Introduction

Diabetes is a global epidemic affecting 422 million people worldwide. A common complication of diabetes is diabetic retinopathy. This is a condition of the small vessels in the retina and presents in different stages. In 2011, 93 million people worldwide suffered from diabetic retinopathy. In Ireland alone, 1 person a week goes blind due to this condition (diabetes.ie).

Diabetic retinopathy is detected through screening. The most common method of obtaining retinal images is fundus photography. Normally, images obtained from screening are analysed manually by trained, accredited human graders. It takes a human grader 1.5 times longer to decide an image is normal than to spot disease (Tan, 2014). This is because the markers of disease can be very obvious e.g haemorrhages, hard exudates, cotton wool spots, microaneurysms and growth of new blood vessels. Automated retinal image analyses (ARIA) could be used as a workload reduction tool in the grading of retinal images. An automated system that could safely reduce the number of ‘no disease cases’ would make the grading process more efficient and cost effective.

Our aim is to compare the IDx DR and Medalytix automated systems grades to the gold standard of trained accredited human graders.

Materials and methods

We obtained two datasets from a local ophthalmologist in the form of Excel spreadsheets. The first dataset consisted of 95 cases, all of which were Type 2 diabetic, Irish patients. Their retinal images had been graded by trained accredited Irish graders and also by the IDx DR automated system. Similarly the second dataset contained 535 cases, all of which were Irish patients. Their retinal images were graded by the same Irish human graders and the Medalytix automated system. We began with the IDx dataset and created a new Excel spreadsheet containing the two grades, from the manual graders and the automated system. We then changed these grades to numerical grades based on the scale seen here. We used Excel operations to find and sort the agreements, false positives, false negatives, true positives and true negatives. From these figures we used the following formulae to find the IDx DR statistical values.

Similarly, we created a new spreadsheet from the original Medalytix dataset. We also converted the manual and automated grades to numbers using another grading scale (shown here). Then we used the same operations to find the systems agreements, true positives and negatives, false positives and negatives. We subsequently found the Medalytix automated systems sensitivity, specificity, accuracy, PPV and NPV using the equations as above. After analysing the IDx spreadsheet, an ophthalmologist reviewed the results and changed some cases, so we changed the statistical values accordingly. We also did another analysis of the IDx grades using a grading scale similar to the one we used for the Medalytix data for additional results.

LITERATURE CITED


RESULTS:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDx-DR</td>
<td>0.915</td>
<td>0.954</td>
<td>0.880</td>
<td>0.933</td>
<td>0.857</td>
</tr>
<tr>
<td>Medalytix</td>
<td>0.915</td>
<td>0.954</td>
<td>0.880</td>
<td>0.933</td>
<td>0.857</td>
</tr>
</tbody>
</table>

The results show that both the IDx-DR and Medalytix automated systems have similar accuracy, specificity, sensitivity, positive and negative predictive values. The IDx-DR system had a slightly higher accuracy and specificity than the Medalytix system. However, both systems perform well and can be used as a workload reduction tool in the grading of retinal images.

Conclusions

To conclude our results show that both ARIA and Manual grading methods can be used as an effective method for removing “no disease” images from the manual grading queue. The negative predictive values of both systems are over 93% and the specificities of both are over 93% demonstrating that both systems can identify “true negative” cases well.

However, with relatively low sensitivities (both below 59%) it is questionable if an automated system would be entirely effective and clinically safe in the grading process of urgent referral cases. The original sensitivities and specificities of both systems don’t meet the requirements of the World Health Organization (WHO) guidelines. Additionally, in the grading process of minimal diabetic retinopathy cases a lower sensitivity and specificity may be acceptable.

Medalytix did appear to be the most effective at grading of the two systems from our comparative statistical graph as all of the statistical values found from Medalytix, other than the sensitivity, were higher than those of IDx-DR. This could be due to the fact that the protocol Medalytix is modelled on (NSC) is similar to the protocol the manual graders are trained on (ENDPS) in comparison to the protocol IDx follows (CDR).

The statistical scale we used for the Medalytix spreadsheet may have caused the results to be higher in general as having only two grades made it easier for results to agree. We used our grading scale on the IDx dataset, we found that this helped the results agree also and gave very high sensitivities and specificities.

Overall an automated system would be a cost effective, time saving and safe alternative to identifying negative disease cases manually. A solution to the low sensitivities of the systems would be human review of all positive cases. The deep learning feature on automated systems could help improve their performance into the future (Gulshan) Our suggestion is for automated retinal image analysis to be used in synergy with manual graders. The increasing burden of diabetic screening will only worsen as the epidemic of diabetic grows in Ireland and worldwide.