Performance of an AI-based cancer diagnosis system in France's largest network of pathology institutes

Daphna Laifenfeld[1] Judith Sandbank[1,2] Chaim Linhart[1] Lilach Bien[1] Cobi Reouven[1]
Roei Harduf[1] Mahul Amin[3] Olivier Levrel[4] Stéphane Rossat[4] & Delphine Raoux[4]

1 Ibex Medical Analytics, Israel 2 Institute of Pathology, Maccabi Healthcare Services, Israel 3 University of Tennessee Health Science Center, USA 4 MédiPath, France

BACKGROUND

- Prostate cancer is the second leading cause of cancer-related deaths, compounded by complexity of diagnosis, insufficient diagnostic reproducibility, and a growing shortage in pathologists
- Ibex Medical Analytics, focusing on AI-based cancer diagnostics, has developed an algorithm that identifies various cell types and tissue structures within whole slide images of prostate core needle biopsies (PCNBs), such as cancerous glands (of Gleason patterns 3, 4 and 5), high-grade PIN, inflammation and atrophic glands
- MediPath is the largest network of pathology institutes in France, servicing 370,000 patients including 5,000 prostate biopsies annually

Objective

This study was aimed to assess the performance of Ibex's algorithm on benign cases from 4 labs within the network, and to detect and characterize missed cancer cases

ALGORITHM FOR PROSTATE BIOPSIES

- The algorithm utilizes state-of-the-art Artificial Intelligence (AI) and Machine Learning techniques and was trained on hundreds of thousands of image samples, the majority taken from hundreds of manually annotated PCNBs from Maccabi Healthcare Services Pathology Institute
- The algorithm was validated on a large set of slides from Maccabi Pathology Institute and demonstrated very high performance, as shown in the table below:

Accuracy of Prostate Cancer Detection Algorithm

AUC	SENSITIVITY	SPECIFICITY	# SLIDES
0.9957	96%	97.5%	3,466 (2,787 benign ,679 cancer)

Note: All validation data is independent of training data. AUC: area under ROC curve.

METHODS

Data

- Calibration set included 150 anonymized cases (100 benign, 50 cancer) including 1,140 H&E slides with various clinical features (e.g., Gleason grades)
- 40 slides from the Calibration set were annotated and used as part of the algorithm's training data
- Study set included 100 benign cases (801 H&E slides)

Algorithmic Analysis

- Calibration set was used to determine a threshold score optimizing for specificity and sensitivity
- Ibex prostate algorithm was run on the Study set

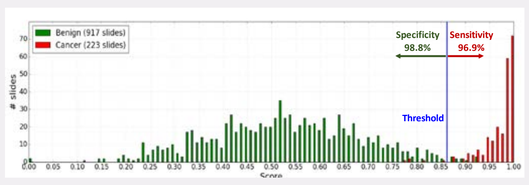
Pathologists Review

- Slides that passed the threshold were sent to two independent pathologists (M.A. and O.L.) for review

RESULTS

DISTRIBUTION OF CANCER SCORES

Calibration Set (100 benign & 50 cancer cases)

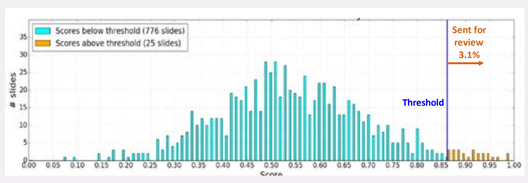


Threshold: was set to 0.86255 to optimize for sensitivity & specificity

Specificity = 98.8%
(11 benign slides out of 917 pass the threshold)

Sensitivity = 96.9%
(218 cancer slides out of 225 pass the threshold)

Study Set (100 benign cases)



25 of the Study set slides, reported as benign, passed the threshold and were sent to review.

After correcting diagnoses based on the review, the specificity of the Study set was similar to that of the Calibration set.

RESULTS OF EXPERTS REVIEW

Diagnoses Distribution

Reviewers' Diagnosis	# slides
Benign/ Benign	7
ASAP/ Benign	2
ASAP/ ASAP	1
Cancer/ Benign	1
Cancer/ ASAP	4
Cancer/ Cancer	10
Total	**25**

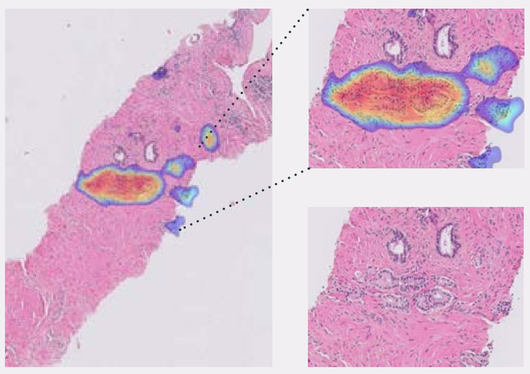
Cancer was diagnosed in 12 unique cases by at least one reviewer.

Characterization of Missed Cancers

Case	Original Diagnosis	Review				
		# slides with cancer	Reviewer #1	Reviewer #2	Gleason Grade Group	Tumor size (mm)
A03278	Benign	1	G3+3	G3+3	1	0.5
A03285	Benign	2	G3+3	G3+4	1	1.8
A03312	Benign	1	G3+3	G3+4	1-2	0.5
A03314	Benign	2-3	G4+4	G4+3	3-4	5
A03316	Benign	1	G3+5	G4+3	3-4	1.3
A03329	Benign	1	ASAP	G3+3	1	0.4
A03337	Benign	1	ASAP	G3+3	1	0.3
A03339	Benign	1	G3+3	G3+3	1	1.2
A03341	Benign	1	ASAP	G3+3	1	0.7
A03360	Benign	1	G3+3	G3+3	1	3
A03375	Benign	1	G3+3	G3+3	1	0.5
A03392	Benign	1	G4+3	G4+3	3	3.3

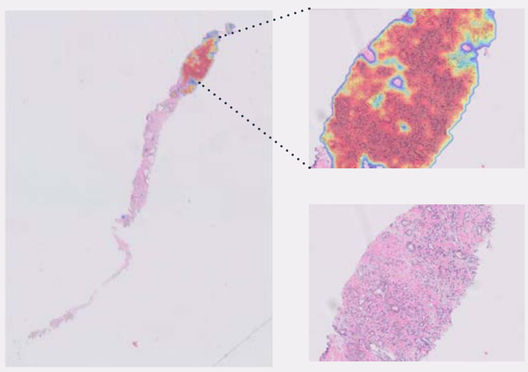
EXAMPLES OF CANCER

Low-Grade Cancer



Diagnosed G3+3 (both reviewers)
Tumor size is 0.5mm

High-Grade Cancer

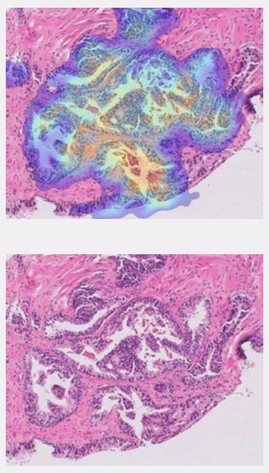


Diagnosed G4+3 (both reviewers)
Tumor size is 3.3mm

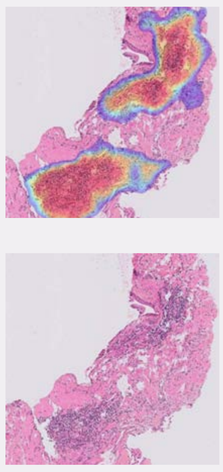
Low Probability — High Probability

EXAMPLES OF OTHER FEATURES

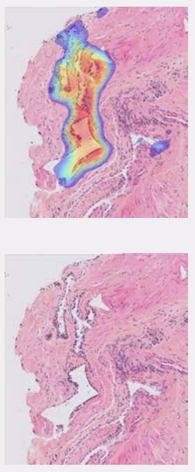
High-Grade PIN



Inflammation



Atrophy



Low Probability — High Probability

CONCLUSIONS

1. The Ibex algorithm for detecting prostate cancer is extremely accurate, and it can identify additional features that are clinically important, such as high-grade PIN, atrophy and inflammation
2. Prostate cancer, typically low-grade tumors, might be missed by pathologists. High-grade tumors are also misdiagnosed, especially when they are small and appear in a single slide
3. Pathology institutes can utilize AI-based systems to increase the accuracy, efficiency and turnaround time of the diagnostic process