LITERATURE SURVEY ON DEEP LEARNING FUNDUS IMAGE OF DIABETIC RETINOPATHY

TEAMID:PNT2022TMID33204

Abstract

Diabetes is a globally prevalent disease that can cause visible micro vascular complications such as diabetic retinopathy and macular in the human eye retina, the images of which are today used for manual disease screening and diagnosis. This intensive task could greatly benefit from automatic detection using deep learning technique. Here we present a deep learning system that identifies referable diabetic retinopathy comparably or better than presented in the previous studies, although we use only a small fraction of images (<1/4) in training but are aided with higher image resolutions. We also provide novel results for five different screening and clinical grading systems for diabetic retinopathy and macular classification including state-of-the-art results for accurately classifying images according to clinical five-grade diabetic retinopathy and for the first time for the four-grade diabetic macular scales. These results suggest that a deep learning system could increase the cost-effectiveness of screening and diagnosis, while attaining higher than recommended performance, and that the system could be applied in clinical examinations requiring finer grading.

Introduction

Diabetic retinopathy is the most common micro vascular complication in diabetes, for the screening of which the retinal imaging is the most widely used method due to its high sensitivity in detecting retinopathy. The evaluation of the severity and degree of retinopathy associated with a person having diabetes is currently performed by medical experts based on the fundus or retinal images of the patient's eyes. As the number of patients with diabetes is rapidly increasing, the number of retinal images produced by the screening programmes will also increase, which in turn introduces a large intensive burden on the medical experts as well as cost to the healthcare services. This could be alleviated with an automated system either as support for medical experts' work or as full diagnosis tool. There are two recent studies that have investigated the use of deep learning systems in automated detection of diabetic

retinopathy. Both show that an automated system, based on the deep learning artificial neural network approach, can achieve high sensitivity with high specificity in detecting the referable diabetic retinopathy, defined as moderate or worse diabetic retinopathy. There are also other referable eye complications that have recently been investigated with this approach, such as diabetic macular and possible glaucoma and age-related macular degeneration.

Methods

Original fundus image dataset

The research of present study was done in collaboration with Digifundus Ltd, an ISO 9001:2015 certified provider of diabetic retinopathy screening and monitoring services in Finland. Digifundus Ltd provided a non-open, anonymized retinal image dataset of patients with diabetes, including 41122 graded retinal colour images from 14624 patients. The images were taken with Canon CR2 retinal camera after inducing with tropicamide 5 mg/ml eye drops. Two 45 degree colour fundus photographs, centered on fovea and optic disc were taken from the patient's both eyes. The output images were of variable resolutions, ranging from 3888 × 2592 to 5184 × 3456 pixels.

The present study is a methodological study with anonymized medical data and without any intervention in the integrity of a person such as contact with a person. In Finnish law this is not considered as a medical study requiring approval by an ethics committee or a written consent of a person.

Retinal image grading systems and gradability

Each of the retinal images had been graded with respect to three different criteria, (i) diabetic retinopathy, (ii) macular edema, and (iii) gradability. Images are graded with the proposed international clinical diabetic retinopathy and macular edema disease severity scales denoted later as PIRC and PIMEC, respectively. The image gradability is a two-stage system, which considers an image to be either gradable or not. All personnel participating in retinopathy assessment had over 10 years' experience in diabetic retinopathy grading. Retinal images with no lesions or mild diabetic lesions were graded by an optometrist and an

M.D. trained for retinopathy grading. All images with moderate or worse changes were graded by two ophthalmologist both with more than 10 years of experience in grading fundus images. If there was a disagreement in grading, such an image was not included in this study.

PIRC and PIMEC grades were further used to obtain additional three types of grading systems: (i) a binary system of *nonreferable/referable diabetic retinopathy* (NRDR/RDR), (ii) a binary system of *nonreferable/referable diabetic macular edema* (NRDME/RDME), and (iii) three-class system of ungradable/NRDR/RDR. The NRDR/RDR system considers the cases with no diabetic retinopathy and mild diabetic retinopathy as nonreferable diabetic retinopathy, and the cases with moderate or worse diabetic retinopathy as referable diabetic retinopathy. This system has been used in recent works investigating automated detection of diabetic retinopathy. The NRDME/RDME system here is defined such that the absence of macular edema is defined as nonreferable diabetic macular edema and any level of macular edema as referable diabetic macular edema. Note that only the gradable images were graded for diabetic retinopathy and macular edema. Ungradable images were included in a single task, in combination with referable diabetic retinopathy classification, which constitutes the grading system QRDR, in which each image is considered to be either ungradable, depicting nonreferable diabetic retinopathy, or depicting referable diabetic retinopathy (ungradable/NRDR/RDR).

Image preprocessing and dataset division

In the model training and subsequent primary validation, we used preprocessed versions of the original images. The preprocessing consisted of image cropping followed by resizing. Each image was cropped to a square shape which included the most tightly contained circular area of fundus. The procedure removed most of the black borders and all of the patient related annotations from the image data. Each of the cropped images were then resized to five different standard input image sizes of 256×256, 299×299, 512×512, 1024×1024, and 2095×2095 pixels. The largest image size was the smallest native resolution of the retinal cameras after the preprocessing steps. Here the creation of multiple resolutions was done for the purposes of analyzing the effect of the input image resolution on the classification performance.

The obtained processed datasets were divided into three

sets: *training*, *tuning*, and *primary validation* set in the 70%, 10% and 20% proportions of the whole image dataset, respectively, separately for each of the grading systems used in the experiments. In the division per a particular grading system, the different sets were to have similar grade distributions, and that the dataset data per patient to not reside in multiple but only in one of the three sets (of training, tuning, and primary validation), in order to prevent the possibility of obtaining over-optimistic results due to data memorization. Table 1 shows the statistics of the resulting divisions that were used in the experiments. Note that the grade distributions across the different sets were similar, with respect to each grading system, for example, when we consider the NRDR/RDR-system, the proportion of images associated with referable diabetic retinopathy in the training, tuning and primary validation set 44%, 43.9% and 43.4%, respectively.

Results

In the binary classification tasks, i.e. NRDR/RDR and NRDME/RDME, our algorithm achieved the best results using the largest 2095 × 2095 pixels input image size. In the NRDR/RDR classification on our primary validation set having 7118 images, our algorithm achieved the sensitivity of 0.896 (with 95% CI: 0.885–0.907) and specificity 0.974 (with 95% CI: 0.969–979) and AUC of 0.987 (with 95% CI: 0.984–0.989). Our model performance was evaluated at the operating point where the tuning set achieved 0.900 sensitivity, in a similar manner to Ting *et al.* while Gulshan *et al.* had two operating points namely at a high specificity (0.980) point and at a high sensitivity (0.975) point. In Table we present the AUC values of our model, along with the AUC values reported by Gulshan *et al.* and Ting *et al.* The Two other recent studies, Krause *et al.* and Guan *et al* also explored the NRDR/RDR classification, but as they do not report res ults close to the 0.900 sensitivity point,

References

Marshall, S. M. & Flyvbjerg, A. Prevention and early detection of vascular complications of diabetes. *BMJ.* **333**(7566), 475–480 (2006).

Hutchinson, A. *et al.* Effectiveness of screening and monitoring tests for diabetic retinopathy–a systematic review. *Diabet Med.* **17**(7), 495–506 (2000).

Taylor, R. & Batey, D. *Handbook of Retinal Screening in Diabetes: Diagnosis and Management*. Wiley (2012).

Gulshan, V. *et al.* Development and Validation of a Deep Learning

Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* **316**(22), 2402–2410 (2016).

Ting, D. S. W. *et al.* Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA.* **318**(22), 2211–2223 (2017).

Wilkinson, C. P. *et al.* Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology.* **110**(9), 1677–1682 (2003).

Summanen, P. *et al.* Update on Current Care Guideline: Diabetic retinopathy. *Duodecim.* **131**(9), 893–894 (2015)

Krause, J. *et al.* Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. *Ophthalmology.* **125**(8), 1264–1272 (2018).

Guan, M. Y., Gulshan, V., Dai, A. M. & Hinton, G. E. Who Said What: Modeling Individual Labelers Improves Classification. *arXiv e-prints.*, https://ui.adsabs.harvard.edu/\#abs/2017arXiv170308774G. (Accessed March 01, 2017).

Wang, F., Casalino, L. P. & Khullar, D. Deep Learning in Medicine-Promise, Progress, and Challenges. *JAMA Intern Med.*

Medical Research Act, https://www.finlex.fi/en/laki/kaannokset/1999/en19990488_20100794.pdf

Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning. MIT Press

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv e-prints.* 2015, https://ui.adsabs.harvard.edu/\#abs/2015arXiv151200567S

Russakovsky, O. *et al.* ImageNet Large Scale Visual Recognition Challenge.

Decencière, E. *et al. Feedback on a publicly distributed image database: The Messidor database.* Vol 02014.