Computers in Biology and Medicine 88 (2017) 100–109

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A tool for automated diabetic retinopathy pre-screening based on retinal image computer analysis

Manuel E. Gegundez-Arias [b], Diego Marin [a,*], Beatriz Ponte [c], Fatima Alvarez [d], Javier Garrido [e], Carlos Ortega [e], Manuel J. Vasallo [a], Jose M. Bravo [a]

[a] Department of Electronic, Computer Science and Automatic Engineering, University of Huelva, Spain
[b] Department of Mathematics, University of Huelva, Spain
[c] Virgen de Macarena University Hospital of Seville, Andalusian Health Service, Spain
[d] Juan Ramon Jimenez Hospital of Huelva, Andalusian Health Service, Spain
[e] North Area of Sanitary Management of Cordoba, Andalusian Health Service, Spain

ARTICLE INFO

Keywords:
Diabetic retinopathy
Early detection system
Automated screening
Retinal image processing
Computer-aided diagnosis

ABSTRACT

Aim: This paper presents a methodology and first results of an automatic detection system of first signs of Diabetic Retinopathy (*DR*) in fundus images, developed for the Health Ministry of the Andalusian Regional Government (Spain).

Material and methods: The system detects the presence of microaneurysms and haemorrhages in retinography by means of techniques of digital image processing and supervised classification. Evaluation was conducted on 1058 images of 529 diabetic patients at risk of presenting evidence of *DR* (an image of each eye is provided). To this end, a ground-truth diagnosis was created based on gradations performed by 3 independent ophthalmology specialists.

Results: The comparison between the diagnosis provided by the system and the reference clinical diagnosis shows that the system can work at a level of sensitivity that is similar to that achieved by experts (0.9380 sensitivity per patient against 0.9416 sensitivity of several specialists). False negatives have proven to be mild cases. Moreover, while the specificity of the system is significantly lower than that of human graders (0.5098), it is high enough to screen more than half of the patients unaffected by the disease.

Conclusion: Results are promising in integrating this system in *DR* screening programmes. At an early stage, the system could act as a pre-screening system, by screening healthy patients (with no obvious signs of *DR*) and identifying only those presenting signs of the disease.

1. Introduction

Diabetes Mellitus (*DM*) is a health disorder in which a malfunction in the secretion of insulin causes hyperglycaemia. It is also a disorder of social importance due to its high incidence - *DM* is estimated to affect 592 million people in 2035 [1]. In the long term, it can cause microvascular complications that affect the retina, resulting in Diabetic Retinopathy (*DR*), which is the leading cause of blindness in active population [2]). There are three forms of *DR*: 1) Macular edema, i.e. the build-up and spill of fluid in the macula; (2) weakening and changes in retinal blood vessels, resulting in microaneurysms, haemorrhages, malformation and vascular tortuosity (Non-Proliferative Diabetic Retinopathy, *NPDR*) that can subsequently cause an abnormal growth of retinal blood vessels (Proliferative Diabetic Retinopathy, *PDR*); and 3) retinal capillary closure, a form of vascular change considered as a potential blinding complication in diabetes [3].

DR severity is well known to have a close association with the duration of *DM*. However, most diabetic patients have no symptoms in early retinopathy stages, when treatments to prevent vision loss (laser therapy and/or vitrectomy and/or intraocular pharmacological intervention) are effective. By analysing the cumulative evolution of *DR* throughout 25 years, the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) found that patients who were diagnosed most recently with a similar duration of diabetes had a lower prevalence of *PDR* regardless of glycosylated haemoglobin levels, blood pressure levels and the presence of proteinuria [4]. Among individuals with type 1 diabetes, *DR* is not

* Corresponding author. Escuela Técnica Superior de Ingeniería, Campus de La Rábida, Universidad de Huelva, Carretera Palos de La Frontera s/n, 21819, Palos de La Frontera, Huelva, Spain.
E-mail address: diego.marin@diesia.uhu.es (D. Marin).

http://dx.doi.org/10.1016/j.compbiomed.2017.07.007
Received 27 March 2017; Received in revised form 23 June 2017; Accepted 6 July 2017

often present at the time of diagnosis. However, as many as 21% [5] (nearly 40% in UKPDS [6] of individuals with type 2 diabetes already have retinopathy by the time diabetes is diagnosed). This difference is the basis for the American Diabetes Association's recommendation of initial screening within 5 years after the onset of type 1 diabetes and shortly after diagnosis for type 2 diabetes. In case no evidence of retinopathy is found in one or more eye exams, examination every 2 years may then be considered. If *DR* is present, subsequent examinations for type 1 and type 2 diabetic patients should be repeated annually by an ophthalmologist or optometrist. If the retinopathy is progressing or sight threatening, more frequent examinations will then be required [7]. Therefore, systematic screening programmes in the diabetes community are needed for early *DR* detection, and to reduce blindness in diabetic patients.

Nowadays, *DR* screening programmes are performed by fundus imaging analysis. While this procedure does not replace the examination performed by an ophthalmologist, it does provide reliable information as to whether a patient is beginning to be affected by the disease [8,9]. However, due to the high incidence of *DM*, the number of daily examinations required for the proper control and follow-up of patients is very high. For instance, based on 2035 estimations, 3.2 million eyes would need to be examined every day. In this context, integrating systems that analyse fundus images and automatically detect signs of *DR* will ensure sustainability and quality in the prevention of the disease. This is why there has been a great interest in the development of systems that detect early ophthalmic signs of illness in recent years (see Refs. [10–13]).

Some prototypes of such systems have already been presented [14–19,21]. They are mainly evaluated by measuring sensitivity and specificity of automated retinopathy detection, regarding clinical diagnoses provided by ophthalmologic specialists. The system proposed by Usher et al. [14] classified a set of 1406 fundus images from 703 diabetic patients as abnormal or normal (affected by *DR* or not) according to the presence or absence of microaneurysms, haemorrhages and exudates. At the 0.890 sensitivity setting, the system obtained a specificity of 0.624. Philip et al. [15] presented another disease/no disease grading system based on the detection of red lesions (microaneurysms and haemorrhages). The sensitivity and specificity of automated grading (detecting technical failures or any retinopathy) of a set of 14,406 images from 6722 patients attending a *DR* screening programme were 0.862 and 0.768, respectively. Abramoff et al. [21] evaluated the performance of a *DR*-related lesion detection system (red and bright lesions) built from published algorithms in a large screening population (10,000 patients from the EyeCheck diabetic screening project [22]), reporting a sensitivity of 0.84 and a specificity of 0.64. Niemeijer et al. [16] extended the evaluation to 15,000 eye exams (60,000 images) to propose information fusion methods and evaluate their effect on the performance of a complete comprehensive automatic *DR* screening system. The best performing fusion method obtained an area under the receiver operator characteristic curve of 0.881, reaching a sensitivity of 0.929 at a fixed specificity of 0.6. Dupas et al. [17] evaluated an algorithm for the automated detection of microaneurysms on 761 fundus images included in the Messidor database [23]. The sensitivity and specificity of the algorithm to detect *DR* were 0.839 and 0.727, respectively. On this same database, Antal et al. [19] evaluated an ensemble-based *DR* screening method based on microaneurysm and exudate detection, reporting a sensitivity of 0.90 and a specificity of 0.91. Abramoff et al. [18] validated a computer *DR* detection programme on 874 people (1748 images) with diabetes at risk for *DR*. The programme is based on published methodologies for image quality assessment and detection of microaneurysms and haemorrhages, exudates and cotton wool spots, as well as, irregular lesions. Authors reported a sensitivity of 0.968 and specificity of 0.594, none of the 6 false-negative results had sight-threatening diseases. Finally, deep learning techniques are being used recently to design *DR*-detection algorithms, with very promising results. For example, in Ref. [20], a convolutional neural network was used to detect *DR* in EyePACS-1 (9963 images, 4997 patients) and Messidor-2 (1748 images, 874 patients) data sets. Sensitivity/specificity results were: 90.3%/98.1% and 96.1%/93.9%, respectively. Results are shown in Table 1.

This paper presents a new *DR* detection system based on the detection of microaneurysms and haemorrhages. The system is funded by the Health Ministry of the Andalusian Regional Government (Spain). The structure of the paper is as follows: The material used to evaluate the *DR* detection system will be described in the next section, together with the procedure used to design it. The methodology and the generation of automated diagnoses are explained and illustrated in Section 3, while the results from comparing system and human diagnoses are presented in Section 4. Finally, the authors' conclusions and discussion conclude this paper.

2. Materials and clinical diagnoses

The system has been evaluated using eye-fundus color images from the Messidor database [23]. Specifically, the study used two macula-centered retinographies of the posterior pole, taken from 529 patients with diabetes at risk for *DR* (1058 images, one per eye in each patient). Images were independently diagnosed by three ophthalmology specialists. This allowed for a consensus diagnosis, which is reliable for further evaluation of the system. Next, the procedure used to come up with the reference clinical diagnosis is described.

2.1. Generation of ophthalmologists' diagnoses

Ophthalmology specialists from the Andalusian Health Service (Spain) manually and independently mark microaneurysms and haemorrhages found in each image. These specialists work at the following medical centres: the North Area of Sanitary Management of Cordoba (NASMC), Virgen de Macarena University Hospital of Seville (VMUHS) and Juan Ramon Jimenez Hospital of Huelva (JRJHH). Their marks are used to generate three diagnoses per image, according to the grading criteria described in Table 2 (designed in Ref. [23] based on recommendations suggested by Refs. [24–27]). Each image diagnosis is used to generate patient diagnoses, by assigning the maximum gradations of the two fundus images available.

2.2. Generation of ground-truth clinical diagnoses

Diagnoses generated by each medical centre –NASMC, VMUHS and JRJHH– are used to generate a reference clinical diagnosis per image and patient. To do so, the consistency of diagnoses from the different experts needs to be checked. Inter-rater reliability metrics calculated for this purpose show a high agreement among diagnoses. The intra-class correlation [28] value (0.844 with a 95% confidence interval, *CI*, of $0.820 - 0.865$) shows excellent agreement according to [29], while Fleiss' kappa (0.603 with a 95% *CI* of $0.595 - 0.612$) indicates substantial agreement according to [30] (detailed results can be seen in Ref. [31]). Thus, to evaluate the proposed system, the following reference clinical diagnoses were used:

- Ground-truth diagnosis per image, generated by applying consensus criteria to the diagnoses provided by specialists for each of the 1058 fundus images. The consensus diagnosis was obtained by assigning

Table 1
State of art: performance of *DR*-detection systems.

Method	Images/Database	Lesions	Sens/Spec
Usher et al. [14]	1406/Private	$\mu A, H, Ex$	0.890/0.624
Philip et al. [15]	14,406/Private	$\mu A, H$	0.862/0.768
Niemeijer et al. [16]	60,000/EyeCheck	$\mu A, H, Ex, \ldots$	0.929/0.600
Dupas et al. [17]	761/Messidor	μA	0.839/0.727
Abramoff et al. [18]	1748/Messidor-2	$\mu A, H, Ex, \ldots$	0.968/0.594
Antal et al. [19]	1200/Messidor	$\mu A, Ex$	0.900/0.910
Gulshan et al. [20]	9963/EyePACS-1	–	0.903/0.981
	1748/Messidor-2	–	0.961/0.939