Part 1

Interesting study on the use of a deep learning system (DLS) for Gleason grading of Prostate cancer. The basic findings were that the DLS for grading tumour containing biopsy specimens was in agreement with subspecialists (expert urologists) 71.7% of the time. The DLS was also significantly better at detecting tumour containing biopsy’s then general pathologists 58.0% (p<0.001).

The DLS went through two training stages. The first stage was detailed region-level annotations from prostatectomy and biopsy specimens, which generated 114 million labelled image patches. Labelled data is used for training the DLS to recognise key aspects of the image that might pertain to prostate cancer.

The problem with this approach to training the DLS is that some of the specimens used were taken from prostatectomy’s (removed prostate glands), this training does not generalize well because the predictive model will be applied to patient biopsy’s only, not prostatectomy’s. The authors of the study may not have had a choice in using the prostatectomy data, as DLS require huge amounts of data the train the predictive model. The author of the study touched on this point in the limitations.

Part 2

The second stage of training involved of the DLS involved using 580 biopsy specimens which was definitely appropriate considering the population the DLS is going to be applied to.

The DLS neural network architecture was adapted for Gleason grading via Neural Architecture Search (NAS), which is a way of automatically building a neural network to meet the task that needs to be performed. It can offer significant advantages over hand-built neural network designs in terms of performance.

The tumour detection between the DLS and the general pathologists (GP’s) were similar (DLS 94.3% vs 94.7%). The DLS detected more tumours (increased sensitivity) than the GP’s although that was at the expense of more false positives.

False positives=1-Specificity, DLS=8.3% vs GP’s=3.0% (Have the disease when actually they don’t)

False negatives=1-Sensitivity, DLS=4.5% vs GP’s=7.2% (Don’t have the disease when they actual do).

The high false positive rate may be due to the training data for DLS using more prostatectomy specimens than prostate biopsy’s.

Part 3

The agreement rate for Gleason grading (as outlined at the start of this comment) in the DLS was higher in the subspecialist group 71.7% than in the GP’s group 58.0%. This finding may help GP’s spot more tumours which would have been otherwise missed.

In summary a good study, the author cites more work to be done on determining how DLS can be used to alter the current decision making process in terms of increasing accuracy and consistency. May be ‘tweaking’ the neural network design manually to the one obtained through NAS may be a way to improve the predictive model.