Artificial intelligence determined reference value (“rAIght value”) included in virtual histopathology EQA scheme: Comparison of participating pathologists and a trained image analysis algorithm

Jonna Pelanti[1*], Pia Eloranta[1], Juuso Juhila[2**], Anniina Wester[2], Marita Laurila[3], Heidi Berghäll[1]

1. Aurevia, Helsinki, Finland 2. Aiforia Technologies, Helsinki, Finland (**at time of this research project) 3. Department of Pathology, Fimlab Laboratories, Tampere, Finland *jonna.pelanti@aurevia.com

Introduction

The development of digital pathology and artificial intelligence (AI) has made it possible to utilize whole slide imaging (WSI) in addition to an expert evaluation of a pathology slide.

Correct identification of prostate cancer is important to help patients correctly and on time. Prostate cancer samples are evaluated with the Gleason score. The most common and most aggressive grades are added together, resulting in an overall Gleason score for the sample. The overall Gleason score determines the Grade Group (GG) from 1 to 5, where 5 is the most aggressive.

Aurevia organizes a virtual histopathology external quality assessment (EQA) scheme twice a year. In round 2-2023, the topic was Prostate. Aiforia's AI model was used to produce “rAIght values” as additional information, however, the evaluation of the participant performance was based on a reference diagnosis by the scheme expert.

Conclusions

Artificial intelligence tools can support the user's visual interpretation and assist the pathologist in making a diagnosis.

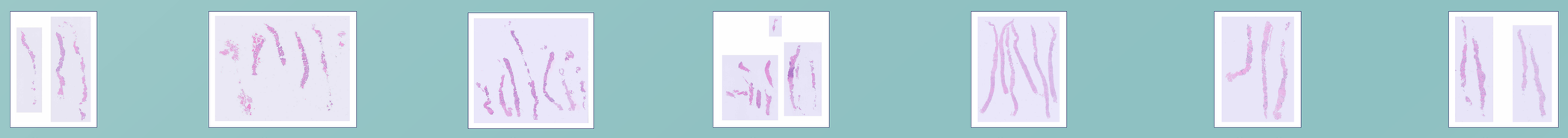
As AI models are able to analyze the whole slide images quickly, they can help to reduce the workload of the medical professionals.

In the final report of the current scheme round, a comparison of results by pathologists and an AI produced "rAIght value” was added to introduce AI in EQA for histopathology.

In this study, the grading of the samples differs somewhat between the participants and the AI model, however, there is also variability in Gleason scoring and GG between the participants indicating that there are challenges in making a diagnosis. In all 7 cases, both the participants and the AI model graded the clinical outcome of the samples such that the patient could have received similar treatment.

Method

Participants were provided 7 scanned slides for analysis using virtual microscopy (cases 1-7). Case specific relevant clinical patient history was included.



Altogether 148-150 pathologists from 15 countries participated in this EQA round.

Risk Group*	Gleason Score	Grade Group
Low/Very Low	Gleason Score ≤ 6	Grade Group 1
Intermediate Favorable/Unfavorable	Gleason Score 7 (3 + 4)	Grade Group 2
	Gleason Score 7 (4 + 3)	Grade Group 3
High/Very High	Gleason Score 8	Grade Group 4
	Gleason Score 9-10	Grade Group 5

*The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Epstein JI, et al. Am J Surg Pathol. 2016;40(2):244-52.

Aiforia AI model produces a “rAIgth value” according to Gleason Grading for comparison.

The evaluation of the participant performance was based on a reference diagnosis by the scheme expert.

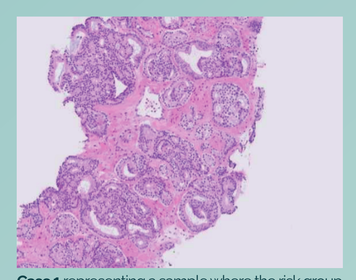
Pathologists were requested to grade the samples according to Gleason Grading:
- most common Gleason score
- most aggressive Gleason score
- Grade Group (GG).

In this study, only GG results were used.

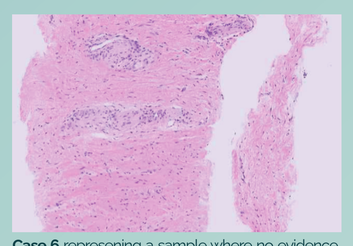
Comparison of results by pathologists and "rAIght value" in the final report as additional information and an introduction to how AI could be used and further developed to ease the workload of pathologists in the future.

Results

	Most of the participants responded	Reference diagnosis by scheme expert	AI model's “rAIght value”		How many (%) of the participants agreed with “rAIght value”	Risk Group by participants	Risk Group by “rAIght value”
Case 1	4	4	4	Gleason Score, most common			
	4	4	4	Gleason Score, most aggressive			
	3	4	3	GG	47 %	Intermediate	Intermediate
Case 2	3	3	3	Gleason Score, most common			
	3	3	3	Gleason Score, most aggressive			
	1	1	1	GG	82 %	Low/Very Low	Low/Very Low
Case 3	3	3	4	Gleason Score, most common			
	4	4	4	Gleason Score, most aggressive		Intermediate	Intermediate
	2	2	3	GG	28 %		
Case 4	4	4	4	Gleason Score, most common			
	4	3	5	Gleason Score, most aggressive		Intermediate	High
	3	3	5	GG	19 %		
Case 5	4	3	3	Gleason Score, most common			
	5	5	4	Gleason Score, most aggressive		High/Very High	Intermediate
	4	4	2	GG	0 %		
Case 6	No evidence of malignancy	No evidence of malignancy	No evidence of malignancy	Gleason Score, most common			
				Gleason Score, most aggressive		No evidence of malignancy	No evidence of malignancy
				GG	90 %		
Case 7	3	3	3	Gleason Score, most common			
	3	3	4	Gleason Score, most aggressive		Low/Very Low	Intermediate
	1	1	2	GG	32 %		



Case 1 representing a sample where the risk group was graded Intermediate by most of the participants and the "rAIght value".



Case 6 represening a sample where no evidence of malignancy was found by the majority of participants and he "rAIght value".

Labquality EQAS by Aurevia