1016/j.euf.2019.11.003.
22. Renshaw AA, Gould EW. Measuring errors in surgical pathology in real-life practice: defining what does and does not matter. Am J Clin Pathol 2007;127(1):144–152. https://doi.org/10.1309/5KF89P63F4F6EUHB.
23. Priebe M, Markin R. *Acta Scientific Cancer Biology Review of Anatomic Pathology and Diagnostic Radiology Quality Assurance Tools to Reduce Diagnostic Discordance in Cancer*. 2019.
24. Van Der Kwast T, Lopes C, Martikainen P, et al. Report of the pathology committee: false-positive and false-negative diagnoses of prostate cancer. BJU Int 2003;92(2):62–65.
25. Yang C, Humphrey PA. False-negative histopathologic diagnosis of prostatic adenocarcinoma. Archives of Pathology and Laboratory Medicine. College of American Pathologists; 2020. p. 326–334. https://doi.org/10.5858/arpa.2019-0456-RA.
26. Senevirathna P, Pires D, Capurro D. Data-driven overdiagnosis definitions: scoping review. J Biomed Inform 2023;147, 104506. https://doi.org/10.1016/j.jbi.2023.104506. ISSN 1532-0464.
27. Bera K, Schalper KA, Rimm DL, et al. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019;16:703–715. https://doi.org/10.1038/s41571-019-0252-y.