**The use of fundus images to discriminate prediabetes and type 2 diabetes. The Maastricht Study**

Friso G. Heslinga MSc1, Tos T.J.M. Berendschot PhD1,2, Miranda T. Schram PhD3, Coen D.A. Stehouwer MD PhD3, Marleen J. van Greevenbroek PhD3, Josien P.W. Pluim PhD1, A.J.H.M. Houben PhD3,\*, Mitko Veta PhD1,\*

1Eindhoven University of Technology, Department of Biomedical Engineering, the Netherlands

2Maastricht University Medical Center+, University Eye Clinic Maastricht, the Netherlands

3Maastricht University Medical Center+, Department of Internal Medicine and CARIM, the

Netherlands

\*contributed equally

Correspondence author:

Mitko Veta

Email: m.veta@tue.nl

Address: Department of Biomedical Engineering, Eindhoven University of Technology, PO Box 513 5600 MB Eindhoven, the Netherlands

Running title: (pre)diabetes detection using fundus images

Word count: 2,589 / 4,000

Number of figures: 3

Number of tables: 2

Number of supplementary: 1

Number of suppl. e-figures: 2

**ABSTRACT (248 of 250 words)**

**Objective:** To study to what extent fundus images can be used to discriminate between individuals with normal glucose metabolism, prediabetes, and type 2 diabetes. In addition, the discriminative value of fundus images was compared with typical risk factors for type 2 diabetes.

**Research Design and Methods:** A deep learning model was developed to determine a glucose metabolism score (GMScore) from fundus images using data from The Maastricht Study. The discriminative power of this GMscore for classifying individuals with type 2 diabetes versus normal glucose metabolism was assessed for a hold-out dataset by the area under receiver operating characteristic (AUROC) curve. For comparison, AUROCs for risk factors such as age, waist circumference, and family history were also calculated and combined with the GMScore to evaluate the additional value of the fundus images.

**Results:** The GMScore based on the fundus images obtained an AUROC of 0.757 (95% CI 0.731 – 0.783), which is higher than 5 out of 6 risk factors for diabetes. Only waist circumference resulted in a higher AUROC. When information from the fundus images was combined with other risk factors, the AUROC increased to 0.895 (95% 0.878 - 0.912). Moreover, prediabetes individuals were found to have a distinct GMScore distribution, approximately half-way between normal and type 2 diabetes individuals.

**Conclusions:** Fundus images are informative for discriminating individuals with prediabetes and type 2 diabetes from those with normal glucose metabolism. When combined with other typical risk factors, fundus imaging can contribute modestly to screening for early-stage type 2 diabetes.

**INTRODUCTION**

Still half of all people living with diabetes worldwide are undiagnosed (1) while it is well known that late detection can lead to long-term complications, including blindness (2) and cardiovascular disease (3). Research efforts into non-invasive screening techniques are ongoing, using risk factors such as waist circumference (4), family history (5) and a combination of factors (6,7).

Type 2 diabetes affects the cardiovascular system and early changes in the retinal vascular tree are associated with type 2 diabetes, such as vessel caliber (8–11) and vascular tortuosity (12,13). Fundus photography allows for non-invasive visualization of the retina and automated retinal image analysis has become increasingly popular (14,15). Specifically, the use of deep learning (16) has been promising, and several studies have shown excellent results for detection of diabetic retinopathy (17–19).

So far, limited research has been done on the value of deep learning on fundus images for early type 2 diabetes detection. This could be due to the challenging nature, as early signs of type 2 diabetes are much more subtle than retinopathy, but also due to the lack of a large good-quality data set.

Since 2010, a large set of fundus images from individuals with (pre)diabetes has been collected as part of The Maastricht Study (20). Here, we aim to utilize these fundus images to develop an automated method to discriminate between individuals with normal glucose metabolism and type 2 diabetes. We also compare the discriminative power of the fundus images with that of some typical risk factors such as age, sex, and family history. In addition, we investigate to what extent fundus images can be used to discriminate between individuals with normal glucose metabolism and individuals with prediabetes.

**RESEARCH DESIGN AND METHODS**

*Data*

We used data from The Maastricht Study, an observational prospective population-based cohort study. The rationale and methodology have been described previously (20). In brief, the study focuses on the etiology, pathophysiology, complications, and comorbidities of type 2 diabetes mellitus and is characterized by an extensive phenotyping approach. Eligible for participation were all individuals aged between 40 and 75 years and living in the southern part of the Netherlands. Participants were recruited through mass media campaigns and from the municipal registries and the regional Diabetes Patient Registry via mailings. Recruitment was stratified according to known type 2 diabetes status, with an oversampling of individuals with type 2 diabetes, for reasons of efficiency. The present report includes cross-sectional data from the first 7,689 participants, **from whom fundus photography was available for 6,539 participants**, and who completed the baseline survey between November 2010 and December 2017. The examinations of each participant were performed within a time window of three months. The study has been approved by the institutional medical ethical committee (NL31329.068.10) and the Minister of Health, Welfare and Sports of the Netherlands (Permit 131088-105234-PG). All participants gave written informed consent.

A total of 58,722 color fundus images (FF450; Carl Zeiss AG, Jena, Germany) from 6,539 individuals were initially included. The data set comprised images from both eyes, fixated on the optic disc, macula, or periphery. Fundus image quality was assessed automatically, and low-contrast images (*N* = 12,000) were excluded (Supplementary eFigure 1). In addition, we excluded individuals with other types of diabetes than type 2 (*N* = 41), resulting in a final set of 46,371 images from 6,453 individuals.

*(Pre)diabetes classification*

Glucose metabolism status was based on the World Health Organization definitions for fasting glucose, 2-hour oral glucose tolerance test (21), and use of glucose-lowering medication. Consequently, we distinguished individuals with normal glucose metabolism, type 2 diabetes, and prediabetes. Prediabetes was defined as impaired fasting glucose and/or impaired glucose tolerance. Individual-level glucose metabolism statuses were extended to the fundus images, meaning that every fundus image from an individual with e.g. type 2 diabetes was labeled as type 2 diabetes, independent of the actual information in the image.

*Image preprocessing and experimental setup*

All fundus images were cropped, resized, and preprocessed (22) to increase contrast for more efficient training of the image analysis algorithms (Supplementary eFigure 2). We split the data randomly on an individual level into a set for model development (N = 28,153 images) and a hold-out set for final validation (N = 14,476 images). Additionally, a separate set (N = 3,742 images) was created for a matching cohort experiment. This set includes 275 individuals that were newly identified as having type 2 diabetes during The Maastricht Study, matched with non-diabetic individuals based on typical type 2 diabetes risk factors.

*Deep learning model*

Our image analysis algorithm is based on deep learning (16). A detailed description of the development is provided in Supplementary 1. In short, a convolutional neural network with an EfficientNet B4 (23) architecture was trained to classify the glucose metabolism status of individual fundus images. During training, only normal glucose metabolism cases and known type 2 diabetes cases were used. Prediabetes cases were not used at this point. The output of the model is a glucose metabolism score (GMScore) in the range 0 to 1 where an output value closer to 0 represent normal glucose metabolism, and close to 1 represents type 2 diabetes.

*Evaluation*

For evaluation, the fundus images of the validation set were processed by the trained network to obtain glucose metabolism scores. For an individual-level model GMScore, image-level scores were combined by averaging over all fundus images of the left and right eye. Our primary evaluation was the discriminative power of the GMScore prediction for classifying individuals with normal glucose metabolism versus type 2 diabetes individuals, measured by the area under receiver operating characteristic (AUROC) curve. For comparison, risk factors for type 2 diabetes (sex, age, waist circumference, smoking, hypertension, and family history) were used to train logistic regression classifiers to discriminate normal glucose metabolism from type 2 diabetes individuals. Smoking was divided into 3 categories (non-smoker, former, and current), based on self-reported data. Hypertension was defined as an office blood pressure greater 140/90 mm Hg or use of blood pressure lowering medication. Family history represents self-reported data about a first or second degree relative with diabetes. Ninety-five percent confidence intervals (CI) for the AUROC were obtained using 1,000 bootstrap samples. Prediabetes individuals in the validation set were also processed and glucose metabolism scores for this group were compared with normal glucose metabolism individuals. All p-values reported are based on the Welch's unequal variances two-sided t-test and describe the probability that compared groups are similar.

*Matching cohort experiment*

Risk factors for diabetes such as age and sex can, to some extent, be extracted from fundus images using deep learning (24). Since these risk factors are easily obtained in a screening setting and can be strong confounders for prediction of type 2 diabetes from fundus images, a matching cohort experiment was designed. For this cohort, all individuals were selected that were newly identified as having type 2 diabetes in The Maastricht Study (*N* = 275), meaning that these individuals were previously unaware of having type 2 diabetes. Each newly identified case was matched with an individual with normal glucose metabolism, identical sex, and similar age and waist circumference.

**RESULTS**

The study population demographics and fundus image details are displayed in Table 1. Selection for the match-based cohort set was done before randomly assigning the remainder of individuals to the development set or validation set. Of the 6,453 individuals included in this study 1,502 (23.3%) had type 2 diabetes, while 967 (15.0%) had prediabetes. Most of the fundus images were centered on the optic disc or macula (64.8%). The remainder of the images was either fixated on the temporal periphery or ‘other’ (e.g. superior of the optic disc).

The results for discriminating individuals with known type 2 diabetes from individuals with normal glucose metabolism are presented in Table 2. The GMScore obtained with the deep learning model achieves an AUROC of 0.757 (95% CI 0.731 – 0.783) (see also Figure 1) which is higher than the individual risk factors age, sex, smoking, hypertension, and family history with AUROCs in the range 0.607 – 0.727. In contrast, waist circumference has a stronger predictive value as compared to the GMScore (AUROC of 0.832 (95% CI 0.810 - 0.853)). The GMScore also provides additional predictive value when combined with risk factors. For example, the AUROC for age, sex and waist circumference increases from 0.853 (95% CI 0.832 - 0.873) to 0.867 (95% CI 0.846 - 0.888) when combined with GMSscore. Even when all six individual factors are combined, the addition of the GMScore still provides extra predictive power with the AUROC increasing from 0.888 (95% CI 0.870 - 0.906) to 0.895 (95% CI 0.878 - 0.912) (p-value < 0.001).

Prediabetes individuals were excluded for the calculations of AUROCs in Table 2. This means that AUROCs obtained here are somewhat higher than when the prediabetes group would have been included. For example, the AUROC of 0.757 obtained with the GMScore decreases to 0.736 (95% CI 0.710 - 0.762), when the prediabetes group is added to those with normal glucose metabolism. Similarly, the AUROC of 0.832 for waist circumference would decrease to 0.792 (95% CI 0.770 - 0.815).

In this study, waist circumference was included as a risk factor since multiple studies have shown waist circumference to be a stronger discriminator for type 2 diabetes than Body Mass Index (BMI) (4,25). A post-hoc analysis shows that this is also true for our dataset, where the AUROC for BMI was 0.773 (95% CI 0.748 - 0.798).

On average, the data set contains 7.2 (± 3.6) fundus images per individual and the final glucose metabolism score is obtained by averaging across images of the left and right eye. The use of multiple fundus images per individual potentially improves the accuracy of the GMScore for two reasons: (1) different images focus on different parts of the retina and (2) averaging across multiple examples makes the prediction more robust. However, additional images also require more time to collect. We therefore studied the effect of the number of fundus images per individual by selecting all individuals for whom at least five images were available (83% of individuals). For this subset we recalculated the AUROC using 1, 2, 3, 4 or 5 images which were sampled randomly without replacement. Results are shown in Figure 1. When only one image is used per individual, the AUROC decreases to 0.715 (95% CI 0.685 – 0.745). Starting at three images per individual, the AUROC is similar to the one obtained when all images are used.

A picture containing chart

Description automatically generated

Figure 1: Receiver operating characteristic (ROC) curve. The light-blue area represents the 95% CI for the ROC calculated with all images.

The performance stratified per different fixation of the images varied. The image-level AUROC for images centered on the optic disc was found to be 0.731 (95% CI 0.713 - 0.750); macula: 0.737 (95% CI 0.720 - 0.754); periphery: 0.715 (95% CI 0.690 - 0.740); and other: 0.713 (95% CI 0.688 - 0.738).

Figure 2 shows boxplots of the glucose metabolism scores as computed by the deep learning algorithm for the individuals of the validation set. Even though the algorithm was trained using only examples from individuals with normal glucose metabolism or type 2 diabetes, the prediabetes group has a distinct distribution, in between the two other groups. The means of the normal and prediabetes groups are significantly different (p-value < 0.001). The AUROC for discriminating prediabetes individuals from individuals with normal glucose metabolism was 0.611 (95% CI 0.575 - 0.646).

Chart, box and whisker chart

Description automatically generated

Figure 2: Boxplots of glucose metabolism scores for individuals with normal glucose metabolism, prediabetes, and known type 2 diabetes.

For the matching cohort experiment, the mean matching distance was 0.98 years (age) and 1.5 cm (waist circumference). The AUROC for discriminating between individuals with normal glucose metabolism and newly discovered type 2 diabetes in this set was found to be 0.549 (0.500-0.597).

Heat maps (Figure 3) were constructed to visualize which regions of the fundus image contribute to high glucose metabolism scores (26). Details about the heat maps are provided in Supplementary 1. The examples shown in Figure 3 are from individuals with type 2 diabetes, correctly identified as such (true positives). For these individuals, the algorithm focusses on selective parts of the vascular tree. Although some of the heat maps are more diffuse than the presented examples, in general the focus seems to be on the venules and arterioles and not on the optic disc. We also looked at heat maps of false positives and observed that they looked similar to those of true positives, focusing on parts of the vascular tree.

A picture containing colorful, bubble, red, set

Description automatically generated

Figure 3: Heat maps (bottom) with corresponding fundus image (top) with high glucose metabolism scores from individuals with type 2 diabetes. Red regions in the heat map represent areas that contribute strongly to the glucose metabolism score. (A) Optic disc centered. (B) Macula centered. (C) Periphery centered.

**Conclusions**

In this study, a deep learning model was developed to obtain a glucose metabolism score (GMScore) from fundus images that can be used to discriminate between individuals with normal glucose metabolism and those with known type 2 diabetes. The AUROC obtained with the GMScore is higher than those obtained with age, sex, smoking, hypertension, and family history. Only waist circumference provided a higher AUROC. These results indicate that fundus images could be more informative for discriminating type 2 diabetes than some of the other well-known risk factors. A possible explanation is that the retina contains information about multiple of these factors. For example, Poplin et al. (24) showed that deep learning can be used to determine age, sex, and smoking status from fundus images relatively well. However, fundus images seem to contain additional information about the glucose tolerance status of individuals. This is supported by our results, which show that even when multiple risk factors are combined, the AUROC increases when the GMScore is added. Similarly, this finding is supported by the matching cohort experiment results that showed that the AUROC is modestly higher than 0.5, even though matching was done for age, sex, and waist circumference.

Averaging across multiple fundus images per individual seems to have a beneficial effect on the quality of the GMScore score, as it leads to a higher AUROC. There is a flattening effect from 3 images onwards, and future research could consider including 3 images instead of just one per individual. It should be noted that for this analysis only individuals were included that had at least 5 fundus images available which could introduce a selection bias. However, since the subset for this analysis contains 83% of individuals, the effect on the AUROC should be small. We also found that the fixation of fundus images influences the image-level AUROC. The GMScore obtained from fundus images centered at the optic disc or macula results in a higher AUROC than those obtained from the periphery. A possible explanation for this is that more of the vascular tree is visible in optic disc and macula centered images, compared to other fixations.

Even though the algorithm was trained to separate normal glucose metabolism from type 2 diabetes fundus images, we found that the GMscore can actually be used to discriminate prediabetes from normal glucose metabolism individuals to some extent, with an AUROC of 0.611. The distribution of GMScores for prediabetes individuals falls in between type 2 diabetes and normal glucose metabolism, indicating that the deep learning model finds modest signs of type 2 diabetes for the prediabetes group.

Our deep learning approach does not require any handcrafted features, such as arteriolar width, but learns the relevant features directly from the data. An advantage of this strategy is that no priors are needed and that no unknown features are missed. A disadvantage that is often attributed to deep learning is the limited explainability of the decision making due to the large number of model weights. The use of heat maps (26) in this study allows for insight into the focus of the algorithm. For individuals with a high GMScore, the prediction often seems to originate from selective parts of the vascular tree. This makes sense, since studies have shown the effect of (pre)diabetes on the microvessels, such as the impaired microvascular function and different calibers as described by Sörensen et al. (27) and Li et al. (21).

In contrast to the extensive research on diabetic retinopathy detection (29,30), only limited work has been done on direct classification of type 2 diabetes from fundus images via deep learning. Exploratory studies (31,32) have shown the feasibility of deep learning for type 2 diabetes detection, while one other study tried to determine haemoglobin A1c (HbA1c) levels with limited success (24). We used the results from previous research (31) for model design choice, including weight initialization, image preprocessing, data augmentation, training strategies, and patient-level aggregation of GMScores. To the best of our knowledge, the study presented here is the first clinical evaluation of the value of deep learning on fundus images for type 2 diabetes detection.

We showed that the AUROC obtained with waist circumference decreased from 0.832 to 0.792 if the prediabetes group was added to those with normal glucose metabolism. Interestingly this is still higher than reported by others. For example, in a meta-analysis by Lee et al. the AUROC for waist circumference was reported to be 0.70 for men and 0.74 for women, although with great disparity between ethnic groups (25)

One limitation of our study is that we developed our methods and validated our results on data from the same distribution (e.g. demographics, fundus photography settings). Future research should focus on applying these models to fundus and glucose metabolism status data collected at different centers.

Another limitation is that we did not take into account the length of an individual’s diabetes history, nor the delay between the glucose fasting measurements and the fundus photo acquisition. Both time components can potentially impact the manifestation in the retinal microvessels. Similarly, we did not consider the effect of diabetes treatment on the microvascular system, even though medication reduces the harmful effects of high blood glucose, potentially partly reversing the impaired microvascular function and the changes in the fundus images. The effect of the reversibility and time components could be topic of future research, for example by specifying groups for which medication is (in)effective, or by grouping participants by length of their diabetes history.

In conclusion, it was shown that fundus images are informative for discriminating individuals with normal glucose metabolism, prediabetes, and type 2 diabetes. Using deep learning, a glucose metabolism score was obtained that proved more predictive than other risk factors, except for waist circumference.

**Acknowledgements**

This research is financially supported by the TTW Perspectief program and Philips Research. The clinical data used in the research originates from The Maastricht Study which was supported by the European Regional Development Fund via OP-Zuid, the Province of Limburg, the Dutch Ministry of Economic Affairs (grant 31O.041), Stichting De Weijerhorst (Maastricht, The Netherlands), the Pearl String Initiative Diabetes (Amsterdam, The Netherlands), CARIM School for Cardiovascular Diseases (Maastricht, The Netherlands), Stichting Annadal (Maastricht, The Netherlands), Health Foundation Limburg (Maastricht,The Netherlands) and by unrestricted grants from Janssen-Cilag B.V. (Tilburg, The Netherlands), Novo Nordisk Farma B.V. (Alphen aan den Rijn, The Netherlands), and Sano-Aventis Netherlands B.V. (Gouda, The Netherlands).

The authors have no conflicts of interests to report.

**Contributions**

F.H., T.B., A.H., and M.V. conceptualized the study design. F.H. implemented the deep learning models and conducted the analyses. F.H., T.B., A.H., and M.V. interpreted the analyses, and drafted the manuscript. M.S., C.S., M.G., and J.P. interpreted the analyses and revised the manuscript. All authors approved the final version of the article. F.G.H. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**References** (I’m aware of the formatting bug for second names that start with a ‘V’)

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019;157.

2. Congdon NG, Friedman DS, Lietman T. Important causes of visual impairment in the world today. JAMA. 2003;290(15):2057–60.

3. Gu K, Cowie CC, Harris MI. Diabetes and decline in heart disease mortality in US adults. JAMA. 1999;281(14):1291–7.

4. Browning LM, Hsieh SD, Ashwell M. A systematic review of waist-to-height ratio as a screening tool for the prediction of cardiovascular disease and diabetes: 0.5 could be a suitable global boundary value. Nutr Res Rev. 2010;23(2):247–69.

5. van ’t Riet E, Dekker JM, Sun Q, Nijpels G, Hu FB, van Dam RM. Role of adiposity and lifestyle in the relationship between family history of diabetes and 20-year incidence of type 2 diabetes in U.S. women. Diabetes Care. 2010;33(4):763–7.

6. Lee YH, Bang H, Kim HC, Kim HM, Park SW, Kim DJ. A simple screening score for diabetes for the Korean population: Development, validation, and comparison with other scores. Diabetes Care. 2012;35(8):1723–30.

7. Cho E, Min D, Lee HS. Development and validation of an undiagnosed diabetes screening tool: Based on the Korean national health and nutrition examination survey (2010–2016). Healthcare. 2021;9(9):1138.

8. Nguyen TT, Wang JJ, Richey Sharrett A, Amirul Islam FM, Klein R, Klein BEK, et al. Relationship of retinal vascular caliber with diabetes and retinopathy: the Multi-Ethnic Study of Atherosclerosis (MESA). Diabetes Care. 2008;31:544–9.

9. Nguyen TT, Wang JJ, Wong TY. Retinal vascular changes in pre-diabetes and prehypertension: new findings and their research and clinical implications. Diabetes Care. 2007;30(10):2708–15.

10. Li W, Schram MT, Berendschot TTJM, Webers CAB, Kroon AA, van der Kallen CJH, et al. Type 2 diabetes and HbA1c are independently associated with wider retinal arterioles: the Maastricht study. Diabetologia. 2020;63(7):1408–17.

11. Quinn N, Jenkins A, Ryan C, Januszewski A, Peto T, Brazionis L. Imaging the eye and its relevance to diabetes care. J Diabetes Investig. 2021;12(6):897–908.

12. Sasongko MB, Wong TY, Nguyen TT, Cheung CY, Shaw JE, Wang JJ. Retinal vascular tortuosity in persons with diabetes and diabetic retinopathy. Diabetologia. 2011;54(9):2409–16.

13. Sasongko MB, Wong TY, Nguyen TT, Cheung CY, Shaw JE, Kawasaki R, et al. Retinal vessel tortuosity and its relation to traditional and novel vascular risk markers in persons with diabetes. Curr Eye Res. 2016;41(4):551–7.

14. ter Haar Romeny BM, Bekkers EJ, Zhang J, Abbasi-Sureshjani S, Huang F, Duits R, et al. Brain-inspired algorithms for retinal image analysis. Mach Vis Appl. 2016;27(8):1117–35.

15. Heslinga FG, Pluim JPW, Dashtbozorg B, Berendschot TJM, Houben AJHM, Henry RMA, et al. Approximation of a pipeline of unsupervised retina image analysis methods with a CNN. Proceedings of SPIE 10949, Medical Imaging 2019: Image Processing. 2019;10949.

16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.

17. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016 Dec 13;316(22):2402–10.

18. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci. 2016 Oct 1;57(13):5200–6.

19. Ting DSW, Cheung CY-L, Lim G, Tan GSW, Quang ND, Gan A, 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. 2017;318(22):2211–23.

20. Schram MT, Sep SJS, van der Kallen CJ, Dagnelie PC, Koster A, Schaper N, et al. The Maastricht Study: An extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. Eur J Epidemiol. 2014;29(6):439–51.

21. World Health Organization, International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. World Health Organization; 2006.

22. Foracchia M, Grisan E, Ruggeri A. Luminosity and contrast normalization in retinal images. Med Image Anal. 2005;9(3):179–90.

23. Tan M, Le Q v. EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning. 2019;PMLR 97:6105–14.

24. Poplin R, Varadarajan A v, Blumer K, Liu Y, McConnell M v, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng. 2018;2:158–64.

25. Lee CMY, Huxley RR, Wildman RP, Woodward M. Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: a meta-analysis. J Clin Epidemiol. 2008 Jul;61(7):646–53.

26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. IEEE International Conference on Computer Vision (ICCV). 2017;618–26.

27. Sörensen BM, Houben AJHM, Berendschot TTJM, Schouten JSAG, Kroon AA, van der Kallen CJH, et al. Prediabetes and type 2 diabetes are associated with generalized microvascular dysfunction: The Maastricht Study. Circulation. 2016;134(18):1339–52.

28. Li W, Schram MT, Sörensen BM, van Agtmaal MJM, Berendschot TTJM, Webers CAB, et al. Microvascular phenotyping in The Maastricht Study: Design and main findings, 2010–2018. Am J Epidemiol. 2020;189(9):873–84.

29. Bellemo V, Lim G, Rim TH, Tan GSW, Cheung CY, Sadda S v, et al. Artificial intelligence screening for diabetic retinopathy: the real-world emerging application. Curr Diab Rep. 2019;19(9):Article number 72.

30. Asiri N, Hussain M, al Adel F, Alzaidi N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artif Intell Med. 2019;99:Article ID 101701.

31. Heslinga FG, Pluim JPW, Houben AJHM, Schram MT, Henry RMA, Stehouwer DA, et al. Direct classification of type 2 diabetes from retinal fundus images in a population-based sample from The Maastricht Study. Proceedings of SPIE 11314, Medical Imaging 2020: Computer-Aided Diagnosis. 2020;11314.

32. Abbasi-Sureshjani S, Dashtbozorg B, ter Haar Romeny BM, Fleuret F. Exploratory study on direct prediction of diabetes using deep residual networks. VipIMAGE2017. 2018;27:797–802.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Development** | **Validation** | **Matching cohort** |
| Individuals | 3,903 | 2,000 | 550 |
| Normal (%) | 2,460 (63.0) | 1,249 (62.5) | 275 (50.0) |
| Prediabetes (%) | 632 (16.2) | 335 (16.8) | - |
| Type 2 diabetes (%) | 811 (20.8) | 416 (20.8) | 275 (50.0) |
| Newly identified type 2 diabetes individuals | - | - | 275 (50.0) |
| Females (%) | 1977 (50.7) | 1044 (52.2) | 216 (39.3) |
| Age [years] (std) | 59.4 (8.8) | 59.3 (8.6) | 63.4 (7.9) |
| Waist [cm] (std) | 94.5 (13.5) | 94.0 (13.3) | 102.7 (12.4) |
| Hypertension\* (%) | 2,031 (52.1) | 1,039 (52.1) | 368 (66.9) |
| Smoking†; Current (%) | 497 (12.8) | 249 (12.5) | 60 (11.0) |
| Former (%) | 1,912 (49.4) | 967 (48.6) | 284 (52.0) |
| Non-smoking (%) | 1,464 (37.8) | 773 (38.9) | 202 (37.0) |
| Fundus images | 28,153 | 14,476 | 3,742 |
| Left eye (%) | 14,314 (50.8) | 7,339 (50.7) | 1,872 (50.0) |
| Optic disc centered (%) | 8,763 (31.1) | 4,409 (30.5) | 1,242 (33.2) |
| Macula centered (%) | 9,468 (33.6) | 4,862 (33.6) | 1,316 (35.2) |
| Periphery centered (%) | 5,242 (18.6) | 2,620 (18.1) | 663 (17.7) |
| Other (%) | 4,680 (16.6) | 2,585 (17.9) | 521 (13.9) |

Table 1: Study population demographics and fundus image details. Normal = Normal glucose metabolism. \* Missing data for 7 individuals. † Missing data for 45 individuals.

|  |  |
| --- | --- |
| **Factors used for classification** | **AUROC (95% CI)** |
| GMScore | 0.757 (0.731 - 0.783) |
| Age | 0.691 (0.663 - 0.719) |
| Sex | 0.607 (0.580 - 0.634) |
| Age, Sex | 0.711 (0.682 - 0.740) |
| Waist | 0.832 (0.810 - 0.853) |
| Age, Sex, Waist | 0.853 (0.832 - 0.873) |
| Smoking\* | 0.581 (0.551 - 0.610) |
| Hypertension\* | 0.727 (0.704 - 0.749) |
| Family history\* | 0.657 (0.625 - 0.689) |
| Age, Sex, Waist, Smoking, Hypertension, Family history\* | 0.888 (0.870 - 0.906) |
| GMScore, Age, Sex | 0.773 (0.747 - 0.799) |
| GMScore, Age, Sex, Waist | 0.867 (0.846 - 0.888) |
| GMScore, Age, Sex, Waist, Smoking, Hypertension, Family history\* | 0.895 (0.878 - 0.912) |

Table 2: Performance of the algorithm in comparison with other risk factors. AUROC = Area under receiver operating characteristic. GMScore = glucose metabolism score, obtained using the deep learning algorithm. \*Some individuals (< 1.3%) were excluded for AUROC calculation because of missing data points

**Supplementary 1: Deep learning model development**

The development set was split on an individual-level into a set for model training and a set for model selection. Data augmentation was used to increase the variety in the training set: translation (≤ 20 pixels), rotation (< 360 degrees), scaling (0.95-1.05), horizontal and vertical reflection, intensity shift (≤ 10/255), color (≤ 10/255) and contrast shift (0.95-1.05) and addition of white noise (≤ 3/255). For training, only fundus images from individuals with normal glucose metabolism (label 0) and type 2 diabetes (label 1) were included.

EfficientNet B4 (23) was implemented in Keras with a Tensorflow backend to process images of size 512 x 512 pixels on a single GPU (Nvidia Titan Xp). The final model layer was replaced by a single node with a sigmoid activation to output a probability value that is interpreted as glucose metabolism score (GMscore). The model was initialized with random weights, which were iteratively updated by minimizing the binary cross-entropy loss between the labels and the model prediction using the Adam optimizer. The model was trained for 245,760 iterations using batches of eight images and an initial learning rate of 0.001. The learning rate was multiplied with a factor of 0.3 every 61,440 iterations. For evaluation, fundus images of the validation set were processed resulting in a glucose metabolism score between 0 and 1.

For the heat maps, grad-CAM was used. Since the horizontal and vertical dimension of the final convolutional layer were only 16×16, the main model provided a heat map with limited detail. Therefore, a separate model was trained for the grad-CAM, using the same EfficientNet B4 architecture but with the removal of the final downscaling operation (originally implemented as a stride = 2 in a convolutional layer). To train this model, the batch size was decreased to six to fit in GPU memory. The resulting heat maps were sized 31×31 which were upscaled and overlayed on a gray-level image of the original fundus image. Heat maps were normalized per image, such that regions with high activations are always shown in red and low activations are shown in blue. A mask was used to remove any heat map signal in the background.

A picture containing logo

Description automatically generated

Supplementary eFigure 1: Exclusion of low-contrast fundus images. Image contrast was calculated by subtracting a blurred version from the original image and summing over all non-background pixels of the difference map. Blurring was done using a 2-D Gaussian filter (sigma = 3×3). A threshold was manually selected to exclude 12,000 (20.4%) images. (A) Low-contract image. (B) Example just below the threshold for inclusion. (C) Example just above the threshold. (D) High-quality image.

A picture containing bubble chart

Description automatically generated

Supplementary eFigure 2: Fundus image preprocessing. (A) The original image (3744×3744 pixels) is cropped and resized (B) to 512×512 pixels. (C) Image luminosity and contrast is normalized channel-wise, similar to Foracchia et al. (22).