

A Deep Learning Model for Screening Type 2 Diabetes from Retinal Photographs

Short running title: Screening Diabetes using Deep Learning Model

Jae-Seung Yun, M.D., Ph.D.^{1,2§}, Jaesik Kim, B.S.^{1,3,4§}, Sang-Hyuk Jung, B.E.^{1,4,5}, Seon-Ah Cha, M.D., Ph.D.², Seung-Hyun Ko, M.D., Ph.D.², Yu-Bae Ahn, M.D., Ph.D.², Hong-Hee Won, Ph.D.⁵, Kyung-Ah Sohn, Ph.D.^{3,7*}, Dokyoon Kim, Ph.D.^{1,4*}

¹Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

²Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

³Department of Computer Engineering, Ajou University, Suwon, Republic of Korea.

⁴Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

⁵Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea

⁶Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA

⁷Department of Artificial Intelligence, Ajou University, Suwon, Republic of Korea

§Jae-Seung Yun and Jaesik Kim contributed equally as co-first authors.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

*Kyung-Ah Sohn and Dokyoon Kim contributed equally as co-corresponding authors.

Correspondence to:

Kyung-Ah Sohn

206, World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, 16499, Republic of Korea

Phone: +82-031-219-2434;

E-mail: kasohn@ajou.ac.kr

Dokyoon Kim

B304 Richards Building, 3700 Hamilton Walk, University of Pennsylvania, Philadelphia, PA
19104-6021

Phone: +1-215-573-5336; Fax: +1-215-573-3111;

E-mail: dokyoon.kim@penncmedicine.upenn.edu

Word count: 3,928

Number of tables and figures: 4

1 **Abstract**

2 **Objective**

3 We aimed to develop and evaluate a non-invasive deep learning algorithm for screening type 2
4 diabetes in UK Biobank participants using retinal images.

5 **Research Design and Methods**

6 The deep learning model for prediction of type 2 diabetes was trained on retinal images from
7 50,077 UK Biobank participants and tested on 12,185 participants. We evaluated its
8 performance in terms of predicting traditional risk factors (TRFs) and genetic risk for diabetes.
9 Next, we compared the performance of three models in predicting type 2 diabetes using 1) an
10 image-only deep learning algorithm, 2) TRFs, 3) the combination of the algorithm and TRFs.
11 Assessing net reclassification improvement (NRI) allowed quantification of the improvement
12 afforded by adding the algorithm to the TRF model.

13 **Results**

14 When predicting TRFs with the deep learning algorithm, the areas under the curve (AUCs)
15 obtained with the validation set for age, sex, and HbA1c status were 0.931 (0.928-0.934), 0.933
16 (0.929-0.936), and 0.734 (0.715-0.752), respectively. When predicting type 2 diabetes, the
17 AUC of the composite logistic model using non-invasive TRFs was 0.810 (0.790-0.830), and
18 that for the deep learning model using only fundus images was 0.731 (0.707-0.756). Upon
19 addition of TRFs to the deep learning algorithm, discriminative performance was improved to
20 0.844 (0.826-0.861). The addition of the algorithm to the TRFs model improved risk
21 stratification with an overall NRI of 50.8%.

22 **Conclusions**

1 Our results demonstrate that this deep learning algorithm can be a useful tool for stratifying
2 individuals at high risk of type 2 diabetes in the general population.

3

4 **Key Words:** Obesity and Type 2 Diabetes; Retinal Imaging; Prediction Model; Machine
5 Learning; Screening

6

7 **Non-Standard Abbreviations and Acronyms**

8 AI, artificial intelligence

9 AUC, area under the curve

10 CE, Cross-entropy

11 CVD, Cardiovascular disease

12 NPV, negative predictive value

13 PPV, positive predictive value

14 R^2 , R-squared

15 TRF, traditional risk factor

1 **Introduction**

2 Type 2 diabetes is one of the fastest-growing diseases in the world and poses a major threat to
3 health globally (1). Increased prevalence of type 2 diabetes and its complications leads to
4 increased risk of mortality in individuals with diabetes, and so adds to the already-profound
5 social and economic burden (2). Type 2 diabetes has a long asymptomatic period before actual
6 onset (3); however, several reports indicated that awareness of diabetes or prediabetes is not
7 high in the general population (4,5). Therefore, regular screening for diabetes is important,
8 particularly as the disease can be effectively prevented by selecting high-risk groups and
9 performing appropriate interventions for prevention.

10 Funduscopy examination is a rapid, non-invasive, and effective tool for screening
11 retinopathy, especially in the context of diabetes. Many studies have investigated automated
12 screening of retinopathy using deep learning (6). Deep learning algorithms in artificial
13 intelligence (AI) have particular strengths in analyzing image data, and a couple of studies have
14 demonstrated equivalent or even better detection performance using deep learning compared
15 to medical practitioners (7). It has been suggested that algorithmic disease prediction using
16 fundus images can be applied not only to retinopathy, but also to nephropathy, neuropathy, and
17 cardiovascular disease (8-11).

18 It is unclear whether a predictive algorithm comprised of a deep learning model trained
19 on funduscopy images can detect features related to diabetes and can furthermore screen for
20 diabetes itself. It is known that overt retinal lesions such as microaneurysms, cotton-wool spots,
21 hard exudates, and venous beading generally appear 10-15 years after overt diabetes occurs
22 (12). However, according to previous studies, analysis of fundus images not only captures
23 retinal vascular or neural damage but also has considerable capacity to detect metabolic

1 biomarkers related to type 2 diabetes such as age, blood pressure, and BMI (13). Therefore,
2 given the accuracy of AI-based image recognition, screening for type 2 diabetes on the basis of
3 retinal images poses a reasonable challenge for deep learning research.

4 The UK Biobank project is a nationwide, prospective, population-based cohort study
5 that provides a variety of genetic, lifestyle, and clinical data, and furthermore includes over
6 100,000 retinal images. We primarily aimed to establish a screening model for type 2 diabetes
7 using clinical data and funduscopy images from the UK Biobank. In addition, we aimed to
8 investigate whether there are significant associations between the traditional risk factors (TRFs)
9 of diabetes and funduscopy images, and also whether the diabetes risk associated with a fundus
10 image captures either or both genetic and acquired components for diabetes. We finally tested
11 the added value of the deep learning model and TRFs for prediction of type 2 diabetes using
12 discrimination and reclassification methods.

13

14 **Research Design and Methods**

15 **Study population**

16 We used the UK Biobank dataset to develop and validate a deep learning algorithm for
17 prediction of type 2 diabetes using retinal fundus photographs. The UK Biobank project is a
18 prospective observational study that recruited 505,025 UK participants, aged 40-69 years at
19 baseline, between 2006 and 2010. Each participant provided informed consent, completed a
20 touchscreen and in-person interview with trained staff, and underwent a series of physical
21 examinations. Extensive information was collected, including lifestyle, sociodemographic
22 factors, medical history, biologic samples, imaging, and genome-wide genotype data. Detailed

1 protocols for obtaining the data are available on the UK Biobank website at
2 www.ukbiobank.ac.uk. The UK Biobank has ethical approval from the National Research
3 Ethics Committee (June 17, 2011 [RES reference 11/NW/0382]), which was further extended
4 (May 10, 2016 [RES reference 16/NW/0274]). Use of the UK Biobank Resource in the current
5 study was approved under Application Number 67855. The external validation set consisted of
6 6,575 retinal fundus images from type 2 diabetes patients of the University-affiliated Diabetes
7 Center of St. Vincent's Hospital. Use of the external validation set was approved by the
8 Catholic Medical Center Ethics Committee and conducted in accordance with the Declaration
9 of Helsinki.

10 A total of 62,262 participants underwent retinal fundus imaging using a Topcon 3D
11 OCT-1000 MKII (Topcon Corporation) during 2009-2010. All such images were obtained with
12 a 45° primary field of view. The patients were randomly divided into three groups with no
13 overlapping individuals: the training set, $n = 37,904$; the tuning set, $n = 12,173$; and the
14 validation set ($n = 12,185$) (Supplemental Fig 1). The training and tuning sets together
15 comprised the development set. These sets accounted for approximately 12% of patients in the
16 UK Biobank dataset. Images of poor quality were filtered out before training and validation;
17 only gradable retinal fundusoscopic images were used in this study. Detailed methods of image
18 preprocessing for quality control are summarized in the Supplemental Materials.

19

20 **Assessment of major variables**

21 Definition of prevalent type 2 diabetes at baseline was based on a self-report touchscreen
22 questionnaire, nurse-led interview at enrollment, or diagnostic codes recorded across all
23 hospital visits. We excluded subjects who reported type 1 diabetes in their verbal interview and

1 those who had a diagnostic code for type 1 diabetes (E10). These definitions are presented in
2 Supplemental Table S1.

3 Blood pressure was measured twice using the Omron HEM-7015IT digital blood
4 pressure monitor (Omron Healthcare) or, exceptionally, a manual sphygmomanometer. We
5 defined hypertension as systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90
6 mmHg. We considered four lifestyle factors: current smoking, obesity, physical activity, and
7 dietary pattern, as recommended by the strategic goals of the American Heart Association (14).
8 Family history of diabetes was defined as positive if either mother or father had diabetes and
9 negative otherwise. Additional definitions and details regarding lifestyle factors and polygenic
10 risk score for type 2 diabetes are provided in Supplemental Table S2 and Supplemental
11 Materials, respectively.

12

13 **Algorithm development**

14 We used ResNet18 as the deep learning algorithm, pre-trained on ImageNet data, with a
15 resolution of 578×578 and taking fundus images as input (15). Since this model uses a single
16 input image, cases where at least one eye of the patient was available were used for
17 development. Mean squared error loss was used for the continuous target variable and cross-
18 entropy loss for the categorical target variable during training. To prevent overfitting, we
19 applied common augmentation methods such as flip, rotation, and crop, and implemented early
20 stopping based on the average loss of the tuning set. Details of the model architecture and
21 development are given in Supplemental Materials.

22

1 Statistical analysis

2 For performance metrics, we used R-squared (R^2) for continuous target variables and the area
3 under the receiver operating characteristic curve (AUC) for binary target variables. In addition,
4 we determined the sensitivity, specificity, positive predictive value (PPV), and negative
5 predictive value (NPV) at an optimal threshold that maximized sensitivity plus specificity for
6 binary classification. To evaluate the statistical significance of performance values, non-
7 parametric bootstrapping with replacement was used to obtain metric distribution and the 95%
8 confidence interval. We repeated 2,000 times, sampling from the validation set.

9 We first evaluated the capacity of the deep learning algorithm to predict TRFs
10 associated with diabetes. Based on the literature (16), a total of eleven TRFs were selected for
11 analysis: age, sex, hypertension, BMI, waist circumference, history of CVD, unfavorable
12 lifestyle, triglyceride, HDL cholesterol, serum glucose, and HbA1c. Given the lack of standard
13 criteria for determining acceptance in terms of R^2 , and assuming that some risk factors may
14 have little effect on the retina before reaching a certain level, we also examined binary variables
15 with discretized versions of continuous TRF variables defined on the basis of established cut-
16 off levels for metabolic disorder: age, <60 years, ≥ 60 years; hypertension, <140/90 mmHg,
17 $\geq 140/90$ mmHg; BMI, <30 kg/m², ≥ 30 kg/m²; waist circumference, <102 cm (men) or <88 cm
18 (women), ≥ 102 cm (men) or ≥ 88 cm (women); triglyceride, <1.70 mmol/L, ≥ 1.70 mmol/L;
19 HDL cholesterol, ≥ 1.03 mmol/L (men) or ≥ 1.29 mmol/L (women), <1.03 mmol/L (men) or
20 <1.29 mmol/L (women); HbA1c (prediabetes), <5.7% or $\geq 5.7\%$; and HbA1c (diabetes), <6.5%
21 or $\geq 6.5\%$. Next, we trained and evaluated logistic regression models for predicting type 2
22 diabetes using non-invasive TRFs and the deep learning algorithm. To fairly compare
23 diagnostic performance, we determined the logistic model performances of each non-invasive

1 TRF, the algorithm alone, and the TRFs plus the algorithm. The development set was used for
2 logistic model learning and the validation set for performance comparison. For patients for
3 which the algorithm made an incorrect prediction, we evaluated their characteristics as a group
4 using cross-entropy (CE) loss. To quantify the improvement obtained by incorporating the
5 algorithm into the TRF model, we calculated the continuous net reclassification improvement
6 (NRI). Details concerning the calculation of NRI and CE loss are addressed in Supplemental
7 Materials. To further assess the accuracy of our algorithm, we used follow-up data from the
8 UK Biobank (median 8 years) to analyze the future incidence of type 2 diabetes in participants
9 who the algorithm screened as high-risk but did not have diabetes at baseline ($n = 8,889$).

10 An additional validation test was performed using an external dataset comprised of
11 6,575 retinal fundus images from patients with type 2 diabetes at the University-affiliated
12 Diabetes Center of St. Vincent's Hospital (Supplemental Materials). To exclude the effect of
13 overt retinal lesions in patients with diabetic retinopathy on the performance of the deep
14 learning algorithm, we carried out sensitivity analysis after excluding participants who had
15 diagnostic codes for diabetic retinopathy. We also performed subgroup analysis with 39,473
16 participants in our dataset who were eligible for both genetic analyses and funduscopy images
17 to evaluate the association between retinal images and inherited risk for type 2 diabetes. As
18 genetic risk markers, we used the family history of diabetes and polygenic risk score (PRS)
19 (Supplemental Materials) (17).

20 All analyses were performed using Python (version 3.8.1) with the following libraries:
21 numpy (version 1.18.5); pandas (version 1.0.5); pytorch (version 1.5.1); sklearn (version 0.23.1)

22

23 **Results**

1 **Baseline characteristics of study population**

2 We developed predictive algorithms for type 2 diabetes using 69,639 retinal fundus images
3 from 37,904 patients, tuned using 22,342 images from 12,173 patients, and validated using
4 22,394 images from 12,185 patients, totaling 62,262 participants, all from the same UK
5 Biobank dataset. At the time of enrollment, 7,891 (14.9%) participants had prediabetes, 2,691
6 (4.8%) had type 2 diabetes, and 4,149 (6.7%) had cardiovascular disease. There were no
7 differences in clinical characteristics between the training, tuning, and validation cohorts,
8 which were randomly divided. The demographic characteristics of the study population are
9 shown in Table 1.

10

11 **Deep learning algorithm for prediction of TRFs associated with type 2 diabetes**

12 First, we tested and validated the capability of the algorithm to predict well-known risk factors
13 for type 2 diabetes from fundus images. The resulting area under the curve (AUC) values
14 obtained for the validation set are listed in Table 2. The prediction algorithm showed fair
15 performance for HbA1c >6.5% and FBS >7.0 mmol/L, with AUC values of 0.734 (0.715-0.752)
16 and 0.672 (0.648-0.694), respectively; however, its predictive power for triglyceride and HDL
17 cholesterol was relatively low. Relatively high R^2 values were obtained per one unit of age and
18 blood pressure, and relatively low R^2 values per one unit of BMI, waist circumference, lipid
19 profile, glucose, and HbA1c (Table 2).

20

21 **Deep learning algorithm for prediction of type 2 diabetes**

22 Next, we compared the risk factor and deep learning algorithm models in terms of diagnostic

1 performance, and further investigated whether the deep learning model can effectively screen
2 for the presence of type 2 diabetes based on a fundus image (Figure 1, Supplemental Table S3).
3 In terms of performance, the composite diagnostic model using traditional non-invasive risk
4 markers achieved an AUC of 0.810 (0.790-0.830), while that for the deep learning model using
5 only fundus images was 0.731 (0.707-0.756). The AUC of univariate model using glucose
6 alone was 0.795 (0.764-0.826), which is one of the diagnostic criteria for diabetes.

7 Among participants without type 2 diabetes, those having higher CE loss (top 20
8 percentile) were older, were more commonly women, were less often of European ethnicity,
9 were more likely to have hypertension and a history of cardiovascular disease, more commonly
10 had prediabetes, and had poorer metabolic profiles than those with lower CE loss (bottom 20
11 percentile). They also had higher levels of metabolic risk factors than those with lower CE loss
12 (Supplemental Table S4). Supplementary Figure 2 shows that participants in the higher CE loss
13 group are more similar in characteristics to the type 2 diabetes group than to those with lower
14 CE loss (Detailed descriptions can be found in Supplemental Materials). During the median 8-
15 year follow-up period of the UK Biobank project, type 2 diabetes developed in 134 of 8,889
16 participants in the validation cohort who did not have the disease at baseline. Participants with
17 a higher CE loss tended to have higher incident rate of type 2 diabetes, with borderline
18 significance (Supplemental Table S4, P for trend = 0.051).

19 After excluding 382 individuals with diabetic retinopathy, model performance
20 remained essentially unaltered (AUC 0.726 [0.699-0.752]). In validation with the mixed
21 internal and external dataset, sets with balanced labels (50% cases and 50% control) and
22 imbalanced labels (4.7% cases and 94.3% controls) both achieved performance similar to the
23 original results (balanced AUC, 0.703 [0.691-0.715]; imbalanced AUC, 0.703 [0.679-0.727])

1 (Supplemental Table S5).

2 We also investigated the performance of the deep learning algorithm in predicting
3 prediabetes in our study cohort. When targeting prediabetes participants who were HbA1c \geq 5.7%
4 and <6.5% and without diabetes (n = 1723), the AUC was 0.647 (0.631-0.662) (Supplemental
5 Table S6).

6

7 **TRFs plus deep learning algorithm model for prediction of type 2 diabetes**

8 Adding the TRFs to the deep learning algorithm significantly improved predictive power,
9 yielding an AUC of 0.844 (0.826-0.861). We evaluated individual classifications when adding
10 the deep learning algorithm to the TRFs model and ultimately reclassified 3.0% of type 2
11 diabetes patients and 47.8% of non-diabetes patients. Overall, the addition of algorithm to the
12 TRF model improved prediction of type 2 diabetes, with a continuous NRI of 50.8% (95% CI
13 40.9%-60.9%) (Table 3). Among participants who were classified in the top 10 percentile high-
14 risk subgroup by the TRF model, overall NRI was 51.2% (95% CI 19.5-82.9).

15

16 **Subgroup analysis of the association between retinal images and genetic risk**

17 To confirm whether funduscopy images can capture information concerning genetic risk for
18 type 2 diabetes, we additionally analyzed a subgroup of 39,473 participants for whom genetic
19 data was available after quality control (PRS, family history of diabetes). When using the deep
20 learning algorithm for prediction of PRS, the R^2 was low (R^2 , -0.0001 [-0.0008-0.0005]); the
21 AUC of the algorithm for the high-genetic-risk top 20 percentile was also poor (AUC, 0.499
22 [0.488-0.509]) (Supplemental Table S7). Similarly, this subgroup yielded low AUC values for

1 the prediction of type 2 diabetes when using either a PRS-based model (0.596 [0.566-0.627])
2 or the algorithm (0.711 [0.684-0.738]). When PRS information was added to the deep learning
3 algorithm, predictive power was slightly improved (AUC, 0.721 [0.693-0.749]), and more so
4 upon incorporating all genetic factors, clinical factors, and algorithms (AUC, 0.845 [0.823-
5 0.865]) (Supplemental Table S8).

6

7 **Conclusion**

8 The purpose of this study was to develop and validate a funduscopy image-based deep learning
9 model for screening type 2 diabetes using the UK Biobank dataset. The main finding of our
10 study was that the status of TRFs for type 2 diabetes, including age, sex, blood pressure, and
11 metabolic parameters, was modestly predictable by a deep learning algorithm on the basis of
12 retinal funduscopy images. The algorithm also showed moderate predictive performance for
13 type 2 diabetes, which was strengthened by combining fundus imaging with TRFs; thus,
14 incorporating fundus imaging alongside TRFs that do not require an invasive blood test can
15 improve the discriminative power of diagnostic models for type 2 diabetes. Notably, our deep
16 learning screening algorithm appears to capture acquired features rather than inherited features
17 to predict type 2 diabetes.

18 According to major diabetes guidelines, including those from the American Diabetes
19 Association, routine screening should be considered to identify prediabetes or type 2 diabetes
20 in adults who have one or more risk factors, including age >45 years, family history of diabetes,
21 history of cardiovascular disease, hypertension, dyslipidemia, and poor lifestyle habits (3).
22 However, such screening necessitates blood sampling tests that are invasive, along with two or
23 more visits to the clinic for medical examination, blood sampling, and confirmation of the test

1 results. In addition, if there is no unequivocal hyperglycemic symptom, two or more abnormal
2 results in two different samples are needed to confirm diabetes (3). On the other hand,
3 funduscopy examination is a rapid, inexpensive, non-invasive screening tool that can be
4 utilized even without pharmacologic pupil dilatation in an outpatient clinical setting. In
5 previous studies, the application of a deep learning model to retinal photographs has yielded
6 promising results in predicting systemic disease or biomarker levels. Specifically, Gerritis et al.
7 suggested that a deep neural network can predict age, sex, blood pressure, HbA1c, and fat mass
8 at an acceptable level (18), and Poplin et al. demonstrated that a deep learning model using
9 only funduscopy images from the UK Biobank and EyePACS datasets achieved moderate
10 performance at predicting six major cardiovascular risk factors, including age, sex, blood
11 pressure, and BMI (10). Another recent publication evaluated the capacity of deep learning
12 algorithms to predict 47 systemic biomarkers with seven Asian and European cohorts (13).
13 These studies reported deep learning models to predict age, sex, and blood pressure well, which
14 is in line with our results; they also reported promising performance of the models in predicting
15 body composition indices and kidney function, measures that are associated with development
16 of type 2 diabetes. These studies support the feasibility of funduscopy examination as a
17 screening tool for detecting type 2 diabetes.

18 Several possible links are known that connect retinal structure and the risk of diabetes.
19 For one, diabetic retinopathy shares risk factors with diabetes and its vascular complications
20 based on the common pathophysiology; in addition, epidemiological and mechanistic evidence
21 has consistently proposed interaction between diabetic complications (19-21). Retinal vessel
22 tortuosity has been suggested as associated with metabolic risk factors (22), and early retinal
23 changes such as wide retinal venular caliber have been consistently observed in prediabetes

1 and metabolic syndrome (23). Nonetheless, to the best of our knowledge, no previous attempt
2 has been made to predict diabetes status itself from fundus photographs using a deep learning
3 model. Since the deep learning algorithms in our study can detect systemic biomarkers closely
4 related to diabetes and can predict cardiovascular diseases that share common risk factors with
5 diabetes, it is plausible to predict diabetes from fundus photographs. However, because changes
6 in retinal structure that relate to diabetes are a long-term consequence of overt diabetes, it is
7 challenging to determine whether deep learning can detect fundus changes related to diabetes
8 itself. In our study, the deep learning algorithm demonstrated moderate performance in type 2
9 diabetes screening. Combining the deep learning model with TRFs significantly improved the
10 discriminative power for diabetes, which means that this combined model captures additional
11 information from retinal structure that is independent of conventional risk factors, thereby
12 improving risk stratification. Meanwhile, we observed only a weak association of funduscopy
13 images and genetic risk for diabetes as represented by family history or polygenic risk score.
14 A few prior studies have considered the association between genetic determinants and retinal
15 traits (24,25), but our results indicate that this prediction model captures features corresponding
16 to the acquired risk of type 2 diabetes rather than the inherited risk.

17 We compared the performance in discrimination and risk classification of a prediction
18 model using non-invasive TRFs and another that combines TRFs with an algorithm trained on
19 retinal images. The new model classified nearly half of the participants more appropriately,
20 even when only subjects in the high-risk group were analyzed. Namely, among participants
21 who were traditionally classified as high-risk for type 2 diabetes, 47.8% of those who did not
22 have diabetes were reclassified as lower-risk. All told, our new non-invasive predictive model
23 incorporating a deep learning algorithm and retinal images can be used conveniently and would

1 save time and medical costs by reducing the number of additional unnecessary blood tests or
2 medical visits.

3 We further investigated per-sample CE loss to determine the characteristics of the
4 group in which the deep learning model failed to diagnose diabetes. CE reflects the degree to
5 which a deep learning algorithm fails to predict (26). Among participants without type 2
6 diabetes, those having the top 20% CE possessed more risk factors for diabetes than those in
7 the bottom 20%. The ratio of metabolic syndrome or prediabetes in the high CE group was also
8 significantly higher than in the low CE group. Therefore, there is a possibility that the deep
9 learning algorithm did not fail to predict due to error; rather, this algorithm may have classified
10 participants who are at high risk of diabetes but have not yet developed overt disease in the
11 ‘diabetes’ group. However, our verification of the increased risk of overt diabetes in this group
12 achieved only borderline significance, mainly due to the limited incidence of diabetes in UK
13 Biobank follow-up data. Future longitudinal follow-up studies including larger samples are
14 needed to clarify this hypothesis.

15 The results of this study should be interpreted with the following limitations. First,
16 since the external dataset included only cases (type 2 diabetes) and we could not evaluate AUC
17 with this dataset alone, we verified our result with a mixed internal and external validation
18 dataset. In addition, participants in the UK Biobank are mostly European descent and middle-
19 aged; thus, our results may not be generalizable to other ethnicities or a younger older
20 population. However, a retinal funduscopy exam is generally used to screen for retinopathy in
21 patients with diabetes, making it difficult to obtain a balanced large-scale retinal funduscopy
22 dataset that includes sufficient patients without diabetes. Moreover, very few large-scale
23 datasets exist that include comprehensive genetic, clinical, and funduscopy image data, except

1 for the UK Biobank dataset. Second, the diabetes prevalence among participants in the UK
2 Biobank funduscopy image dataset was low. UK Biobank participants were recruited from the
3 UK general population, rather than a diabetes population that needs regular screening for
4 diabetic retinopathy (27). The prevalence of diabetes in the UK from 2009-2010, when the
5 retinal images were collected, was reported to be approximately 5%; this is comparable with
6 the diabetes prevalence in our study population (18). Because the prevalence of type 2 diabetes
7 in our UK Biobank dataset was low, our algorithm exhibited high negative predictive value and
8 low positive predictive value. There is a clear need for larger, more diverse populations in
9 studies using retinal image recognition with deep learning models for predicting type 2 diabetes.
10 Third, some of the information used in this study, such as lifestyle and family history, comes
11 from participant self-reports, which may be prone to recall bias. Finally, although we tried to
12 aggregate all available information in the UK Biobank dataset, such as questionnaires, medical
13 records, medication history, and laboratory findings, misclassification bias may possibly have
14 occurred.

15 In conclusion, we confirmed that a deep learning algorithm using only retinal fundus
16 photographs showed considerable performance in the detection of patients with diabetes. Our
17 results suggest a utility for funduscopy examination in combination with a deep learning model
18 for type 2 diabetes screening in the general population. AI technology can lower the cost burden
19 and increase the productivity of screening programs or monitoring in populations with high
20 risk of diabetes. This is critical in light of the growing prevalence of type 2 diabetes; we expect
21 that an AI algorithm can reduce the burden of simple screening for the disease, allowing
22 medical practitioners to focus on more complicated aspects of care. While this technique for
23 the detection of diabetes still needs further research and external validation in various situations,

1 based on the results reported herein, evaluating funduscopy images with a deep learning model
2 is a promising approach that would enable efficient screening for multiple complications
3 related to diabetes as well as for diabetes itself.

4 5 **Acknowledgments**

6 This work was supported by the National Research Foundation of Korea Grant funded by the
7 Korean Government (NRF-2016R1C1B1009262) and the National Research Foundation of
8 Korea Grant (NRF-2019R1A2C1006608) funded by the Korea government. This work was
9 also supported by NLM R01 NL012535 and NIGMS R01 GM138597.

10 11 **Author contributions**

12 J.-S.Y. is the guarantor of this work, as such, had full access to all the data in the study and
13 takes responsibility for the integrity of the data and the accuracy of the data analysis, and
14 developed the study design, wrote the manuscript, analyzed and interpreted data. J.K. analyzed
15 data, wrote the manuscript, interpreted data and contributed to discussion. S.-H.J. analyzed data,
16 interpreted data and contributed to discussion. S.-A.C., Y.-B.A., S.-H.K., H.-H.W. revised the
17 manuscript, contributed to discussion. K.-A.S. and D.K. interpreted results, revised the
18 manuscript, contributed to discussion.

19 20 **Duality of Interest**

21 No potential conflicts of interest relevant to this article were reported.

22

1 Reference

- 2 1. International Diabetes Federation. IDF Diabetes Atlas. 8th ed.: International Diabetes
3 Federation; 2017.
- 4 2. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care* 2018;41:917-928
- 5 3. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes 2021.
6 *Diabetes Care* 2021;44:S15-s33
- 7 4. Shahraz S, Pittas AG, Saadati M, Thomas CP, Lundquist CM, Kent DM. Change in Testing,
8 Awareness of Hemoglobin A1c Result, and Glycemic Control in US Adults, 2007-2014. *JAMA*
9 2017;318:1825-1827
- 10 5. Jung CH, Son JW, Kang S, Kim WJ, Kim HS, Kim HS, Seo M, Shin HJ, Lee SS, Jeong SJ,
11 Cho Y, Han SJ, Jang HM, Rho M, Lee S, Koo M, Yoo B, Moon JW, Lee HY, Yun JS, Kim SY,
12 Kim SR, Jeong IK, Mok JO, Yoon KH. Diabetes Fact Sheets in Korea, 2020: An Appraisal of
13 Current Status. *Diabetes Metab J* 2021;45:1-10
- 14 6. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer
15 L, Keane PA, Wong TY. Artificial intelligence and deep learning in ophthalmology. *Br J*
16 *Ophthalmol* 2019;103:167-175
- 17 7. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R,
18 San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein
19 EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng
20 CY, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY. Development and Validation of a Deep
21 Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images
22 From Multiethnic Populations With Diabetes. *JAMA* 2017;318:2211-2223
- 23 8. Sabanayagam C, Xu D, Ting DSW, Nusinovici S, Banu R, Hamzah H, Lim C, Tham YC,

- 1 Cheung CY, Tai ES, Wang YX, Jonas JB, Cheng CY, Lee ML, Hsu W, Wong TY. A deep
2 learning algorithm to detect chronic kidney disease from retinal photographs in community-
3 based populations. *Lancet Digit Health* 2020;2:e295-e302
- 4 9. Williams BM, Borroni D, Liu R, Zhao Y, Zhang J, Lim J, Ma B, Romano V, Qi H, Ferdousi
5 M, Petropoulos IN, Ponirakis G, Kaye S, Malik RA, Alam U, Zheng Y. An artificial
6 intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using
7 corneal confocal microscopy: a development and validation study. *Diabetologia* 2020;63:419-
8 430
- 9 10. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster
10 DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning.
11 *Nat Biomed Eng* 2018;2:158-164
- 12 11. Son J, Shin JY, Kim HD, Jung KH, Park KH, Park SJ. Development and Validation of Deep
13 Learning Models for Screening Multiple Abnormal Findings in Retinal Fundus Images.
14 *Ophthalmology* 2020;127:85-94
- 15 12. Yun JS, Lim TS, Cha SA, Ahn YB, Song KH, Choi JA, Kwon J, Jee D, Cho YK, Park YM,
16 Ko SH. Clinical Course and Risk Factors of Diabetic Retinopathy in Patients with Type 2
17 Diabetes Mellitus in Korea. *Diabetes Metab J* 2016;40:482-493
- 18 13. Rim TH, Lee G, Kim Y, Tham YC, Lee CJ, Baik SJ, Kim YA, Yu M, Deshmukh M, Lee
19 BK, Park S, Kim HC, Sabayanagam C, Ting DSW, Wang YX, Jonas JB, Kim SS, Wong TY,
20 Cheng CY. Prediction of systemic biomarkers from retinal photographs: development and
21 validation of deep-learning algorithms. *Lancet Digit Health* 2020;2:e526-e536
- 22 14. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, Greenlund
23 K, Daniels S, Nichol G, Tomaselli GF, Arnett DK, Fonarow GC, Ho PM, Lauer MS, Masoudi
24 FA, Robertson RM, Roger V, Schwamm LH, Sorlie P, Yancy CW, Rosamond WD. Defining

- 1 and setting national goals for cardiovascular health promotion and disease reduction: the
2 American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation*
3 2010;121:586-613
- 4 15. He KZ, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. 2016 IEEE
5 Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA 2015
- 6 16. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. Prediction of
7 incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern*
8 *Med* 2007;167:1068-1074
- 9 17. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES,
10 Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases
11 identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 2018;50:1219-
12 1224
- 13 18. Gerrits N, Elen B, Craenendonck TV, Triantafyllidou D, Petropoulos IN, Malik RA, De
14 Boever P. Age and sex affect deep learning prediction of cardiometabolic risk factors from
15 retinal images. *Sci Rep* 2020;10:9432
- 16 19. Lee WJ, Sobrin L, Lee MJ, Kang MH, Seong M, Cho H. The relationship between diabetic
17 retinopathy and diabetic nephropathy in a population-based study in Korea (KNHANES V-2,
18 3). *Invest Ophthalmol Vis Sci* 2014;55:6547-6553
- 19 20. Paneni F, Beckman JA, Creager MA, Cosentino F. Diabetes and vascular disease:
20 pathophysiology, clinical consequences, and medical therapy: part I. *Eur Heart J* 2013;34:2436-
21 2443
- 22 21. Rask-Madsen C, King GL. Vascular complications of diabetes: mechanisms of injury and
23 protective factors. *Cell Metab* 2013;17:20-33
- 24 22. Cheung CY, Zheng Y, Hsu W, Lee ML, Lau QP, Mitchell P, Wang JJ, Klein R, Wong TY.

- 1 Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology*
2 2011;118:812-818
- 3 23. Nguyen TT, Wang JJ, Wong TY. Retinal vascular changes in pre-diabetes and
4 prehypertension: new findings and their research and clinical implications. *Diabetes Care*
5 2007;30:2708-2715
- 6 24. Ikram MK, Sim X, Jensen RA, Cotch MF, Hewitt AW, Ikram MA, Wang JJ, Klein R, Klein
7 BE, Breteler MM, Cheung N, Liew G, Mitchell P, Uitterlinden AG, Rivadeneira F, Hofman A,
8 de Jong PT, van Duijn CM, Kao L, Cheng CY, Smith AV, Glazer NL, Lumley T, McKnight B,
9 Psaty BM, Jonasson F, Eiriksdottir G, Aspelund T, Harris TB, Launer LJ, Taylor KD, Li X,
10 Iyengar SK, Xi Q, Sivakumaran TA, Mackey DA, Macgregor S, Martin NG, Young TL, Bis
11 JC, Wiggins KL, Heckbert SR, Hammond CJ, Andrew T, Fahy S, Attia J, Holliday EG, Scott
12 RJ, Islam FM, Rotter JI, McAuley AK, Boerwinkle E, Tai ES, Gudnason V, Siscovick DS,
13 Vingerling JR, Wong TY. Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the
14 microcirculation in vivo. *PLoS Genet* 2010;6:e1001184
- 15 25. Jensen RA, Sim X, Smith AV, Li X, Jakobsdóttir J, Cheng CY, Brody JA, Cotch MF,
16 McKnight B, Klein R, Wang JJ, Kifley A, Harris TB, Launer LJ, Taylor KD, Klein BE, Raffel
17 LJ, Li X, Ikram MA, Klaver CC, van der Lee SJ, Mutlu U, Hofman A, Uitterlinden AG, Liu C,
18 Kraja AT, Mitchell P, Gudnason V, Rotter JI, Boerwinkle E, van Duijn CM, Psaty BM, Wong
19 TY. Novel Genetic Loci Associated With Retinal Microvascular Diameter. *Circ Cardiovasc*
20 *Genet* 2016;9:45-54
- 21 26. Cover TM, Thomas JA. *Elements of Information Theory* 2006
- 22 27. Diabetes in the UK 2010: Key statistics on diabetes. [https://www.diabetes.org.uk/resources-](https://www.diabetes.org.uk/resources-s3/2017-11/diabetes_in_the_uk_2010.pdf)
23 [s3/2017-11/diabetes_in_the_uk_2010.pdf](https://www.diabetes.org.uk/resources-s3/2017-11/diabetes_in_the_uk_2010.pdf). Accessed April 2021.

1

2 **Table legends**

3 Table 1. Baseline characteristics of the study population.

4 Table 2. Performance of deep learning algorithm and traditional risk factor models for the
5 prediction of type 2 diabetes.

6 Table 3. Net reclassification improvement with addition of the algorithm to traditional risk
7 factors, determined by 2000x bootstrapping with the validation set.

8 Supplemental Table S1. Detailed definitions of type 2 diabetes and cardiovascular disease.

9 Supplemental Table S2. Detailed definitions of lifestyle factors and lifestyle behavior.

10 Supplemental Table S3. Performance of the deep learning algorithm for prediction of
11 traditional risk factors of type 2 diabetes.

12 Supplemental Table S4. Comparison of clinical characteristics according to the level of cross-
13 entropy loss.

14 Supplemental Table S5. Model performance for the prediction of type 2 diabetes using internal
15 and external validation set

16 Supplemental Table S6. Comparison of model performance for the prediction of type 2 diabetes
17 and prediabetes.

18 Supplemental Table S7. Performance of the deep learning algorithm in predicting genetic risk
19 for type 2 diabetes.

20 Supplemental Table S8. Performance of the deep learning algorithm and genetic risk plus deep
21 learning algorithm for the prediction of type 2 diabetes

22

1 **Figure legends**

2 Figure 1. Performance comparison between each non-invasive traditional risk factor (TRF)
3 model, composite TRFs model, deep learning algorithm, and TRFs plus algorithm model for
4 the prediction of type 2 diabetes, determined by 2000x bootstrapping with the validation set.

5 Supplemental Figure S1. Summarization of the study design.

6 Supplemental Figure S2. UMAP visualization of image representations in the last hidden layer
7 of the deep learning algorithm. Red colored points with 2D contour histogram represent type 2
8 diabetes. Blue colored points represent non-diabetes with the three groups of per-sample cross-
9 entropy loss (High: top 20 percentile; Intermediate: 20-79th percentile; Low: bottom 20
10 percentile).

11 Supplemental Figure S3. Representative randomly-selected examples of quality assessment. P
12 indicates the probability of poor quality given an image ($P(\text{poor}|\text{image})$).

13

14

Tables & Figures

1

2 **Table 1. Baseline characteristics of the study population.**

	Total (N=62262)	Training set (N=37904)	Tuning set (N=12173)	Validation set (N=12185)	<i>P</i> value
Age	56.2 ± 8.2	56.2 ± 8.2	56.2 ± 8.1	56.2 ± 8.2	0.759
<60 years	33291 (53.5)	20345 (53.7)	6497 (53.4)	6449 (52.9)	
≥60 years	28971 (46.5)	17559 (46.3)	5676 (46.6)	5736 (47.1)	
Sex					0.856
Women	34220 (55.0)	20819 (54.9)	6717 (55.2)	6684 (54.9)	
Men	28042 (45.0)	17085 (45.1)	5456 (44.8)	5501 (45.1)	
Race					0.555
White	50667 (81.4)	31471 (83.0)	10533 (86.5)	11142 (92.1)	
Asian	2010 (3.3)	1255 (3.3)	396 (3.3)	359 (3.0)	
Black	1751 (2.8)	1060 (2.8)	363 (3.0)	328 (2.7)	
Others	902 (1.5)	561 (1.5)	171 (1.4)	170 (1.4)	
Mixed	516 (0.8)	312 (0.8)	100 (0.8)	104 (0.9)	
Systolic blood pressure	139.7 ± 19.6	139.7 ± 19.6	139.7 ± 19.7	139.6 ± 19.6	0.839
<140 mmHg	33061 (53.1)	20139 (53.2)	6402 (52.6)	6520 (53.5)	
≥140 mmHg	29162 (46.9)	17741 (46.8)	5761 (47.4)	5660 (46.5)	
Diastolic blood pressure	82.0 ± 10.7	82.0 ± 10.7	81.9 ± 10.8	81.9 ± 10.6	0.894
<90 mmHg	47808 (76.8)	29117 (76.9)	9340 (76.8)	9351 (76.8)	
≥90 mmHg	14415 (23.2)	8763 (23.1)	2823 (23.2)	2829 (23.2)	
Body mass index	27.3 ± 4.8	27.3 ± 4.8	27.2 ± 4.7	27.3 ± 4.8	0.189
<30.0 kg/m ²	47199 (76.2)	28667 (76.0)	9301 (76.8)	9231 (76.1)	
≥30.0 kg/m ²	14741 (23.8)	9040 (24.0)	2803 (23.2)	2898 (23.9)	
Waist circumference	90.0 ± 13.5	90.1 ± 13.5	89.9 ± 13.3	90.0 ± 13.6	0.532
<102 cm (men) or <88 cm (women)	41211 (66.4)	25028 (66.2)	8105 (66.8)	8078 (66.5)	
≥102 cm (men) or ≥88 cm (women)	20864 (33.6)	12761 (33.8)	4026 (33.2)	4077 (33.5)	
Family history of diabetes	26773 (43.0)	16395 (43.3)	5163 (42.4)	5215 (42.8)	0.234
Lifestyle behavior					0.841
Unfavorable lifestyle	6694 (11.2)	4091 (11.3)	1319 (11.3)	1284 (11.0)	
Intermediate lifestyle	19938 (33.5)	12101 (33.4)	3890 (33.4)	3947 (33.9)	
Favorable lifestyle	32872 (55.2)	20025 (55.3)	6439 (55.3)	6408 (55.1)	
Fasting plasma glucose (mmol/L)	5.14 ± 0.98	5.14 ± 0.99	5.15 ± 0.99	5.13 ± 0.90	0.219
<7.0 mmol/L	53199 (97.6)	32343 (97.5)	10430 (97.8)	10426 (97.5)	
≥7.0 mmol/L	1334 (2.4)	826 (2.5)	238 (2.2)	270 (2.5)	
HbA1c (%)	5.4 ± 0.6	5.4 ± 0.6	5.4 ± 0.6	5.4 ± 0.6	0.376
HbA1c (mmol/mol)	35.9 ± 6.4	36.0 ± 6.6	35.9 ± 6.2	35.9 ± 6.2	0.376
<5.7%	45146 (85.1)	27442 (85.1)	8871 (85.5)	8833 (84.9)	
5.7-6.4%	8776 (15.6)	5374 (15.7)	1679 (15.3)	1723 (15.7)	
≥6.5%	1848 (3.3)	1135 (3.3)	351 (3.2)	362 (3.3)	
Prediabetes	7891 (14.9)	4818 (14.9)	1508 (14.5)	1565 (15.1)	0.516
Type 2 diabetes mellitus	2691 (4.8)	1677 (4.9)	504 (4.6)	510 (4.7)	0.297
Total cholesterol (mmol/L)	5.69 ± 1.13	5.69 ± 1.13	5.68 ± 1.12	5.71 ± 1.13	0.126
Triglyceride (mmol/L)	1.67 ± 0.96	1.67 ± 0.95	1.67 ± 0.98	1.66 ± 0.97	0.255
<1.70 mmol/L	35996 (62.9)	21726 (62.4)	7190 (64.1)	7080 (63.2)	
≥1.70 mmol/L	21231 (37.1)	13090 (37.6)	4026 (35.9)	4115 (36.8)	
HDL cholesterol (mmol/L)	1.48 ± 0.39	1.48 ± 0.39	1.48 ± 0.39	1.49 ± 0.39	0.073
≥1.03 mmol/L (men) or ≥1.29 mmol/L (women)	44312 (81.2)	26866 (80.9)	8776 (82.2)	8670 (81.0)	
<1.03 mmol/L (men) or <1.29 mmol/L (women)	10268 (18.8)	6330 (19.1)	1904 (17.8)	2034 (19.0)	
LDL cholesterol (mmol/L)	3.54 ± 0.86	3.53 ± 0.86	3.53 ± 0.85	3.55 ± 0.86	0.131
Cardiovascular disease	4149 (6.7)	2526 (6.7)	807 (6.6)	816 (6.7)	0.928

3 Data are n (%) or mean (SD).

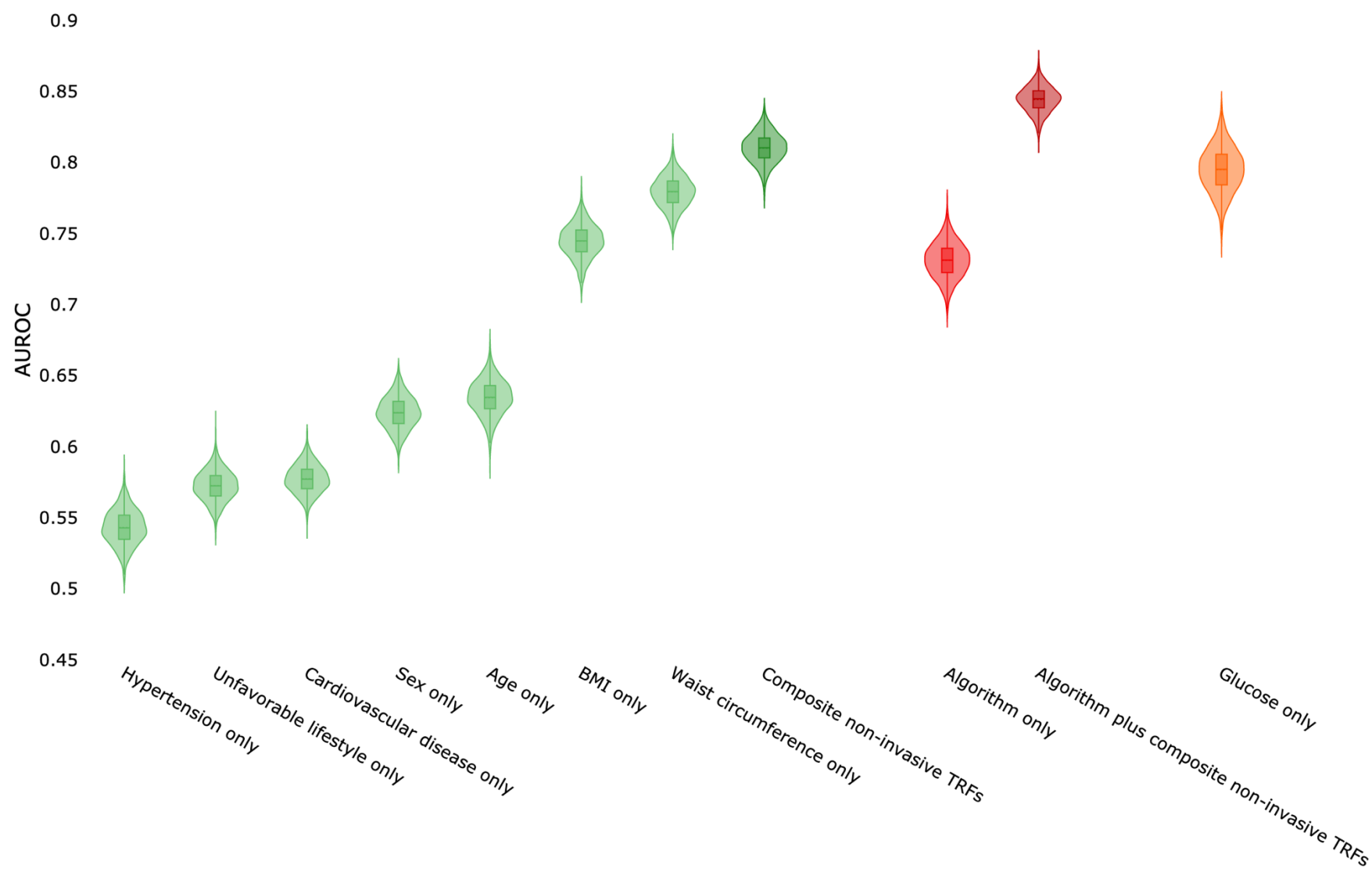
4

1 **Table 2. Performance of deep learning algorithm for the prediction of traditional risk factors for type 2 diabetes**

Predicted risk factor	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	R ² (95% CI) per 1 unit
Age (<60 vs. ≥60 years)	0.931 (0.928-0.934)	0.852 (0.847-0.856)	0.852 (0.847-0.856)	0.834 (0.827-0.840)	0.868 (0.863-0.873)	0.783 (0.778-0.789)
Sex (Women vs. Men)	0.933 (0.929-0.936)	0.854 (0.849-0.859)	0.854 (0.849-0.859)	0.826 (0.819-0.833)	0.878 (0.873-0.883)	NA
Hypertension (BP <140/90 mmHg vs. ≥140/90 mmHg)	0.763 (0.757-0.769)	0.693 (0.687-0.699)	0.693 (0.687-0.699)	0.689 (0.681-0.697)	0.698 (0.689-0.706)	Systolic BP 0.304 (0.294-0.314) Diastolic BP 0.233 (0.223-0.244)
Obese (BMI <30 kg/m ² vs. ≥30 kg/m ²)	0.648 (0.639-0.656)	0.604 (0.597-0.612)	0.604 (0.597-0.612)	0.322 (0.312-0.332)	0.831 (0.825-0.837)	0.074 (0.067-0.081)
Central obesity (WC <102 cm (men) or <88 cm (women) vs. ≥102 cm (men) or ≥88 cm (women))	0.628 (0.620-0.635)	0.590 (0.583-0.597)	0.590 (0.583-0.597)	0.418 (0.409-0.428)	0.741 (0.734-0.749)	0.144 (0.134-0.154)
Cardiovascular disease (made available under a CC-BY-NC-ND 4.0 International license)	0.623 (0.608-0.637)	0.623 (0.608-0.637)	0.623 (0.608-0.637)	0.105 (0.098-0.113)	0.959 (0.955-0.962)	NA
Unfavorable lifestyle (no vs. yes)	0.511 (0.498-0.523)	0.508 (0.498-0.517)	0.508 (0.498-0.517)	0.894 (0.888-0.899)	0.112 (0.107-0.118)	NA
Triglyceride (<1.70 mmol/L vs. ≥1.70 mmol/L)	0.546 (0.538-0.554)	0.534 (0.527-0.541)	0.534 (0.527-0.541)	0.399 (0.389-0.408)	0.664 (0.655-0.673)	-0.028 (-0.035- -0.021)
HDL cholesterol (≥1.03 mmol/L (men) or ≥1.29 mmol/L (women) vs. <1.03 mmol/L (men) or <1.29 mmol/L (women))	0.534 (0.524-0.545)	0.523 (0.514-0.532)	0.523 (0.514-0.532)	0.205 (0.197-0.214)	0.823 (0.815-0.830)	0.082 (0.073-0.091)
HbA1c (<5.7% vs. ≥5.7%)	0.648 (0.637-0.658)	0.600 (0.590-0.610)	0.600 (0.590-0.610)	0.201 (0.192-0.211)	0.899 (0.894-0.904)	0.061 (0.054-0.069)
HbA1c (<6.5% vs. ≥6.5%)	0.734 (0.715-0.752)	0.676 (0.656-0.693)	0.676 (0.656-0.693)	0.066 (0.059-0.073)	0.984 (0.982-0.986)	
Glucose (<7.0 mmol/L vs. ≥7.0 mmol/L)	0.672 (0.648-0.694)	0.631 (0.608-0.651)	0.631 (0.608-0.651)	0.043 (0.038-0.048)	0.985 (0.983-0.987)	-0.005 (-0.009- -0.001)

2 AUC, area under the curve; PPV, positive predictive value; NPV, negative predicted value; NA, not applicable; BP, blood pressure; WC, waist circumference

3
4



5

6 **Figure 1. Performance comparison between each non-invasive traditional risk factor (TRF) model, composite TRFs model, deep learning algorithm, and TRFs plus algorithm model for the**
7 **prediction of type 2 diabetes, determined by 2000x bootstrapping with the validation set.**

1 **Table 3. Net reclassification improvement with addition of the algorithm to traditional risk factors, determined by 2000x bootstrapping with the validation set.**

	TRFs plus algorithm model				TRFs plus algorithm model			
	All participants				Top 10 percentile high risk participants in TRFs model			
	Number	Individuals reclassified		NRI (%)	Number	Individuals reclassified		NRI (%)
		Increased (%)	Decreased (%)			Increased (%)	Decreased (%)	
Cases	419	51.5 (46.7-56.4)	48.5 (43.6-53.3)	3.0 (-6.7-12.8)	178	51.7 (36.1-67.4)	48.3 (32.6-63.9)	3.4 (-27.8-34.9)
Non-cases	8889	73.9 (72.9-74.8)	26.1 (25.2-27.1)	47.8 (45.8-49.5)	753	73.9 (70.8-76.8)	26.1 (23.2-29.2)	47.8 (41.6-53.5)
All	9308	-	-	50.8 (40.9-60.9)	931	-	-	51.2 (19.5-82.9)

2 TRF, traditional risk factor; NRI, net reclassification improvement

medRxiv preprint doi: <https://doi.org/10.1101/2021.06.29.21259606>; this version posted July 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).