**Only through international cooperation can AI improve patient lives**

The largest prostate cancer biopsy dataset – involving over 95,000 images – has been created by researchers in Sweden to ensure AI can be trained to diagnose and grade prostate cancer for real world clinical applications.

The researchers will call today, at the European Association of Urology Annual Congress (EAU22), for large-scale clinical trials of artificial intelligence (AI) algorithms and greater global coordination to ensure that AI enhanced diagnostics, prognostication, and treatment selection can help save lives.

There is a shortage of pathologists around the world, both generalists and those specialised in urology. AI can help in detecting prostate cancer at an early stage, but because of the vast differences in the way clinics prepare samples, scan images and in the diverse patient populations they serve, many algorithms do not have universal application.

The team, from Karolinska Institutet (SE), worked with colleagues from Radboud University Medical Center in the Netherlands, University of Turku in Finland and Google Health in the US to run an AI competition involving nearly 1,300 developers from around the world. The developers created algorithms able to grade prostate cancer tumours and trained them using 10,000 international biopsy images. The top performing algorithms outperformed generalist pathologists and matched the average performance of specialist uropathologists.

Dr Kimmo Kartasalo, who will present the results of the competition at EAU22, said: “Grading prostate cancer is a key step in deciding on appropriate
treatment, but it’s a fairly subjective process and differences between pathologists’ assessments can sometimes be large. AI can provide an additional expert opinion, helping to offset the shortage of pathologists and standardize grading. While many algorithms are not widely applicable, those developed in our competition did retain their performance across different patient cohorts.”

PhD Student Nita Mulliqi worked with colleagues at the Karolinska Institutet to prepare the extended dataset of 95,000 prostate biopsy images, the equivalent of more than three years of a single uropathologist’s work. They used biopsies from a clinical trial in Stockholm that lasted around four years from 2012, and obtained images from nine other European laboratories, and many rare disease subtypes from colleagues in Australia.

Mulliqi is now using the dataset to train and test a clinically applicable robust AI based on integrating the best elements of the highest performing entries to the competition into a single, improved algorithm. The extended dataset will ensure that the algorithm can cope with the kind of additional complexity that can be found in a real clinical situation, such as rare cancer types and situations that mimic cancer, but are benign.

Through the research, Mulliqi identified four key areas that require specific attention to ensure better grading and prognosis of prostate and other cancers can be achieved using AI, and that the algorithms can be introduced into clinical use in a responsible manner.

The four areas are:

1. Scanner calibration: ensuring the set-up is the same wherever scans are taking place
2. Improved algorithms: leveraging state-of-the-art AI methodology to ensure robust performance and wide applicability of the algorithms
3. Dataset upscaling: providing larger international datasets to ‘teach’ the AI
4. Modelling morphological heterogeneity: looking at different subtypes of the same disease
Mulliqi will be presenting these findings at EAU22 today. She said: “AI holds great promise and can benefit patients everywhere but in order to achieve this promise, we need an international effort to collect datasets that are representative of the variation in technical approaches and between patients. The combination of our vast database and our colleagues’ algorithms is beginning to show how we can really work together to make a big difference for clinicians and patients.”

Professor Jochen Walz heads the Department of Urology at the Institut Paoli-Calmettes Cancer Centre in Marseille (FR) and is a member of the EAU’s Scientific Congress Office. He said: “AI is going to become a routine tool, which won’t replace pathologists and urologists but will help them reach more consistent decisions. There is currently a lot of variation in the grading of prostate cancers, particularly outside specialist centres.

“This research has used a clever means – crowdsourcing expertise – to develop AI to improve tumour grading and took the next step by validating it against a very varied range of images. This shows that it could be used in general clinical practice.”

“So far, AI has only replicated the grading system used by urologists. But it has the potential to go beyond this – to identify elements within the images that can predict clinical outcomes directly. That is the next challenge for AI.”

-ENDS-

Notes to editors:

About EAU22
Europe’s biggest urology congress will take place from 1-4 July 2022 in Amsterdam, The Netherlands. With nearly 1,300 abstracts presented and moderated live, the 37th Annual Congress of the European Association of Urology (EAU22) will be amongst Europe’s biggest medical congresses in 2022.
Clinicians, scientists, and patients will meet to discuss topics such as:

- Prostate cancer: new developments to improve treatments of the most common male cancer
- Urinary incontinence: a growing concern for the elderly population
- Practice changing treatments for both bladder and kidney cancer
- Prevention and treatment of urinary stones; 1 in 10 people (55 million adults in Europe) will form a stone at some point
- Special track for representatives of patient advocacy group on Monday 4 July

...and many other conditions related to the male and female urinary tract system and male reproductive organs. Review the full scientific programme on the [congress website](#).

**Contact:**
Ruth Francis, Campus PR
Tel: +44 7968 262273
E-mail: [ruth@campuspr.co.uk](mailto:ruth@campuspr.co.uk)

The abstracts, *A robust artificial intelligence approach for histopathological evaluation of prostate biopsies*, and *Crowdsourcing of artificial intelligence algorithms for diagnosis* and *Gleason grading of prostate cancer in biopsies*, are presented to the European Association of Urology Annual Congress (EAU22) in Amsterdam on Sunday 03 July, 2022.

**Abstract:**
**A0613: A robust artificial intelligence approach for histopathological evaluation of prostate**

**Introduction & Objectives**

Examination of biopsies determines the diagnosis and treatment decisions of prostate cancer but is complicated by a global deficiency in pathology expertise and inter- and intra-rater variability. Artificial Intelligence (AI) can aid with these challenges, but its widespread implementation requires tackling diverse patient populations and clinical settings. Biopsies
from different clinics vary greatly in terms of sample preparation and scanning and include morphological heterogeneity (e.g. unusual malignant subtypes or benign cancer mimickers). We propose a comprehensive AI approach for robust clinically applicable prostate biopsy evaluation.

**Materials & Methods**

Our approach integrates: 1) scanner calibration, 2) improved algorithms, 3) dataset upscaling and 4) modelling morphological heterogeneity (Fig. 1). To train robust AI models, clinical samples with linked pathology information are obtained from 9 European laboratories, resulting in a diverse dataset that is an order of magnitude larger than earlier studies (~95,000 whole slide images) and encompasses various scanners (Philips, Hamamatsu, Aperio, 3DHisTech). Further, we apply scanner calibration (Ji et al., abstract AM22-2335) and novel weakly supervised AI algorithms for improved robustness (Kartasalo et al., abstract AM22-0665). Additionally, we model perineural invasion and cribriform morphologies.

**Results**

An initial prototype trained on a single-clinic dataset (N=6682) can detect prostate cancer at a clinically useful accuracy: correctly classifying >85% of benign biopsies while detecting >99% of cancerous cores in a test set (N=1631). Cancer length estimation is highly concordant with the study pathologist (Pearson correlation 0.96). In Gleason grading, concordance of the AI with experts (mean pairwise linear Cohen’s kappa 0.62) is comparable to a standardisation panel of 23 experienced uropathologists.

**Conclusions**

AI can differentiate between malignant and benign prostate biopsies and perform Gleason grading comparably to experts. However, robust performance across laboratories and scanning equipment in real-world clinical settings remain to be improved. We empirically identify 4 key points to be tackled to achieve better prognostication and Gleason grading and allow for evaluating the system in a diagnostic clinical trial.

**Abstract:**

A0611: Crowdsourcing of artificial intelligence algorithms for diagnosis and Gleason grading of prostate cancer in biopsies

**Introduction & Objectives**

Gleason grading of biopsies is crucial for prostate cancer treatment decisions but considerable inter- and intraobserver variability can lead to under- and overtreatment. Moreover, there is a global shortage of pathologists. Artificial intelligence (AI) could potentially mitigate these challenges through partial automation and decision support. While AI has shown promise for diagnosing and grading prostate cancer, the results have
typically not been validated across international cohorts, and robust performance on data from different sites remains a challenge. Competitions have shown to be an efficient way of crowdsourcing medical innovation. To accelerate the development of AI algorithms for Gleason grading, we organized the PANDA Challenge - the largest competition in histopathology to date.

Materials & Methods

We collected the largest public dataset of digitally scanned prostate biopsies to date, consisting of 10,616 specimens from 2,113 patients from Karolinska Institutet and Radboud University Medical Center. The data were provided to algorithm developers through the competition on the Kaggle data science platform (Apr 21 - Jul 23, 2020). Participants could submit algorithms online, and receive performance estimates on a set of 393 biopsies that they did not have direct access to. Performance was evaluated in terms of concordance (quadratically weighted kappa, QWK) with grading by panels of experienced uropathologists. Finally, the contributed solutions were evaluated on a test set of 545 biopsies. Top-ranking algorithms were further assessed on European (330 biopsies) and US (741 biopsies) external validation sets.

Results

In total, 1,290 developers from 65 countries contributed 1,010 algorithms. Top-performing algorithms analysed in detail showed a mean QWK of 0.931 (95% CI: 0.918-0.944) on the internal test set. On US and EU external validation sets the algorithms achieved QWK of 0.862 (95% CI: 0.840-0.884) and 0.868 (95% CI: 0.835-0.900).

Conclusions

We showed that AI algorithms developed as a community effort in a global competition reached pathologist-level performance in Gleason grading of prostate biopsies. The algorithms generalized to intercontinental external cohorts representing different patient populations, laboratories and reference standards, warranting evaluation of AI based Gleason grading in prospective clinical trials. Taken together, this study serves as an example of how important medical problems can be solved through the combination of AI, innovative study designs, and rigorous validation across diverse cohorts.

PANDA Challenge Consortium: Y Cai, DF Steiner, H van Boven, R Vink, C Hulsbergen-van de Kaa, J van der Laak, MB Amin, AJ Evans, T van der Kwast, R Allan, PA Humphrey, H Grönberg, H Samaratunga, B Delahunt, T Tsuzuki, T Häkkinen, L Egevad, M Demkin, S Dane, F Tan, M Valkonen, GS Corrado, L Peng, CH Mermel, PANDA Challenge participant teams