Google AI outperforms general pathologists at validating Gleason grading of prostate cancer biopsies
24 July 2020, by Bob Yirka

A team of researchers from Google Health, working with others from institutions across the U.S. and Canada has found that a Google AI system was able to outperform general pathologists when validating Gleason grading of prostate cancer biopsies. In their paper published in *JAMA Oncology*, the group describes two major experiments they conducted that compared the abilities of general pathologists against an AI system at categorizing prostate cells.

Over the past several years, Google Inc. has been playing an active role in the development of AI systems for use in diagnostics. Thus far, they have been testing systems for detection of breast cancer on mammograms, lung cancer on CT scans and retinopathy on retinal scans. In this new effort, they have developed systems to detect prostate cancer.

In the first experiment, the researchers asked six pathologists (who had an average of 25 years of experience in their profession) to look at 498 slides of prostate cells with different degrees of cancer. Each of the slides had previously been stained (and privately graded) by urologic subspecialist pathologists in ways that highlighted groups of cells (Gleason grading). The pathologists were then asked to identify which Gleason grade each of the slides should be given, a means of identifying which were cancerous. Once the pathologists concluded their work, the Google AI system then performed the same task and the results were compared. The researchers found that the AI system was 72 percent accurate (as compared to the subspecialists), while the pathologists were just 58 percent accurate. They note that general pathologists are not typically asked to identify cancer in biopsied cell tissue—that is the job of the subspecialist pathologists, which is why they were used as the baseline.

In the second experiment, the researchers asked both the general pathologists and the AI system to look at tissue specimens and to answer whether they believed it was cancerous or not. Out of 752 samples, the pathologists and the AI system were nearly identical—94.3 to 94.7 percent respectively. The AI system scored slightly better, but it also returned slightly more false positives.

The researchers suggest that the Google AI system shows promise as an aid to general pathologists working without the assistance of subspecialist pathologists. They further suggest that more work could lead to the development of systems capable of playing a decision-support role in future prostate screening environments.
