Response to ‘Improved Automated Screening of Diabetic Retinopathy’ by Carlos M. Oliveira et al.

Alan D. Fleming a  John A. Olson b  Peter F. Sharp a  Keith A. Goatman a  Sam Philip b

a University of Aberdeen and b Diabetes Retinal Screening Service, Aberdeen, UK

The recent paper by Oliveira et al. [1] describes a reasonable study of their automated microaneurysm detection method. The study also includes automated detection of changes in sequential images to test whether this can be used to improve the accuracy of detection of diabetic retinopathy. However, the assessment of this novel ‘two-step’ approach is flawed.

This is because the study did not evaluate the sensitivity of this new technique. Change between subsequent photographs was assessed only on cases which had been graded as positive by the automated system but which had not been referred for treatment by manual graders (referred to as ‘false positives’). Change was not assessed on the cases referred for treatment by the human grader. Manual grading was used to select the cases used to assess the new technique. The study made an estimate of the specificity but not the sensitivity of the new technique. (The reported sensitivity for the new technique is the sensitivity of automated grading without change assessment but slightly modified, we understand, since 19 cases changed retinopathy status when the later manual gradings were used.)

We also have concerns that the sample of 289 ‘false positives’ is not only very small but may not be a random sample from the total of 2,378 cases. It follows that the specificity estimate of the new technique is unreliable and the decrease in sensitivity, compared to microaneurysm detection alone, is unknown.

The comparisons made between different automated systems [1, table 1] are also flawed. Different automated systems have been evaluated using different image sets and grading methods. Manual graders, even those operating to a common protocol, are known to have different performances [2], so it is not known whether differences in results are due to differences in manual grading, automated grading or both. The concluding sentence of the abstract is not only misleading but also factually wrong in the context of table 1 of Oliveira et al. [1].

We would also point out that the system described is not a fully automated grading system since it has no quality assessment. As manual quality grading of all images would be required, the reduction that this system makes to the burden of manual grading could be less than with other systems [3, 4].

This study has therefore not shown that there is an improvement in automated grading performance by using repeat photographs nor that their system has a better performance than the other reported systems.

Disclosure Statement

Funding for A.D.F. was provided by Medalytix Ltd. under an agreement with Scottish Health Innovations Limited. The authors, the University of Aberdeen and NHS Grampian may receive some remuneration as a result of agreements with Medalytix and Scottish Health Innovations Limited.

References