



# Introducing automated diabetic retinopathy systems: it's not just about sensitivity and specificity

Caroline Jane Styles<sup>1</sup>

Received: 3 July 2019 / Accepted: 5 July 2019 / Published online: 29 July 2019  
© The Royal College of Ophthalmologists 2019

Gulshan et al. [1] present data to demonstrate the feasibility of using an automated grading system to expand diabetic retinopathy (DR) screening programmes. The automated system was used to compare manual grading from two sites in India using 5762 images from 3049 individuals. They have compared the sensitivity and specificity of the automated DR grading system against manual grading for the identification of moderate or worse diabetic retinopathy (DR), and diabetic macular oedema (DME) surrogate markers. They show that the automated system matched or exceeded those of manual grading.

They recommend using deep learning technology in all healthcare systems and emphasise its role in addressing a lack of trained graders in India, and also to allow results to be delivered immediately in other low to medium resource settings. In high resource settings they suggest using machine learning to concurrently read all images with manual grading.

In Scotland we have used an automated grading system since 2012 [2, 3]. We also recognised that trained graders are a scarce resource and as our programme matured, many of our graders had many years of experience but were spending a large proportion of their time on the repetitive task of examining images with no retinopathy. Introducing automated grading was an opportunity to ensure that highly trained manual graders worked at the top of their skill set to ensure timely identification of those with sight-threatening retinopathy changes from the many thousands with no retinopathy.

All retinal images taken in the Scotland diabetic retinopathy screening (DRS) programme have a first pass through the 'autograder'. Those images that are gradable and have

no 'microaneurysms' are given a final grade of ROM0, with an outcome currently set as rescreen 1 year. All other images that do not fit these criteria are then passed to a manual grader. We do not use it as a concurrent read system. In 2018, just over 182,000 images were processed by the autograder with 50.4% graded as ROM0 with no further grading required. Overall, a higher percentage of images have a final grade of ROM0 (after manual grading) reflecting the lower specificity at which we have set the parameters of our system.

The introduction of automated systems to our programme did not just involve discussions about sensitivity and specificity though. Gulshan et al. [1] give elegant examples in Fig. 2 of disagreements between human and manual graders as well as 'missed' pathology by both. Programmes need to debate and come up with robust processes for dealing with these issues. In Scotland the autograder undergoes the same internal quality assurance as a manual grader (500 images per year from each grader are reread by a senior grader). It also is tested with the 100 images used in the biannual external quality assurance that all graders undertake in Scotland. A process for raising issues with automated systems needs to be in place along with an understanding that the sensitivity and specificity of these algorithms is not 100% and there will be disagreements about DR pathology between graders and the automated systems in both ways.

Gulshan et al. also discuss the detection of other retinal disease, such as age related macular degeneration and glaucoma, which might opportunistically be identified during diabetic retinopathy screening by manual graders. The automated processes do not do this, and this was long debated prior to the introduction of automated grading into our programme. DRS services that have functioned for many years prior to the introduction of automated processes are likely to consider that detection of other retinal pathology is part of the programme. The core purpose of diabetic retinopathy screening is to identify sight threatening diabetic eye disease so that treatment can be offered early and

✉ Caroline Jane Styles  
caroline.styles@nhs.net

<sup>1</sup> Department of Ophthalmology NHS Fife, Queen Margaret Hospital, Dunfermline Fife, UK

prevent sight loss. The acceptance of this principle can be challenging, and the involvement of public health screening experts was very important in our discussions.

Just as importantly, people with diabetes eligible for screening must always be informed that regular, full eye examinations are also necessary to detect other eye disease. Following the introduction of the autograder in Scotland, our DRS information leaflet was altered to include the statement:

*'Your screening photographs will be graded either by a health professional or an automated grading system to detect diabetic retinopathy but not any other eye conditions. You should continue to visit your optometrist regularly for a free eye check as well.'*

There are now many papers demonstrating the performance of deep learning in the detection of diabetic retinopathy with high sensitivity and specificity rates published. In Scotland we have adopted available technologies as part of our processes. In doing so, we learnt that our health culture and governance structures play an important part in their adoption. Our blackbox technology is now being overtaken by deep neural network training technologies which promise much more sophisticated DR grading than we currently use. We are starting to consider how they may

fit into our programme in the future. It will be new to us that we don't understand or control exactly what artificial intelligence does, so once again sensitivity and specificity will not be the only issue to debate.

### Compliance with ethical standards

**Conflict of interest** The author declares that she has no conflict of interest.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### References

1. Gulshan V, Rajan R, Widner K, Wu D, Wubbels P, Rhodes T, et al. Performance of a deep learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmology*. 2019. <https://doi.org/10.1001/jamaophthalmol.2019.2004>. Published online 13 Jun 2019.
2. Philip S, Fleming A, Goatman K, Fonseca S, McNamee P, Scotland G, et al. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme. *BJO*. 2007;91:1512–7.
3. Fleming A, Goatman K, Philip S, Prescott G, Sharp P, Olson J. Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts. *BJO*. 2010;94:1606–10.