Pursuing Evaluation of an Automated System to reduce Human Grading Burden in a Diabetic Retinopathy Screening Program

Catarina A. Neves¹A, Carlos M. Oliveira², Maria Luisa Ribeiro¹, Helder J. Ferreira³, Jose G. Cunha-Vaz¹

¹A Coimbra Ophthalmology Reading Center, ¹Association for Innovation and Biomedical Research on Light and Image (AIBILI), Coimbra, Portugal; ²Critical Health, Coimbra, Portugal; ³Health Administration of Central Region of Portugal, Coimbra, Portugal.

PURPOSE:
To evaluate in a prospective study the efficacy of an Automated Software System to reduce the human grading burden in a Diabetic Retinopathy (DR) Screening Program.

INTRODUCTION:
The presented data is related to the DR Screening Program currently ongoing in the central region of Portugal.

An automated system capable of identifying signs of DR pathology, the RetmarkerDR Screening, has been implemented under this Program. Here is presented the current data which demonstrates the efficacy of the automated system in reducing, in a similar manner, the human grading burden related to the grading of color fundus images.

A total of 7986 patients (15866 eyes – more than 30000 images) were included from September 2011 to March 2012. Non-mydratic color fundus photography (4 images per patient) acquired in 3 different local health centers were analyzed and graded at our centralized reading center.

METHODS:
The objective of DR Screening is to identify eyes/patients that need to be referred for a specialist appointment, and the ones that do not need immediate referral, and therefore should return one year later for the annual screening.

The analysis/grading is performed per eye, to evaluate the presence of referable DR (maculopathy and/or proliferative diabetic retinopathy) in two different steps:

* **Automated analysis (1st step)**
  
  The RetmarkerDR Screening analyzes all image sets received and separates those that need human grading from those for which no human grading is needed, based on whether signs of DR pathology were identified or not.

  Human grading is thus reduced only to those image sets identified as having signs of DR pathology present.

* **Human grading (2nd step)**
  
  All image sets identified as having signs of DR pathology, or with comments from the photographer independently of the automated analysis result, are sent to human grading and classified as:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC</td>
<td>Not Classifiable</td>
<td>Repeat if indicated</td>
</tr>
<tr>
<td>R0</td>
<td>DR absent</td>
<td>Annual rescreen</td>
</tr>
<tr>
<td>RL</td>
<td>Non-proliferative DR</td>
<td>Annual rescreen</td>
</tr>
<tr>
<td>M</td>
<td>Maculopathy</td>
<td>Referral ASAP</td>
</tr>
<tr>
<td>RP</td>
<td>Proliferative DR</td>
<td>Urgent referral</td>
</tr>
</tbody>
</table>

Implementation

For quality control and to test the safety of the system about 25% of the image sets marked as not requiring human grading were sent for human grading. The grader was blind to this process.

Two software tools were implemented to assist the grading procedure, a grading tool, and a statistical analysis/reporting tool (portrayed below). The combination of these tools provides a full solution to the reading center.

RESULTS & DISCUSSION:

**Results & Discussion (Cont.):**

The automated system shows a specificity of 48.39% and a sensitivity of 98.11%.

These results from the deployed system currently in use are in agreement with previous published data on a retrospective study by our group on this subject [Ophthalmologica 2011;226(4):191-197].

Furthermore, when comparing the results between the health centers involved in the program, it was possible to identify anomalous situations and take corrective actions, such as training needs or equipment requiring maintenance.

Remark: Patients with specific comments provided by the photographer, independently of the automated analysis results, are additionally sent to human grading. This brings additional safety to the program. The mentioned case not detected by the software had comments from the photographer about low image quality due to Miosis and, therefore, it was sent to human grading.

CONCLUSIONS:

Prospective analysis of the efficacy of the RetmarkerDR Screening, an automated software system to analyse color fundus photographs in the context of a DR Screening Program, shows that it has a potential of human grader burden reduction of 48%.

Such a burden reduction brings a strong economic impact, potentially reducing the screening costs and thus increasing the cost effectiveness of the program. It also enables a larger population to be screened with the same resources and within the same time frame.

The high sensitivity means that the software is safe (only one case was missed but it was still referred to human grading due to photographer comments).

An automated system additionally enables comparisons and early corrective actions when anomalous situations occur.