⮕ **PROSTATE CANCER**

Automated systems comparable to expert pathologists for prostate cancer Gleason grading

> " Correlation between the system and expert pathologists for cancer extent was 0.96 "

Gleason grading is a useful tool in predicting prostate cancer behaviour and recurrence. However, reliable grading requires expertise in microscopic examination of cancer specimens that is not always easily available. Now, two studies in *Lancet Oncology* report on the use of artificial intelligence (AI) and deep learning in automated Gleason grading of prostate tumours.

In the first study, Ström et al. used 6,953 digitized slides from 1,069 individuals (6,682 of these slides were from 976 participants in the population-based STHLM3 study) to train an AI system to assess needle core prostate biopsy samples. In addition to Gleason grade, the system was evaluated on its prediction of presence and extent of malignant tissue in an independent STHLM3 test data set of 1,631 samples (from 246 individuals) and an external validation data set of 330 biopsy samples (from 73 individuals). Grading performance was also evaluated by comparison with 87 samples graded by 23 expert pathologists.

The AI system achieved an area under the curve of 0.997 (95% CI 0.994–0.999) and 0.986 (95% CI 0.994–0.999) for distinguishing malignant from benign samples in the independent and external data sets, respectively. Correlation between the system and expert pathologists for cancer extent was 0.96 (95% CI 0.95–0.97) and 0.87 (95% CI 0.84–0.90). For grading, the system achieved a mean pairwise kappa of 0.62 (expert pathologist range 0.60–0.73).

In the second study, Bulten and colleagues developed a deep-learning system to assign Gleason scores to tissue biopsy specimens. A training data set of 933 slides and 4,712 annotated biopsy samples and a tuning data set of 100 slides and 497 annotated biopsy specimens were used to teach gland delineation, Gleason growth patterns and grading. Validation of the system involved an independent test data set of 210 slides and 535 biopsy samples, which was graded by three expert pathologists, and an external test data set of 886 tissue cores, 245 of which were independently graded by two pathologists.

The deep-learning system closely agreed on Gleason grade with the three expert pathologists (quadratic Cohen's kappa 0.918, 95% CI 0.891–0.941) and achieved high agreement with the expert-graded external test data set cores (quadratic Cohen's kappa 0.723 and 0.707).

Both systems provided Gleason grading similar to expert pathologists.

However, in an associated comment, also published in *Lancet Oncology*, Madabhushi and colleagues raise some limitations of these approaches. In particular, they noted a difference in the thresholds used for a cancer diagnosis, commenting: "Notably, a biopsy was only classified as malignant if 10% or more of the tissue was identified as malignant. Pathologists use a much lower threshold, in some cases making a cancerous diagnosis on the basis of 1% or less of the tissue appearing malignant".

Other limitations include the difficulty in using machine-learning and deep learning models in locations outside of the institutions in which they were developed and in which the models were trained. As noted by Madabhushi and colleagues, "preanalytic variability in staining and scanning can affect images in a way that degrades the performance of automated approaches."

These studies demonstrate the potential of automated systems in clinical practice; such systems could reduce clinician workload, provide second opinions, help standardize grading practice and bring expert grading to locations where it is not yet available.

Tim Thomas
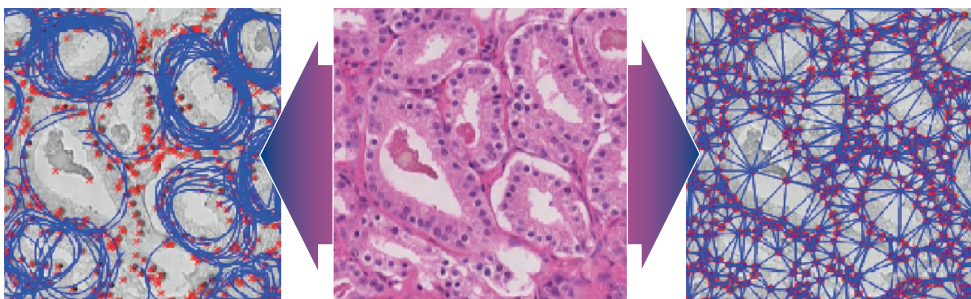
Adapted from Goldenberg, S. L. et al. *Nat. Rev. Urol.* **16**, 391–403 (2019).

ORIGINAL ARTICLES Ström, P. et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol.* **21**, 222–232 (2020) | Bulten, W. et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* **21**, 233–241 (2020) | Madabhushi, A., Feldman, M. D. & Leo, P. Deep-learning approaches for Gleason grading of prostate biopsies. *Lancet Oncol.* **21**, 187–189 (2020)
RELATED ARTICLE Goldenberg, S. L., Nir, G. & Salcudean, S. E. A new era: artificial intelligence and machine learning in prostate cancer. *Nat. Rev. Urol.* **16**, 391–403 (2019)