LABQUALITY DAYS

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)

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

Introduction

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. Labquality organizes a virtual histopathology external quality assessment (EQA) scheme twice a year. In round 2-2023, the topic was prostate cancer and participants were given 7 whole specimen scanned slide for analysis.

Aims

The development in the area of digital pathology and artificial intelligence (AI) has made it possible to utilize whole slide images (WSI) in addition to an expert evaluation of a pathology slide. We compared the image GG analysis done by Aiforia's AI model to the scheme participants' visual GG analysis.

Methods

Participating clients were provided with relevant clinical patient history and instructed to analyze 7 scanned virtual microscopy images of prostate biopsies. The samples were formalin fixed and stained with hematoxylin and eosin. 149 individual participating pathologists were requested to grade the samples according to the most common Gleason score, most aggressive Gleason score and the GG. On this round, the analyses conducted by Aiforia's AI model were used as a "rAIght" values as additional information, however, the evaluation of the participant performance was based on a reference diagnosis by the scheme expert. The Aiforia's AI model automatically detects the tumor epithelium and Gleason patterns from the WSIs.

Results

The AI-produced results were in alignment with the results reported by the participants regarding most common and aggressive Gleason scoring and GG in 3/7 cases (cases 1, 2 and 6). For the other cases, the participants reported slightly different Gleason scores and grade groups. For case 3, the majority (49%) of the participants graded it to GG2 and whereas 28% graded it to GG3 which was the AI-produced grading. Case 4 was graded to GG3 by 55 % of the participants whereas AI determined it to be GG5. Case 5 grading had the most discrepancy among the participants with grading varying roughly even from GG2-5 and 19% graded it to GG2 which was the AI-grading. Case 7 was graded as GG1 by 61% of the participants whereas AI-produced grading was GG2 agreeing with 32% of the participants.

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 WSIs quickly, they can help to reduce the workload of the medical professionals. 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 cases, both the participants and the AI model graded the clinical outcome of the samples such that the patient could have received similar treatment.