TITLE

The evidence for automated grading in diabetic retinopathy screening

AUTHORS

Alan D Fleming[1], Sam Philip[2], Keith A Goatman[1], Gordon J Prescott[2], Peter F Sharp[1], John A Olson[2]

AFFILIATIONS

1. College of Life Science and Medicine, University of Aberdeen, Foresterhill, AB25 2ZD

2. Diabetes Retinal Screening Service, David Anderson Building, Foresterhill Road, Aberdeen AB25 2ZP

CORRESPONDING AUTHOR

Alan Fleming
Medical Physics Building
School of Medicine and Dentistry
Aberdeen University and Grampian University Hospitals
Foresterhill
Aberdeen AB25 2ZD
Scotland
T: (+44) (0)1224 553195
F: (+44) (0)1224 552514
E: a.fleming@abdn.ac.uk

ABSTRACT

Systematic screening for diabetic retinopathy using retinal photography has been shown to reduce the incidence of blindness among people with diabetes. The implementation of diabetic retinopathy screening programmes faces several challenges. Consequently, methods for improving the efficiency of screening are being sought, one of which is the automation of image grading involving detection of images with either disease or of inadequate quality using computer software. This review aims to bring together the available evidence that is suitable for making a judgement about whether automated grading systems could be used effectively in diabetic retinopathy screening. To do this, it focuses on studies made by the few centres who have presented results of tests of automated grading software on large sets of patients or screening episodes. It also considers economic model analyses and papers describing the effectiveness of manual grading in order that the effect of replacing stages of manual grading by automated grading can be judged. In conclusion, the review shows that there is sufficient evidence to suggest that automated grading, operating as a disease / no disease grader, is safe and could reduce the workload of manual grading in diabetic retinopathy screening.

Key words

diabetic retinopathy, screening, computer-assisted image analysis, imaging, telemedicine

INTRODUCTION

Early detection of diabetic retinopathy allows this disease to be treated before it becomes symptomatic so that sight loss can be limited. In the UK, diabetic retinopathy screening is run by the NHS aiming to reduce the incidence of blindness in people with diabetes and is based on digital photography and slit-lamp examination for technical failures [1].

Typically around 70% of screened patients are normal [2], and hence the initial grading task is to perform disease / no disease grading, identifying and removing all normal images and retaining those with some disease or other abnormality for further scrutiny. This is followed by one or more stages of full-disease grading and possibly arbitration grading. The scale of this task is huge; in the UK screening programmes there are approximately two million retinal image sets that require grading each year. Therefore, computer automation has been considered to assist the grading process [3,4].

The intention of this review is to guide potential users of automated retinopathy grading about whether current systems are suitable for introduction into screening. Therefore it only includes papers on automated disease detection that cover studies on large datasets similar to what would be found in a screened population.

A section on manual grading is also included since a decision on which of the alternative systems will provide the greatest benefit requires knowledge of the performance of each option.

Indications for the feasibility of automated grading

Most regional screening programmes generate tens to hundreds of thousands of images to be graded each year. Computers excel at repetitive tasks and, if the necessary infrastructure is in place, it is likely that computer grading will be cheaper than manual grading. There are several indications that computers would be capable of this task;

- Even if a computer is only capable of grading at a disease / no disease level, this would still provide a large reduction in the manual grading workload. This is because a large proportion of the patients attending diabetic retinopathy screening are normal.
- The images are relatively constrained in that normal images show a very standard structure. They are also uncluttered in that lesions are usually separate from normal features.
- The digital network infrastructure is already in place that could allow centralised computer processing of screening programme images.

In addition, over the last two decades, many research groups have shown that it is possible to develop software that, in small tests, can identify retinal lesions to a high accuracy and details may be found in previous reviews that have focussed mainly on the algorithms involved [5-8]. However only large studies can inform decisions on the introduction of automated grading into screening since there is no certainty that the performance found using a small image set will be maintained in practice.

The role that automated grading may play in screening is open to debate. Since most of the available evidence relates to automated "disease / no disease" grading, this will be the main emphasis of this article. However, a section is devoted to other roles.

Terminology

The term "referable retinopathy" is used here to mean retinopathy, including maculopathy, more serious than mild non-proliferative retinopathy as defined by the Early Treatment of Diabetic Retinopathy Study scheme [9] or background retinopathy in the English or Scottish schemes [10,11]. The term "maculopathy" is used to mean the presence of photographic surrogate markers indicative of macula oedema requiring more frequent observation than the standard screening interval or referral to ophthalmology [10,11].

Grading roles of “disease / no disease” grading and “full disease” grading are explained in the UK National Screening Committee Workbook [10]. The process of disease / no disease grading separates episodes showing any disease or abnormality from those which are normal. Full disease grading takes place after disease / no disease grading and identifies episodes requiring a clinical outcome which is not recall at the default screening interval. It assigns a grade according to disease severity.

METHODS - SELECTION OF STUDIES

A search was made for peer-reviewed studies from journals using ISI Web of Knowledge with the specification for paper topic:

(automat OR comput*) AND (detection OR diagnosis OR grading) AND diabet* AND retinopathy*.

The search found 747 papers and these were manually examined using the title and abstract to find those whose main topic was automated grading or computer detection of diabetic retinopathy and its associated signs in retinal photographs. This resulted in 82 papers. The following selection criteria were then used to select those which may be suitable to draw conclusions on the use of automated retinopathy grading within healthcare:

1. The study should be an assessment of an automated image analysis system applied to detection of retinopathy in retinal photographs.
2. The results should be reported for referable retinopathy on a per-patient or per-episode basis, or it should be based on such data.
3. The study should be based on tests with at least 200 subjects with and 200 without referable retinopathy.

Criterion 3 means that sensitivity and specificity for detection of referable retinopathy can be estimated with a confidence interval of less than ±5% at a sensitivity or specificity of 90%. Seven papers satisfied criteria 1 and 2 but were rejected because of study size [12-18] with the largest having 95 [12] patients with referable retinopathy.

The remaining papers were divided into efficacy studies reporting original results for the sensitivity or specificity for detection of referable retinopathy and economic studies describing economic analysis based on efficacy results presented elsewhere.

THE EVIDENCE

Six efficacy studies and two economic studies were identified that satisfy the above criteria and are based on three systems, Utrecht / Iowa, Aberdeen and Brest, as shown in table 1. Only the Utrecht / Iowa and Aberdeen systems can be considered complete automated grading systems since only these have an image quality assessment module. Table 2 shows the performances demonstrated in each paper and includes further study details.

Table 1: Details of the automated systems described in studies listed in table 2. Two of the systems contain multiple modules; however, not all studies used all modules.

References	System	Developed at	Photographic protocol used	Modules
Abramoff 2008 [19] Niemeijer 2009 [20] Abramoff 2010 [21]	Utrecht / Iowa	Image Sciences Institute Utrecht and University of Iowa	Two fields per eye, mydriasis as required.	Red lesion detection Exudate detection Assessment of quality
Philip 2007 [22] Scotland 2007 [23] Fleming 2010 [24] Scotland 2010 [25] Fleming 2010 [26]	Aberdeen	University of Aberdeen	One field per eye, mydriasis as required.	Microaneurysm detection Haemorrhage detection Exudates detection Assessment of quality
Abramoff 2010 [21]	Brest	Brest University Hospital, France	Two fields per eye, mydriasis as required.	Microaneurysm detection

Table 2: Studies on automated image analysis in diabetic retinopathy that satisfy the selection criteria. Study size is given along with the number of cases having referable retinopathy.

Study	Type of study	Automated system details	Study size (referable retinopathy)	Case selection strategy	Detection rates		Specif-icity
Philip *et al* 2007 [22]	Efficacy	Aberdeen system: microaneurysm detection and quality assessment	6672 (330)	Screening programme cohort – all cases included	Referable retinopathy / ungradable cases Referable retinopathy Mild retinopathy Ungradable cases	98.9%* 97.9% 80.9% 99.5%	52.4%*
Scotland *et al* 2007 [23]	Economic modelling	*As for Philip 2007* [17]					
Abramoff *et al* 2008 [19]	Efficacy	Utrecht / Iowa system: haemorrhage/microaneurysm and exudate detection and quality assessment	7689 (378)	Screening programme cohort – ungradable cases removed	Referable retinopathy Ungradable cases	84.4%* 80%	64.3%*
Niemeijer *et al* 2009 [20]	Efficacy	Utrecht / Iowa system: haemorrhage/microaneurysm and exudate detection and quality assessment	15000 (394)	Screening programme cohort – all cases included	Referable retinopathy / ungradable cases	93.0%	60.0%
Fleming *et al* 2010 [24]	Efficacy (automated system comparison)	Aberdeen system: (1) microaneurysm detection and (2) microaneurysm, haemorrhage and exudate detection and quality assessment	7568 (1253)	Screening programme – stratified selection	Referable retinopathy / ungradable cases Referable retinopathy Mild retinopathy Maculopathy Proliferative referable retinopathy Non-proliferative referable retinopathy Ungradable cases	97.3%* 96.9% 79.0% 95.4%* 97.4% 98.9% 98.8%	49.0%
Scotland *et al* 2010 [25]	Economic modelling	*As for Fleming 2010* [21]					
Abramoff *et al* 2010 [21]	Efficacy (automated system comparison)	(1) Utrecht / Iowa system: haemorrhage/ microaneurysm detection (possibly combined with quality assessment*); (2) Brest system: microaneurysm detection.	15980 (793)	Screening programme cohort – ungradable cases removed	*Utrecht/Iowa system* Referable retinopathy *Brest system* Referable retinopathy	 90.0% 90.0%	 47.7% 43.6%
Fleming *et al* 2010 [26]	Efficacy	Aberdeen system: microaneurysm detection and quality assessment	33535 (2214)	Screening programme cohort – all cases included	Referable retinopathy / ungradable cases Referable retinopathy Mild retinopathy Maculopathy Proliferative referable retinopathy Non-proliferative referable retinopathy Ungradable cases	98.8% 97.8%* 83.9% 97.3% 100% 100% 99.8%	41.1%*

*The paper does not make it clear whether quality assessment was included or not.
*Calculated from figures in the paper but not reported directly therein.

The efficacy of automated grading

Table 2 illustrates that the Aberdeen system achieved higher sensitivities than the Utrecht / Iowa system for referable retinopathy (including ungradable cases) though the average specificity over the papers was lower for the Aberdeen system. When comparing sensitivities and specificities it should be borne in mind that there is a trade-off between sensitivity and specificity. (During development, the sensitivity of a system can be increased at the expense of specificity and vice versa.) A comparison between two systems is difficult, except where they are tested in the same study, because the study populations are different and because different people created the reference grading. Only one direct comparison has been made between two systems, the Utrecht/Iowa and Brest systems [23]. This considered disease in gradable images. The Utrecht / Iowa and Aberdeen systems have been designed in the context of different photographic protocols; the Utrecht / Iowa system is based on a protocol used in the Netherlands with two photographs per eye while the Aberdeen system is based on the Scottish protocol with a single field per eye.

Detection rates for proliferative retinopathy have been reported only for the Aberdeen system and was found to be 100% [22,26] and 97.4% [24]. These studies reported detection rates for maculopathy to be 97.3%, 97.8% and 95.4% (calculated figures) respectively.

Three of the studies avoided selection criteria that required examination of the images, one using the Utrecht / Iowa system [20] and two using the Aberdeen system [22,26]. Unfortunately, one of these has limited data on clinical performance because it is published in a technical journal [20]. Two studies with the Utrecht / Iowa system involved removal of ungradable cases in a major part of the analysis and one study with the Aberdeen system involved data stratified according to whether retinopathy was referable or not [24].

One study with the Aberdeen system used 7 graders to arbitrate the false negatives of automated grading compared to the screening programme manual grading and thereby provided a consensus grade for these images [26]. It showed that no cases of non-maculopathy referable retinopathy were missed by the automated system and that the detection rate for maculopathy was 97.8%. An analysis was also made of false negatives with the Utrecht / Iowa system [19]; of 87 missed cases of referable retinopathy according to the manual grading, 24 had large haemorrhages or neovascularisation, 23 had small haemorrhages, 18 had exudates or cottonwool spots and 22 were not diabetic retinopathy according to a second expert. In Philip *et al* 2007, two false negative eyes, graded as having proliferative retinopathy but missed by the Aberdeen system, were found to have only mild diabetic retinopathy at eye clinic examination [22].

Two studies using different systems have looked at the relative benefits of using detectors for the various dot-lesion types. For the Utrecht / Iowa system it was found that detection of referable retinopathy (including maculopathy) is improved by using a combination of bright and red lesion detectors rather than the individual detectors [20]. These results are compatible with those reported for the Aberdeen system which had significantly better detection of referable retinopathy when using microaneurysm, haemorrhage and exudate detection in combination compared to using microaneurysm detection alone [24]. The improvement was specifically for the maculopathy cases. For proliferative retinopathy, the addition of haemorrhage and exudate detection had no effect on performance.

All of the studies listed in table 1 used dot-lesion detection; vascular abnormalities were not explicitly detected. This suggests that detection of dot-lesions, without detection of vascular abnormalities, is sufficient for referable retinopathy detection.

The efficacy of manual grading

The impact of replacing part of manual grading with automated grading can only be known if the efficacy of both manual and automated grading is known.

The evidence concerning the "reliability", meaning the reproducibility, of manual grading has been reviewed in Benbassat and Polak 2009 [27]. The results of the reviewed studies show that reproducibility of grading is very rarely 100%, implying also that sensitivity is below 100%.

Table 3: Results of studies assessing sensitivity and specificity of manual grading of referable retinopathy. Papers using the term "sight threatening retinopathy" were included.

Study	**Total study size (number with referable retinopathy)**	**Sensitivity %**	**Specificity %**	**Notes**
Liesenfeld 2000 [28]	115 (13)	85	90	Median of 6 results
Olson 2003 [29]	545 (55)	93	87	
Stellingwerf 2004 [30]	197 (70)	86 92	93 93	Results for TIFF images
Arun 2006 [31]	498 (62)	93.5	97.8	
Ruamviboonsuk 2006 [32]	400 (44)	93 100	97 73	Best sensitivity and specificity results are given
Philip 2007 [22]	6722 (330)	99.1	73.9*	
Abramoff 2008 [19]	500 (unknown)	62 85	84 89	Best sensitivity and specificity results are given

* Calculated from figures in the paper but not reported directly therein.

Some example results for studies that reported sensitivity and specificity of manual grading for referable retinopathy are given in table 3. Sensitivity and specificity are easier to interpret than kappa agreement coefficient which, though used by many authors, has the disadvantage that it is affected by prevalence. It should be borne in mind that the reference for these assessments is itself a judgement, though hopefully produced in a more robust manner than the grading being assessed. It is likely that there are great differences in performances between professional groups and between individuals as suggested by Ruamviboonsuk *et al* 2006 [32]. This and two other studies compared multiple graders on the same image set [19,30]. Philip *et al* 2007 is the only study to satisfy the size criterion applied in this review for selection of studies on automated grading.

Cases of referable retinopathy are not only missed during grading; they are also missed due to the nature of digital photography. Scanlon *et al* 2003 measured the sensitivity and specificity for digital photography against slit-lamp biomicroscopy [33]. These were 87.8%, and 86.1%, respectively, for mydriatic 2-field photography.

Comparing automated and manual grading

A direct comparison of automated and manual grading, using the same dataset, has been made only in one study [22]. Both automated and manual graders were operating as disease / no disease graders. For the manual system, a sensitivity of 99.1% was reported for detection of referable retinopathy at a specificity of 73.9% (calculated figure). For automated grading the sensitivity was 97.9% at specificity 52.4% (calculated figure). The difference between automated and manual grading sensitivities for referable retinopathy was not statistically significant at the 5% level. For mild retinopathy and for ungradable images, the automated system had higher sensitivity than the manual system.

The sensitivities and specificities of detection of referable retinopathy by manual graders and by automated grading as listed in tables 2 and 3 are displayed in figure 1. Specificity is consistently higher for manual compared to automated grading. This suggests that automated grading cannot completely replace manual grading but, because automated grading has a high sensitivity, it could safely reduce workload.

The following considerations should be taken into account when comparing the results shown in figure 1.

Firstly, some of the manual grading results displayed in table 3 and figure 1 may have been from graders operating at a level of specificity which is suitable as the final specificity of screening; this is a level at which ophthalmology services are not swamped by unnecessary referrals. A trade-off exists between sensitivity and specificity and so this specificity may only be possible by operating at lower levels of sensitivity.

Secondly, the lower specificity of the automated grading systems compared to manual grading means that patients in which automated grading detected any abnormality would have to be referred to manual grading. Therefore, a combined automated and manual grading system would be necessary so that the specificity of the screening programme is acceptable. However, the sensitivity of multiple sequential grading stages is inevitably lower than the sensitivity of any individual stage. A comparison between the combined performances of complete grading systems would be more useful than a comparison of the individual grader performances that are shown in figure 1.

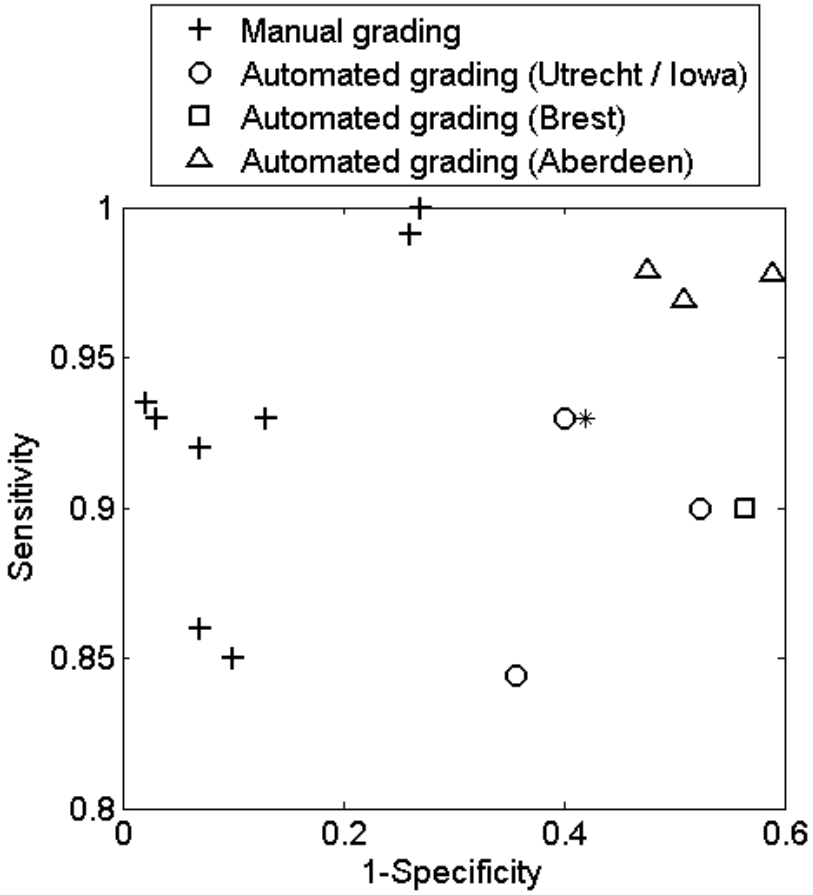

Figure 1. Specificity and sensitivity, presented in Receiver Operator Characteristics curve format, for detection of referable retinopathy as reported in the studies listed in tables 2 and 4. The point labelled by an asterisk is for a study which reported results only for referable retinopathy and ungradable cases grouped together [20].

The economics of automated grading

Modelling of complete grading systems was described in two economic analyses, linked to two of the efficacy studies for the Aberdeen system. These modelled a system consisting of a combination of automated and manual grading and a fully manual system. Scotland *et al* 2007 [23] used the efficacy figures reported in Philip *et al* 2007 [22] and predicted that replacement of manual disease / no disease grading by an automated system would result in a cost saving of around £200,000 when scaled to the population of Scotland [18]. The lower specificity of automated compared to manual grading would mean that a greater number of full disease gradings must be performed. Nonetheless the overall workload would be reduced.

The efficacy results presented in Fleming *et al* 2010 [24] were also subject to economic analysis in Scotland *et al* 2010 [25]. This concluded that inclusion of exudate and haemorrhage detection produces a more cost-effective algorithm compared to microaneurysm detection alone (£68 additional cost per additional case of

referable retinopathy detected). It also found that automated grading would significantly reduce grading costs by £212,695 in Scotland.

Other roles for automated grading

The efficacy results for automated grading are most easily interpreted in terms of a disease / no disease grading role. This section looks briefly at other possible roles.

Full disease grading: A **f**ull-disease grader must operate with specificity above what can be achieved by automated grading. Therefore, this role does not seem to be achievable by current systems.

Verification grading: An automated system working in parallel with and as a verification of manual grading has been suggested as the role of automated image analysis in screening for breast cancer [34]. A similar role may be considered useful for automated retinopathy grading. Given that full-disease grading is not practical by existing automated grading systems, this would have to be at a disease / no disease level. There would be more discrepancies than there would be from manual graders verifying each other since automated grading produces more false positive results than manual grading.

Quality assurance: Another possible role for automated grading is quality assurance since this is a significant component of the work involved in the maintenance of a screening programme [35]. Internal quality assurance requires checking samples of grading results which are large enough to detect suboptimal performance by graders. Automated grading in this role might reduce, but probably not eliminate, the burden of manual quality assurance grading. Before automated grading is used in this role, it would be essential to know the ability of automated grading to detect images which were manual grading failures; on average, these are probably more difficult than routine cases. Such a test would be easy to perform but does not seem to have been reported.

Manual grading assistant: To assess the performance of automated grading as an assistant to manual grading, it would be necessary to test the performance of manual graders with and without this assistance. Does automated lesion detection assist manual graders or does giving attention to the additional information provided by automated grading increase their work? Studies on this issue do not seem to have been reported.

CONCLUSIONS

This review has evaluated the evidence on whether automated grading systems could be used effectively in diabetic retinopathy screening.

Automated grading using detectors of multiple types of dot lesions shows increased performance compared to automated grading using a detector of a single lesion type. However, this improvement may be restricted to the detection of maculopathy; microaneurysm detection is capable of detecting 100% of cases of proliferative retinopathy.

The most likely role for automated grading appears to be disease / no disease grading.

Automated grading has been shown to have very high detection rates for referable retinopathy and for ungradable cases. For the Aberdeen system, the detection rate for proliferative retinopathy approaches 100%.

There is no evidence that the replacement of manual grading by automated grading would reduce the detection rates for disease. Studies on manual grading show that detection rates are well below 100%.

The specificity of automated grading is lower than for manual grading. Despite this, the specificity of automated grading is sufficiently high to provide a substantial workload reduction to manual grading from which cost savings would be expected.

From economic modelling based on large effectiveness studies, it appears that automated disease / no disease grading would result in savings both to running costs and to workload.

CONFLICT OF INTEREST

Commercial implementation associated with some of the referenced work may provide some remuneration for the University of Aberdeen, NHS Grampian, Alan Fleming, John Olson and Peter Sharp. Funding for Alan Fleming was provided by Medalytix Ltd under an agreement with Scottish Health Innovations Limited. Funder sources for all authors made no input into the study design, the collection, analysis and interpretation of data, the writing of the report, nor the decision to submit the paper for publication.

REFERENCES

[1] Scanlon PH. The English national screening programme for sight-threatening diabetic retinopathy. J Med Screen 2008;15:1-4

[2] Sharp PF, Olson J, Strachan F, Hipwell J, Ludbrook A, O'Donnell M, Wallace S, Goatman KA, Grant A, Waugh N, McHardy, K and Forrester JV. The value of digital imaging in diabetic retinopathy, Health Technology Assessment 2003; Vol 7:number 30

[3] Leese GP, Morris AD, Swaminathan K, Petrie JR, Sinharay R, Ellingford A, Taylor A, Jung RT, Newton RW, Ellis JD. Implementation of national diabetes retinal screening programme is associated with a lower proportion of patients referred to ophthalmology. Diabetic Med 2005, 22:8:1112-1115

[4] European conference on screening for diabetic retinopathy in Europe May 2006. http://www.drscreening2005.org.uk/conference_report.doc (accessed 19 April 2011).

[5] Teng T, Lefley M, Claremont D. Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. Med Biol Eng Comput 2002,40:2–13

[6] Patton N, Aslam TM, MacGillivray T, Deary IJ, Dhillon B, Eikelboom RH, Yogesana K, Constable IJ. Retinal image analysis: Concepts, applications and potential. Prog Retin Eye Res 2006, 25:99–127

[7] Winder RJ, Morrow PJ, McRitchie IN, Bailie JR, Hart PM. Algorithms for digital image processing in diabetic retinopathy. Comput Med Imag Grap 2009, 33:608–22

[8] Abràmoff MD, Garvin M, Sonka M. Retinal Imaging and Image Analysis. IEEE Reviews in Biomedical Engineering, 3, 169-208, 2010

[9] Wilkinson CP, Ferris FL III, Klein RE, Lee PP, Agardh CD, Davis M, Dills D, Kampik A, Pararajasegaram R, Verdaguer JT. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 2003;101:1677–82.

[10] UK National Screening Committee, Essential Elements in Developing a Diabetic Retinopathy Screening Programme, Workbook 4.3, June 2009, www.retinalscreening.nhs.uk (accessed 3rd May 2011).

[11] Scottish Diabetic Retinopathy Screening Collaborative. Scottish diabetic retinopathy grading scheme. 2007 v1.1. http://www.ndrs.scot.nhs.uk/ClinGrp/Docs/Grading%20;Scheme%202007%20v1 (accessed 31st May 2011).

[12] Hipwell JH, Strachan F, Olson JA, McHardy KC, Sharp PF, Forrester JV Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. Diabetic Med 2000, 17;8:588-94

[13] Usher D, Dumskyj M, Himaga M, Williamson TH, Nussey S, Boyce J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. Diabetic Med 2004, 21;1:84-90

[14] Hansen AB, Hartvig NV, Jensen MS, Borch-Johnsen K, Lund-Andersen H, Larsen M. Diabetic retinopathy screening using digital non-mydriatic fundus photography and automated image analysis, Acta Ophthalmologica Scandinavica 2004, 82;6: 666-72

[15] Larsen M, Gondolf T, Godt J, Jensen MS, Hartvig NV, Lund-Andersen H, Larsen N, Assessment of automated screening for treatment-requiring diabetic retinopathy. Curr Eye Res 2007, 32:4:331-6

[16] Acharya UR, Chua, CK, Ng EYK, Yu W, Chee C. Application of higher order spectra for the identification of diabetes retinopathy stages. J Med Syst 2008, 32;6: 481-8

[17] Acharya UR, Lim CM, Ng EYK, Chee C, Tamura T, Computer-based detection of diabetes retinopathy stages using digital fundus images. P I Mech Eng H 2009, 223:H5:545-53

[18] Quellec G, Russell SR, Abramoff MD, Optimal Filter Framework for Automated Instantaneous Detection of Lesions in Retinal Images. IEEE T Med Imag 2011:30; 2:523-33

[19] Abràmoff MD, Niemeijer M, Suttorp-Schulten MSA, Viergever MA, Russell SR, van Ginneken B. Evaluation of a System for Automatic Detection of Diabetic Retinopathy From Color Fundus Photographs in a Large Population of Patients With Diabetes. Diabetes Care 2008 31(2):193-198
[20] Niemeijer M, Abramoff MD, van Ginneken B et al. Information Fusion for Diabetic Retinopathy CAD in Digital Color Fundus Photographs, IEEE T Med Imag 2009, 28:775-85
[21] Abràmoff MD, Reinhardt JM, Russell SR, Folk JC, Mahajan VB, Niemeijer M, Quellec G, Automated Early Detection of Diabetic Retinopathy. Ophthalmology 2010, 117:1147-54
[22] Philip S, Fleming AD, Goatman KA, Fonseca S, Mcnamee P, Scotland GS, Prescott GJ, Sharp PF, Olson JA. The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme Br J Ophthalmol 2007,91,1512-7
[23] Scotland GS, McNamee P, Philip S, Fleming AD, Goatman KA, Prescott GJ, Fonseca S, Sharp PF, Olson JA. Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland, Br J Ophthalmol,2007, 91:1518-23
[24] Fleming AD, Goatman KA, Philip S, Williams GJ, Prescott GJ, Scotland GS, McNamee P, Leese GP, Wykes WN, Sharp PF, Olson JA. The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy, Br J Ophthalmol 2010,94:706-711
[25] Scotland GS, McNamee P, Fleming AD, Goatman KA, Philip S, Prescott GJ, Sharp PF, Williams GJ, Wykes W, Leese GP, and Olson JA. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol 2010,94:712-9
[26] Fleming, AD; Goatman, KA; Philip, S, Prescott GJ, Sharp PF, Olson JA. Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts, Br J Ophthalmol 2010;94:1606-10
[27] Benbassat J, Polak BCP. Reliability of screening methods for diabetic retinopathy. Diabet Med 2009,26:783-90
[28] Liesenfeld B, Kohner E, Piehlmeier W, Kluthe S, Aldington S, Porta M et al. A telemedical approach to the screening of diabetic retinopathy: digital fundus photography. Diabetes Care 2000; 23: 345–348.
[29] Olson JA, Strachan FM, Hipwell JH, Goatman KA, McHardy KC, Forrester JV, Sharp PF, A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy, Diabetic Med 2003;20:7,528-34
[30] Stellingwerf C, Hardus PL, Hooymans JM. Assessing diabetic retinopathy using two-field digital photography and the influence of JPEG compression. Doc Ophthalmol 2004; 108: 203-9.
[31] Arun CS, Young D, Batey D, Shotton M, Mitchie D, Stannard KP et al. Establishing ongoing quality assurance in a retinal screening programme. Diabetic Med 2006; 23: 629–34.
[32] Ruamviboonsuk P, Teerasuwanajak K, Tiensuwan M, Yuttitham K, Thai Screening for Diabetic Retinopathy Study Group. Interobserver Agreement in the Interpretation of Single-Field Digital Fundus Images for Diabetic Retinopathy Screening. Ophthalmology 2006;113:826–32
[33] Scanlon PH, Malhotra R, Thomas G, Foy C, Kirkpatrick JN, Lewis-Barned N, Harney B, Aldington SJ. The effectiveness of screening for diabetic retinopathy by digital imaging photography and technician ophthalmoscopy. Diabetic Med 2003;20:467–74
[34] Gilbert FJ, Astley SM, Gillan MGC, Agbaje OF, Wallis MG, James J, Boggis CRM, Duffy SW. Single reading with computer-aided detection for screening mammography. New Engl J Med 2008, 359:1675-84
[35] Nagi DK, Gosden C, Walton C, Winocour PH, Turner B, Williams R, James J, Holt RIG. A national survey of the current state of screening services for diabetic retinopathy: ABCD-Diabetes UK survey of specialist diabetes services 2006. Diabetic Med 2009, 26:1301-5