Using AI-powered solution in Clinical Practice for Primary Diagnosis of Prostate Biopsies

Rainer Grobholz[1,2], Mona Spoerri[2], Philipp Nahm[3], Felice Burn[3], Maciej Kwiatkowski[4], Leigh-Anne McDuffus[5], Shlomtzion Evron[5], Dana Mevorach[5], Manuela Vecsler[5]

[1]Medical Faculty, University of Zurich, Switzerland, [2]Institute of Pathology Kantonsspital Aarau, Switzerland, [3]Department of Informatics, Kantonsspital Aarau, Switzerland, [4]Department of Urology, Kantonsspital Aarau, Switzerland. [5]Ibex Medical Analytics, Tel Aviv, Israel

BACKGROUND

- Adenocarcinoma of the prostate is the most common cancer diagnosed in men with approximately one in nine men diagnosed in their lifetime[1]
- Current screening methods impose substantial laboratory workload and suffer from intra and inter-observer variability, poor reproducibility and potential misses of small cancers[2]
- AI-based diagnostic support solutions are being implemented and may improve diagnostic accuracy, reproducibility and efficiency[3]

OBJECTIVE

This study tested the use of an artificial intelligence (AI)-based solution, Galen™ Prostate[4] (Ibex Medical Analytics) for primary diagnosis of prostate adenocarcinoma using real-world data

METHODOLOGY

89 prostate cancer cases with a total of 674 H&E slides (~1700 cores) from routine diagnostics were retrieved from the archive (Table 1)

Accuracy for cancer detection and Gleason grading was assessed by comparing the AI results with the ground truth (GT) established by the pathologist

Case Diagnosis	No. Cases	No. Slides	Age, Mean ± Std (range)	PSA, Mean ± Std (range)
GG 1	18 (20%)	104 (16%)	64.4 ± 7.64 (50-77)	5.75 ± 2.98 (1.66-12.8)
GG 2	15 (17%)	53 (8%)	67.2 ± 8.26 (51-77)	12,54 ± 6.89 (4.61-30.1)
GG 3	16 (18%)	43 (6%)	71.12 ± 5.11 (62-82)	10.79 ± 9.61 (4.25-42.0)
GG 4	22 (25%)	64 (9%)	71.64 ± 9.17 (55-92)	45.25 ± 122.56 (4.6-529)
GG 5	18 (20%)	64 (9%)	74.61 ± 8.37 (59-87)	236.77 ± 364.63 (6.21-1462)
Benign	0 (0%)	343 (51%)		
Total	**89 (100%)**	**674 (100%)**		

Table 1 | Patient cohort and characteristics.
GG - Gleason Grade Group

RESULTS

Cancer Detection

The AI solution agreed with the ground truth (GT)

in 81 cases (91%) for carcinoma

in 8 cases (9%) areas were detected with a low/medium likelihood for cancer by the AI

On the slide level, 132/674 (20%) slides were marked with a medium likelihood for cancer. Based on this detection, 57/132 (43%) of these marked areas proved to be benign, in 7/132 (5%) slides small additional foci of GG1 cancer were detected, and 60/132 (46%) represented cancer (Table 2)

In 4/343 (0.9%) slides initially no tumor was reported, two were revised as cancer, 2 classified as ASAP

Interestingly, in some cases not the whole tumor area was detected by the AI

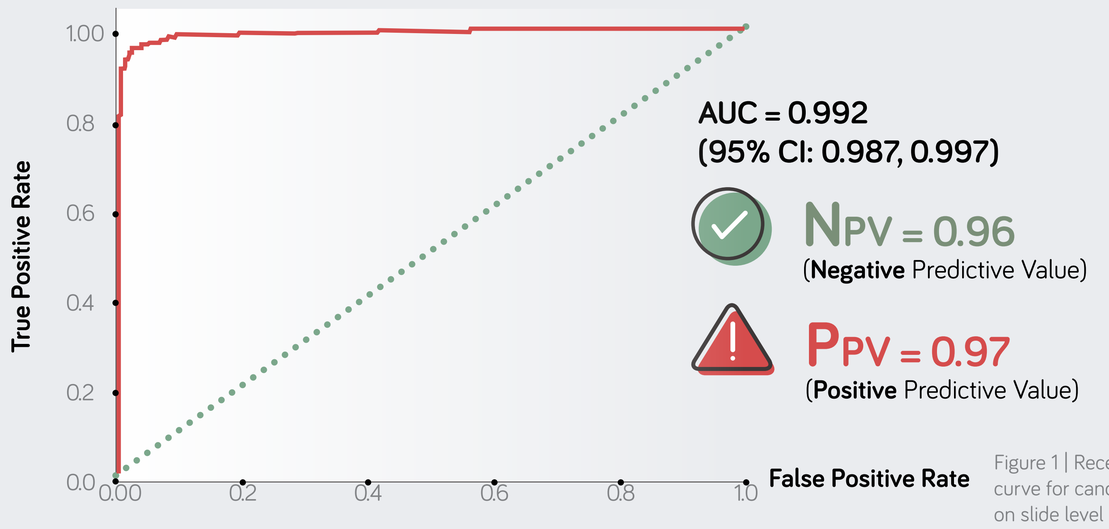

Figure 1 | Receiver operator curve for cancer detection on slide level

Gleason Grade Group Detection

Quadratic Kappa – 0.941
[95% CI: 0.917, 0.959]

ICC - 0.762
[95% CI: 0.683, 0.82]

The AI solution generated identical or +1/-1 Gleason grade groups (GG) to the GT for 88.7% and 89.9% of cases after revision (n=89)

In 8/89 cases (9%) and in 72/674 (10.7%) tumor slides no Gleason score was assigned due to a tumor recognition as "low/medium likelihood of cancer" (Table 3)

The highest concordance rates were in GG1, GG2 and GG5 (Table 3)

Discrepancies resulted due to small cancer foci, tumor quantification method (tumor length versus tumor cell fraction) or undetected tumor areas by the AI

Revisions following the AI diagnosis

Case Level The GT was revised for 12 (13.5%) cases to a higher GG following the AI output

Slide Level The GT was revised for 51 (7.6%) slides following the AI output

- 9 (1.3%) benign slides were revised to cancer
- 7 (1%) benign slides were revised to ASAP
- 32 (4.8%) slides were revised to a higher GG
- 3 (0.4%) slides were revised to a lower GG

Galen Prostate	Pathologist							
	GG1	GG2	GG3	GG4	GG5	ASAP	Benign	No. slides
GG 1	**54 (52%)**	18 (34%)		1 (2%)		1 (25%)	2 (1%)	76
GG 2	3 (3%)	**28 (53%)**	14 (33%)	1 (2%)				46
GG 3	1 (1%)	1 (2%)	**13 (30%)**	9 (14%)	2 (3%)			25
GG 4			10 (23%)	**22 (35%)**	9 (14%)		1 (<1%)	42
GG 5			3 (7%)	24 (38%)	**45 (70%)**		1 (<1%)	73
Medium likely	46 (44%)	5 (9%)	2 (5%)	5 (8%)	2 (3%)	**3 (75%)**	69 (20%)	132
No GS assigned		1 (2%)	1 (2%)	1 (2%)	6 (9%)			9
Benign							270 (79%)	**270**
Total no. slides	**104**	**53**	**43**	**63**	**64**	**4**	**343**	**674**

Table 2. Concordance of Gleason grade groups – slide level

Galen Prostate	Pathologist					
	GG1	GG2	GG3	GG4	GG5	No cases
GG 1	**12 (67%)**	2 (13%)		1 (5%)		15
GG 2		**11 (73%)**	5 (31%)			16
GG 3		1 (7%)	**5 (31%)**	5 (23%)		11
GG 4			5 (31%)	**6 (27%)**	2 (11%)	13
GG 5			1 (6%)	9 (41%)	**16 (89%)**	26
No GS assigned	6 (33%)	1 (7%)		1 (5%)		**8**
Number cases	**18**	**15**	**16**	**22**	**18**	**89**

Table 3. Concordance of Gleason grade groups – case level

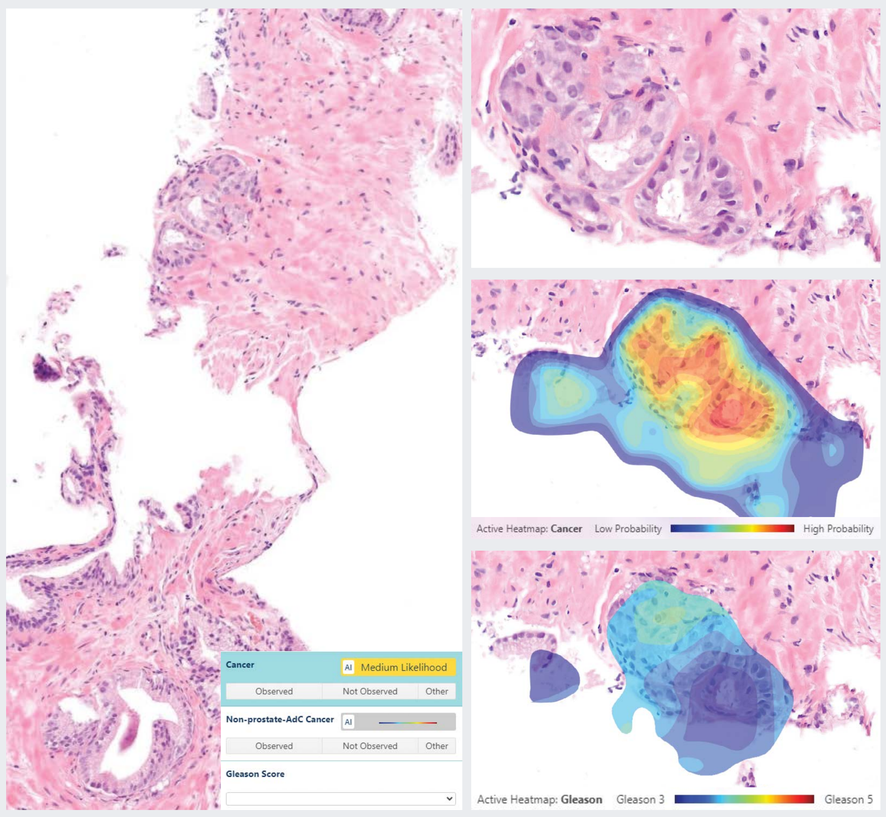

Figure 2. Small cancer focus, missed on H&E, detected by AI with medium likelihood, no Gleason score assigned

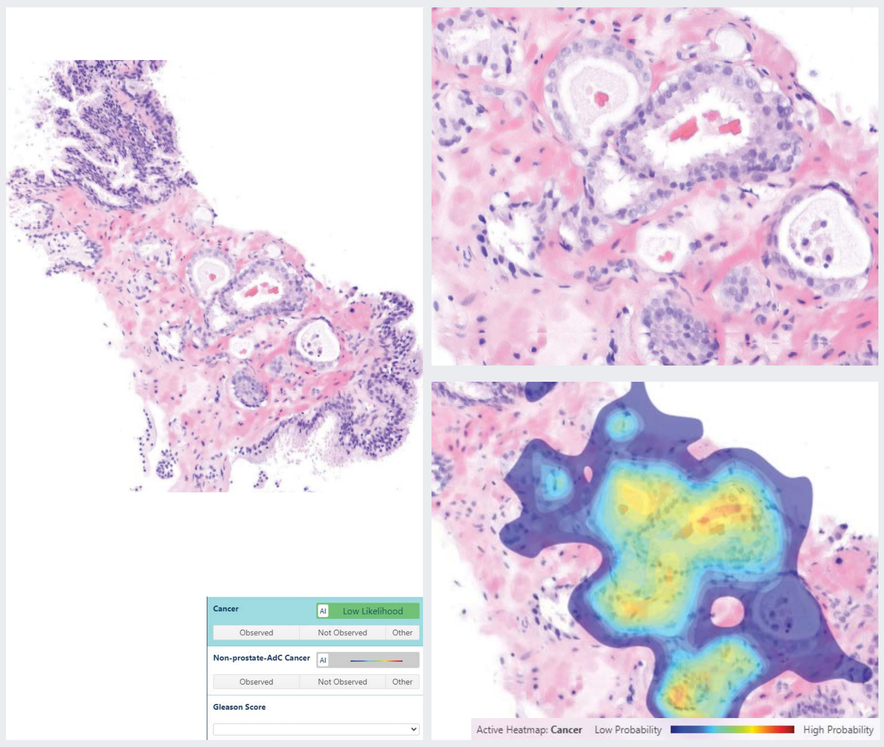

Figure 3. Small cancer focus, detected by AI with low likelihood, no Gleason score assigned

CONCLUSIONS

This study reports high levels of accuracy of an AI-solution that supports pathologists by automatically identifying adenocarcinoma and Gleason grading

The AI-algorithm demonstrated high-accuracy levels:

- The AUC for adenocarcinoma detection was very high
- Agreement levels between the AI and the Gleason grade groups established by a pathologist were high
- 6.9% of slides re-reviewed by pathologists were corrected according to the AI diagnosis
- In 8/89 (9%) cases no Gleason score was assigned by the AI-solution due to tumor recognition as "low" or "medium likelihood"

CLN-0005 Rev. 1.0.

References: 1. Van Booven et al., 2021; 2. Glodenberg et al, 3. Galen Prostate solution is CE-IVD marked; 2019; 4. Pantanowitz et al, 2020. | Disclosures: Leigh-Anne McDuffus, Shlomtzion Evron, Dana Mevorach and Manuela Vecsler are employees of Ibex Medical Analytics and have an ownership interest in stock options. All the other authors have no disclosures