



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Can deep learning on retinal images augment known risk factors for cardiovascular disease prediction in diabetes?

Citation for published version:

Mellor, J, Jiang, W, Fleming, A, McGurnaghan, SJ, Blackburn, L, Styles, C, Storkey, AJ, McKeigue, PM, Colhoun, HM & Scottish Diabetes Research Network Epidemiology Group 2023, 'Can deep learning on retinal images augment known risk factors for cardiovascular disease prediction in diabetes? A prospective cohort study from the national screening programme in Scotland', *International journal of medical informatics*, vol. 175, 105072, pp. 105072. <https://doi.org/10.1016/j.ijmedinf.2023.105072>

Digital Object Identifier (DOI):

[10.1016/j.ijmedinf.2023.105072](https://doi.org/10.1016/j.ijmedinf.2023.105072)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

International journal of medical informatics

Publisher Rights Statement:

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

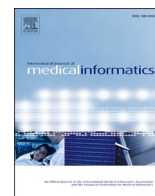
General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Can deep learning on retinal images augment known risk factors for cardiovascular disease prediction in diabetes? A prospective cohort study from the national screening programme in Scotland

Joseph Mellor^{b,*}, Wenhua Jiang^b, Alan Fleming^a, Stuart J. McGurnaghan^{a,b}, Luke Blackburn^a, Caroline Styles^e, Amos J. Storkey^c, Paul M. McKeigue^b, Helen M. Colhoun^{a,d}, on behalf of the Scottish Diabetes Research Network Epidemiology Group

^b The Usher Institute, University of Edinburgh, Edinburgh, UK

^a The Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

^e Queen Margaret Hospital, Dunfermline, Fife, UK

^c School of Informatics, University of Edinburgh, Edinburgh, UK

^d Department of Public Health, NHS Fife, Kirkcaldy, UK

ARTICLE INFO

Keywords:

Type 1 diabetes
Type 2 diabetes
Cardiovascular
Risk prediction
Deep learning

ABSTRACT

Aims: This study's objective was to evaluate whether deep learning (DL) on retinal photographs from a diabetic retinopathy screening programme improve prediction of incident cardiovascular disease (CVD).

Methods: DL models were trained to jointly predict future CVD risk and CVD risk factors and used to output a DL score. Poisson regression models including clinical risk factors with and without a DL score were fitted to study cohorts with 2,072 and 38,730 incident CVD events in type 1 (T1DM) and type 2 diabetes (T2DM) respectively. **Results:** DL scores were independently associated with incident CVD with adjusted standardised incidence rate ratios of 1.14 ($P = 3 \times 10^{-04}$ 95 % CI (1.06, 1.23)) and 1.16 ($P = 4 \times 10^{-33}$ 95 % CI (1.13, 1.18)) in T1DM and T2DM cohorts respectively. The differences in predictive performance between models with and without a DL score were statistically significant (differences in test log-likelihood 6.7 and 51.1 natural log units) but the increments in C-statistics from 0.820 to 0.822 and from 0.709 to 0.711 for T1DM and T2DM respectively, were small.

Conclusions: These results show that in people with diabetes, retinal photographs contain information on future CVD risk. However for this to contribute appreciably to clinical prediction of CVD further approaches, including exploitation of serial images, need to be evaluated.

1. Introduction

Cardiovascular disease (CVD) is one of the leading causes of death worldwide with type 1 and type 2 diabetes as known risk factors and thus many efforts have been made to improve CVD risk prediction [1].

Previous studies have shown that characteristics of the retinal vasculature observed by retinal fundus imaging contain valuable information regarding cardiovascular health. For example, arteriolar and venular widths and tortuosity [2], venular occlusions [3], vascular caliber [4], and a combination of various retinal information [5] have shown strong association with the incidence of future cardiovascular events. Studies in people with diabetes included relatively small cohort

sizes.

It has been shown that a variety of traditional CVD risk factors such as age, sex, and smoking status can be predicted using deep learning (DL) on retinal fundus images [6]. Recently, efforts have been made to assess whether DL predictors of intermediate disease risk markers such as carotid artery atherosclerosis [7], the presence of coronary artery calcium (CAC) [8–10], and of retinal-vessel features [11,12], learned from retinal photographs predict CVD events. However in these studies in the general population, there was little increment in prediction of CVD risk by use of DL models of these intermediate risk markers or factors.

There is considerable evidence that retinopathy and CVD share

* Corresponding author.

E-mail address: joe.mellor@ed.ac.uk (J. Mellor).

<https://doi.org/10.1016/j.ijmedinf.2023.105072>

Received 30 October 2022; Received in revised form 20 February 2023; Accepted 12 April 2023

Available online 18 April 2023

1386-5056/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

common risk factors in diabetes so that predictive information on CVD might be expected from retinal images. In many countries including Scotland screening programmes exist so that retinal photographs may provide a routinely available source of information. Therefore the aim of this prospective cohort study was to examine whether DL predictors (DLP) applied to retinal images could improve the prediction of CVD risk compared with baseline models using clinical covariates, in people with type 1 diabetes (T1DM) and type 2 diabetes (T2DM).

2. Material and methods

2.1. Scottish cohort

The cohorts were constructed using the Scottish Diabetes Research Network dataset (SDRN-NDS [13]) that linked all fundus images between 2005 and 2017 from the Scottish Diabetic Retinopathy Screening (SDRS) programme to a national register of all people with diabetes in Scotland maintained by Scottish Care Information - Diabetes Care (SCI-DC) for primary care data. Data was also linked to Scottish Morbidity Records (SMR) for out- and in-patient records, and to the General Register Office (GRO) for Scotland for death records.

The start date and end date of this study were 1 January 2008 and 1 January 2018. For each individual, the study entry was defined as the latest of: study start date, date of diabetes diagnosis, and date of 18th birthday or date of the individual's first gradable diabetic retinopathy screening episode. An individual's exit date from the study was defined as the earliest of study end date, date of death, date of incident CVD event or ceasing to be under observation in the national register.

The photographic protocol used by SDRS specifies a single fundus photograph from each eye showing the macula and optic disc. A variety of non-mydratric 45 degree fundus cameras were used. For each individual, this study used only images from the baseline retinal screening which was defined as the nearest gradable screening episode before or on the entry date.

CVD events, which constituted the gold-standard outcome for our risk prediction task, were obtained from SMR and GRO. We defined CVD events as any hospital admission or death due to the following conditions; myocardial infarction, stroke, unstable angina, transient ischaemic attack, peripheral vascular disease or acute coronary heart disease, or the following procedures; coronary, carotid, or peripheral artery revascularisations or major associated amputation. International Classification of Diseases version 10 (ICD-10) codes and Office of Population Censuses and Surveys Classification of Interventions and Procedures (OPCS-4) codes within this definition are given in the [supplementary materials](#).

The inclusion criteria were T1DM diagnosed before 50 years of age and T2DM diagnosed between 18 and 100 years of age, and in both cases a gradable screening episode after the age of 12. Subjects were excluded if they had a CVD event prior to their study entry date.

2.2. Candidate covariates

For the T1DM cohort, we included risk factors reported in previous studies including the average HbA1c in the preceding years [14–16]. For the T2DM cohort, we used the same risk factors with the addition of ethnicity and prior drug counts.

Baseline measurements and prescribing data were defined as prior measurements nearest to the entry date but no more than 24 months before that date. All measurements were defined at baseline apart from current age which was time-updated at the beginning of each person-time interval. Covariates with 60 % or more missingness were excluded from the analyses. [Supplementary material](#) contains further details about clinical covariate definitions.

2.3. Deep learning model development

Patients with bilateral gradable retinal fundus images at baseline were used to develop the DL model. For each cohort, we divided the dataset into a training set, validation set, and a test set by patient level in a 50:20:30 ratio. A total of 11,910 and 101,512 bilateral pairs of images were used in the training process of the DL model for T1DM and T2DM, respectively. Each set had a similar proportion of patient with CVD events. The outcomes of the DL model were whether there was a future CVD event, the current DR grade according to SDRS, estimated glomerular filtration rate (eGFR), and systolic blood pressure (SBP). For DL training, CVD followup was split into 5 year intervals - censored at study exit - and for each interval the outcome was a binary variable indicating if any CVD event was observed during the interval. Additionally, the DR outcome was the maximum retinopathy grade (Scottish grading scheme R0-R4) of both eyes; the eGFR outcome was a categorical variable with 4 values "<30 ml/min/1.73 m² or RRT", "30–60 ml/min/1.73 m²", ">60–90 ml/min/1.73 m²", ">90 ml/min/1.73 m²"; and the SBP outcome was a categorical variable with 3 values "<130 mmHg", "130–160 mmHg", ">160 mmHg". A ResNet-101 network architecture (pretrained on ImageNet) was trained to jointly predict the 4 outcomes. To provide a single prediction from bilateral fundus image inputs a multiple-instance learning (MIL) head, as used in [17], was added to the ResNet-101 immediately after the final global average pooling layer, replacing the final fully-connected layer. The MIL module used 4 heads each of dimension 128. Proceeding the MIL module were 4 linear layers, one for each of the 4 outcomes. The CVD outcome linear layer took as input both the MIL output and a binary indicator which was 0 for the first 5 year interval and 1 for the second 5 year interval for each patient. The network structure is illustrated in [Fig. 1](#).

The training objective for the network was a weighted sum of the cross-entropy loss for each of the 4 outcomes (CVD, DR, SBP, eGFR). The CVD loss had a weight of 1, and the remaining losses had a weight of 0.02. The ResNet101 was trained using stochastic gradient descent with a momentum of 0.9 for 100 epochs. The initial learning rate was 0.02 and was reduced at each subsequent epoch using a cosine-annealing learning rate schedule. Mini-batches of size 368 were equally balanced between examples with follow-up CVD and those without. Mixed-precision training was used on a DGX-1 machine using 8 32 GB V100s. Every 10 epochs a Poisson model including age, duration of diabetes, sex, and DL CVD predictions was fitted to 1 year person-time interval validation data and its log-likelihood (LL) calculated. The DL model weights at the epoch with the highest validation LL for the Poisson model became the final model weights. Fundus images were preprocessed to remove black borders and during training were augmented randomly by flip, rotation, resize, cropping, colour jitter, and gridded cutout. Processed images had dimension 448 × 448.

The DL CVD prediction is then included in the baseline Poisson regression model containing known CVD risk factors. We compared the predictive performance between baseline models (a restricted model including age, diabetes duration, and sex and a full baseline using further known risk factors) and Poisson models that included the DL CVD predictor. The pictogram representing the training and evaluation pipeline is illustrated in [Fig. 2](#).

2.4. Statistical analyses

Missing data were imputed [18,19]. An average over 10 imputation runs was used as the imputed values. Poisson risk prediction models to predict incident CVD were fit to observations constituting one year person-time intervals. Two baseline models were used: a restricted base Poisson model including the baseline and time-updated age, sex and baseline diabetes duration and an offset term for intervals in which censoring occurred and a full baseline model using forward selection to add further risk factors until the AIC did not fall by at least the number of extra parameters. Predictors with a skewed distribution (eGFR, Total:

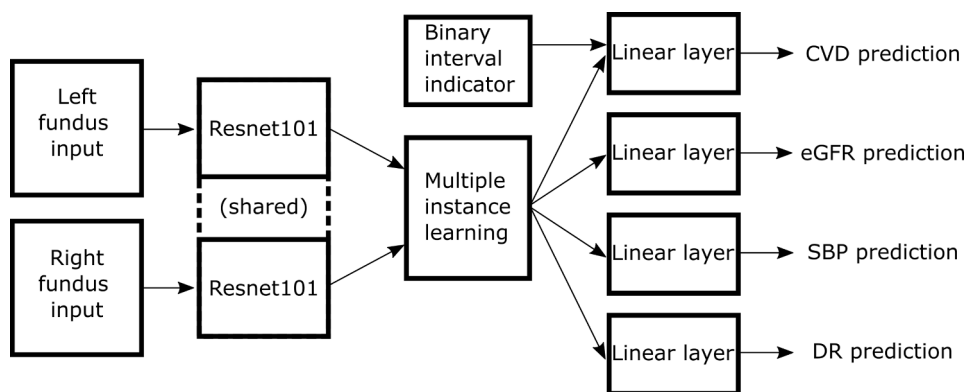


Fig. 1. Diagram visualising the structure of the proposed deep learning model. A single ResNet101 is used to process both left and right fundus images. The ResNet101 outputs are then input into a Multiple Instance Learning module. Those outputs are then input to linear layers - one for each outcome - to determine the predictions for those outcomes.

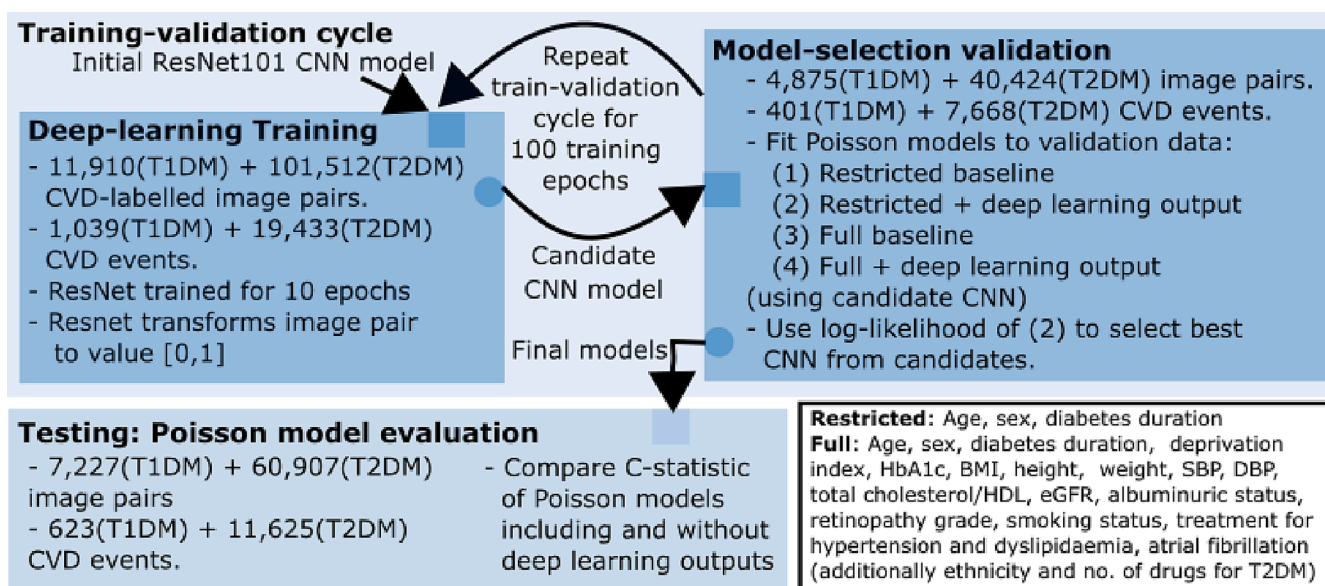


Fig. 2. Pictogram visualising the proposed deep learning training and evaluation pipeline.

HDL cholesterol ratio and BMI) were log-transformed. Quadratic and cubic terms were entered for age. Interactions between candidate covariates and age and sex were considered for inclusion in the model.

Predictive performance was examined via the test set. Patients whose data were included in the training or validation procedures of the DL model were excluded from the test set. The increments in discrimination achieved by Poisson regression models that included DLP compared to the baseline models were quantified using the C-statistic (also known as the area under the receiver operator characteristic curve, or AUC) and the expected information for discrimination, Λ [20]. Increments in Λ are interpretable in absolute units whereas increments in C-statistic are not. The strength of evidence that the final model improved the predictive performance on top of the base model was assessed by the increment in test LL; a difference in test LL of 6.7 natural log units is asymptotically equivalent to a p-value < 0.005 for comparison of nested models [21].

3. Results

Using Scottish health records and retinal images from 24,012 and 202,843 people with T1DM and T2DM, respectively, from the SDRS programme, DLP of CVD and of the intermediate CVD risk factors eGFR, SBP, and DR were jointly trained. We evaluated if these DLPs could improve the predictive performance for CVD risk – in people with T1DM

and T2DM when compared with Poisson regression models using clinical covariates of known risk factors. Data between 2008 and 2018 included 2,072 and 38,730 incident CVD events during 172,481 and 1,273,785 person-years of follow up for T1DM and T2DM respectively.

The most strongly associated risk factor of CVD was HbA1c for T1DM and age at study entry for T2DM. Supplementary Table B.2 shows many common risk factors - examined separately and adjusted for age, sex, and diabetes duration - display significant associations with CVD. Full details of missingness are shown in Supplementary Table B.3 where only albuminuria grade and HDL cholesterol had >10 % missingness. Supplementary Table B.4 shows the age-standardised rates of CVD.

There was a highly statistically-significant independent association between CVD DLP and incident CVD; standardised incidence rate ratios for the CVD DLP after adjusting for all clinical risk factors, estimated using data from the test set not used to train the DL model, were 1.14 ($P = 3 \times 10^{-04}$ 95 % CI (1.06, 1.23)) and 1.16 ($P = 4 \times 10^{-33}$ 95 % CI (1.13, 1.18)) in T1DM and T2DM cohorts respectively.

These were similar in magnitude to the adjusted standardised IRRs for smoking in T1DM (Standardised IRR 1.19P = 2×10^{-04} 95 % CI (1.09, 1.31)) and smoking in T2DM (Standardised IRR 1.15P = 5×10^{-40} 95 % CI (1.12, 1.17)). For comparison the standardised incidence rate ratios for HbA1c after adjusting for all clinical risk factors on the same data were 1.63 ($P = 1 \times 10^{-23}$ 95 % CI (1.48, 1.79)) and 1.14 ($P =$

6×10^{-45} 95 % CI (1.12, 1.17)) in T1DM and T2DM cohorts respectively. We found no interaction effect between follow-up time and the CVD DLP in either cohort.

In the T1DM cohort, the addition of the CVD DLP to the restricted baseline model increased the LL by 8.5 natural log units, and the C-statistic increased from 0.816 to 0.820. For the full clinical model, the increment in LL when adding the CVD DLP was 6.7 natural log units and the C-statistic went from 0.820 to 0.822. Predictive performance of CVD in people with T2DM was lower than for T1DM. Addition of the CVD DLP to the restricted clinical model in the T2DM cohort increased the C-statistic from 0.707 to 0.709 and increased LL by 64.3 natural log units. For the full clinical model it increased the C-statistic from 0.709 to 0.711 and increased LL by 51.1 natural log units. In both cases the CVD DLP increased test LL, while increase in C-statistic was small. Results are shown in Table 1.

Despite clear association of DLPs with CVD, of similar magnitude to that seen for smoking, the addition of DLPs to baseline models did not increase C-statistic substantially even when stratifying either by age or by sex (see Supplementary Table B.5). Adding the eGFR, SBP, and DR DLPs individually to baseline models resulted in very similar small increases in C-statistic (see Supplementary Table B.6). ROC curve comparisons between predictions with and without DL CVD are shown in Supplementary Fig. B.1 and calibration plots for prediction models using DL CVD are shown in Supplementary Fig. B.2. Further performance measures - including net classification improvement, sensitivity, specificity, and F1 score - comparing baseline models to models including DL CVD are shown Table 2. Supplementary Table B.7 shows the performance of the full model excluding the clinical risk factors SBP, log eGFR, log total:HDL cholesterol ratio, and HbA1c (a “near-full” model) and the performance of this model with the addition of each of these risk factors separately and Supplementary Table B.8 gives standardised incident risk ratios adjusted for age, sex, and diabetes duration evaluated on the T1DM and T2DM datasets.

4. Discussion

We have shown that though DL scores are predictive of CVD in models with age, sex, and duration only, the increment in predictive performance when adding DL scores to a model with other clinical risk factors is small, although statistically significant. Stratification by age and sex did not show any subgroup in which the increment in predictive performance was larger.

Our findings are consistent with previous studies. Poplin et al. [6] compared both the SCORE risk prediction model and a restricted prediction model including age, sex, BMI, SBP and smoking status for prediction of 5 year major adverse cardiovascular events (MACE) in the UK Biobank (UKBB) cohort to models that included DL-based predictions

from fundus images. They showed no evidence of predictive improvement over SCORE by including DL predictions. Inclusion of DL predictions to their restricted model yielded a small increase in C-statistic from 0.72 to 0.73. Only 5 % of UKBB participants in the study had diabetes, whereas our study evaluated prediction in T1DM and T2DM cohorts. Since in diabetes the pathogenesis of CVD and retinopathy are partly shared (both being strongly influenced by glycaemia and blood pressure) we may expect that DL from fundus images are more predictive in diabetes cohorts. However we do not find evidence for this.

Readers may be surprised that the predictive performance of CVD risk models is relatively modest: even in T1DM, the best models have a C-statistic of only about 0.8, equivalent to information for discrimination, Λ , of only 1.0 bit. However, this is within the range that we would expect, given that CVD risk is determined by a relatively small number of independent risk factors of modest effect size. For established CVD risk factors such as blood cholesterol the rate ratio associated with increment of one standard deviation of a single risk factor is typically about 1.5, which corresponds approximately to an increment of 0.12 bits in Λ , or an increment from a baseline model with C-statistic of 0.75 (equivalent to Λ of 0.66 bits) to a C-statistic of 0.77. Thus even what are regarded as strong risk factors of CVD only provide a modest increment to prediction of CVD risk. For weaker risk factors with standardised rate ratios of 1.2 the approximate increment in Λ is even smaller (0.02 bits) and the typical C-statistic after inclusion of this risk factor would still only be 0.75. For stronger risk factors with standardised rate ratios of 2.5 (increment of 0.61 bits) it would be from 0.75 to 0.83. Increments in C-statistic are not easily interpretable and so our paper uses increments in Λ . Information in retinal images relevant to CVD prediction is, in part, already captured by other risk factors and so, after adjustment for these risk factors, the standardised rate ratios for CVD DLP are similar to smoking status. A good clinical predictor may have a Λ of 3.0 bits (approximately a C-statistic of 0.93) [20]. Given our full baseline model for T1DM a single independent risk factor would require at least an Λ of 1.74 bits corresponding to a standardised rate ratio of 4.7, well above that of many risk factors including CVD DLP.

A major contribution of our analysis is that it has focused on diabetes, had a large sample size and a large number of incident events. To our knowledge this is the first study to examine the performance of DL models using fundus images applied to CVD risk in such cohorts. This study has been informed by the IJMI checklist [22]. There are limitations to our study. We used only one image at entry whereas serial images are available from SDRS. A series of images may be more informative for DL prediction. For instance, monitoring of changes and their location (via image registration) may be more informative than is possible with a single fundus image. Second, our DL models were trained to predict incident CVD and just three CVD risk factors. This is similar to previous studies that trained DL models to predict multiple CVD risk

Table 1
Performance of baseline and deep learning models in Scottish T1DM and T2DM test datasets.

Model	T1DM			T2DM		
	C-statistic (AUC)	Λ	ΔLL	C-statistic (AUC)	Λ	ΔLL
Restricted baseline						
Restricted baseline	0.750(0.732,0.768)	0.69	0.0	0.671(0.666,0.676)	0.28	0.0
+ all DLPs	0.767(0.749,0.784)	0.77	24.8	0.678(0.674,0.683)	0.31	167.5
+ DL CVD	0.764(0.747,0.782)	0.74	23.3	0.678(0.673,0.683)	0.31	148.0
+ HbA1c	0.795(0.778,0.811)	0.96	112.6	0.676(0.671,0.681)	0.31	131.9
+ log eGFR	0.750(0.733,0.768)	0.70	5.1	0.673(0.668,0.678)	0.29	74.6
+ SBP	0.751(0.733,0.769)	0.69	1.2	0.672(0.667,0.676)	0.29	12.0
+ retinopathy	0.769(0.752,0.786)	0.77	39.6	0.675(0.670,0.680)	0.30	94.6
+ log total:HDL cholesterol ratio	0.766(0.748,0.783)	0.77	36.5	0.677(0.672,0.681)	0.30	133.4
Full baseline						
Full baseline	0.820(0.805,0.835)	1.26	0.0	0.709(0.705,0.714)	0.43	0.0
+ all DLPs	0.821(0.806,0.837)	1.28	0.0	0.711(0.706,0.715)	0.44	56.6
+ DL CVD	0.822(0.807,0.838)	1.28	6.7	0.711(0.706,0.715)	0.44	51.1

Λ is the expected information for discrimination measured in bits. ΔLL is given with respect to the appropriate baseline model (that does not include any DLP). 95 % confidence intervals are given in brackets for C-statistic/AUC.

Table 2

Continuous NRI, NRI, Sensitivity, Specificity, PPV, and F1 score when adding DL CVD to baseline models for Scottish T1DM and T2DM test datasets. We set the decision threshold as twice the mean person-interval risk with respect to the baseline models. 95 % confidence intervals are given for NRI.

Model	Continuous NRI	NRI	Sensitivity	Specificity	PPV	F1
T1DM Restricted model	–	–	0.876	0.387	0.992	0.931
T1DM Restricted + DL CVD	0.337	0.024(–0.009,0.057)	0.884	0.403	0.993	0.935
T1DM Full model	–	–	0.864	0.557	0.994	0.925
T1DM Full + DL CVD	0.085	0.002(–0.015,0.021)	0.867	0.557	0.994	0.926
T2DM Restricted model	–	–	0.929	0.184	0.976	0.952
T2DM Restricted + DL CVD	0.176	0.011(0.006,0.016)	0.921	0.203	0.976	0.948
T2DM Full model	–	–	0.908	0.273	0.978	0.942
T2DM Full + DL CVD	0.122	0.006(0.002,0.01)	0.905	0.282	0.978	0.940

factors simultaneously [6]. In this study it is an example of multi-task learning where the CVD risk factors take the role of related auxiliary tasks to aid training of the primary task [23]. If tasks are similar but with independent signal noise, then learning the tasks jointly can increase the effective sample size due to the combined information in all tasks [24]. However, training against other CVD risk factors may lead to DL models which improve the overall predictive performance of CVD. Thirdly, the removal of ungradable images meant that the gradability of retinal images, which may be affected by ocular opacity, was not taken into account although there are known links between cataract and CVD [25]. We do not externally validate our models in this study, although we did not find evidence that improvements in prediction of CVD using DLPs were large enough to necessitate it at this time. Finally improvements to DL architecture within a similar DL pipeline to our own may provide performance improvements, but we expect they would not produce the standardised rate ratio we state above.

There has been much optimism about the improvement DL using fundus images might make to risk prediction of complications of diabetes such as CVD [26]. Our findings suggest that although there is information relevant to CVD within fundus images increment in prediction for decision making is modest. However, small improvements in predictive performance of models may still be relevant for clinical trial cohort enrichment [27]. In this setting small improvements may substantially reduce the numbers who have to be randomized for the study design to have adequate statistical power.

5. Summary table

Question	
What was already known on the topic?	<ul style="list-style-type: none"> - Deep Learning of fundus images can predict risk factors of CVD in a general cohort from UK Biobank - Increment in predictive performance above that of predictors using clinical risk factors is small in general populations
What this study added to our knowledge?	<ul style="list-style-type: none"> - We find strong evidence that fundus images contain information useful for CVD risk prediction in people with diabetes - Addition of a DL score to a model with clinical risk factors marginally improves CVD prediction for people with diabetes - The risk ratio of a DL score required to produce meaningful improvements in CVD risk prediction is large

Author statement

JM conceived and designed the study. HC, PM, AS, and CS made important contributions to study design. SM and LB were involved in the cleaning up, harmonization, quality-control and databasing of data in Scotland. JM and WJ performed the analyses. JM developed data analysis methods. JM, WJ, and AF contributed to code preparation. JM and WJ drafted the initial manuscript. All authors made critically important contributions to manuscript revision. All authors approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

SDRN-EPI are not data custodians and are not permitted to directly provision data externally. However, the component datasets can be obtained by data governance trained bone fide researchers through the Public Benefit and Privacy Panel for Health and Social Care. See <https://www.informationgovernance.scot.nhs.uk/pbpphsc/> for how to apply.

Acknowledgements

We thank the Scottish Diabetes Research Network for the role in data generation. This work was supported by JDRF [grant 2-SRA-2019-857-S-B].

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijmedinf.2023.105072>.

References

- [1] S.J. Livingstone, D. Levin, H.C. Looker, et al., Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008–2010, *J. Am. Med. Assoc.* 313 (2015) 37–44, <https://doi.org/10.1001/jama.2014.16425>.
- [2] E. Sandoval-Garcia, S. McLachlan, A.H. Price, et al., Retinal arteriolar tortuosity and fractal dimension are associated with long-term cardiovascular outcomes in people with type 2 diabetes, *Diabetologia* 64 (2021) 2215–2227.
- [3] K.H. Frederiksen, L. Stokholm, P.H. Frederiksen, et al., Cardiovascular morbidity and all-cause mortality in patients with retinal vein occlusion: a Danish nationwide cohort study, *Br. J. Ophthalmol.* (2022), <https://doi.org/10.1136/bjophthalmol-2022-321225>.
- [4] H. Ho, C.Y. Cheung, C. Sabanayagam, et al., Retinopathy signs improved prediction and reclassification of cardiovascular disease risk in diabetes: a prospective cohort study, *Sci. Rep.* 7 (2017) 1–8.
- [5] V.Y. Guo, J.C.N. Chan, H. Chung, et al., Retinal information is independently associated with cardiovascular disease in patients with type 2 diabetes, *Sci. Rep.* 6 (2016) 1–8.
- [6] R. Poplin, A.V. Varadarajan, K. Blumer, et al., Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning, *Nat. Biomed. Eng.* 2 (2018) 158–164.
- [7] J. Chang, A. Ko, S.M. Park, et al., Association of cardiovascular mortality and deep learning-funduscopy atherosclerosis score derived from retinal fundus images, *Am. J. Ophthalmol.* 217 (2020) 121–130.
- [8] T.H. Rim, C.J. Lee, Y.-C. Tham, et al., Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs, *The Lancet Digital Health* 3 (2021) e306–e316.
- [9] J. Son, J.Y. Shin, E.J. Chun, et al., Predicting high coronary artery calcium score from retinal fundus images with deep learning algorithms, *Transl. Vis. Sci. Technol.* 9 (2020) 28.
- [10] R.G. Barriada, O. Simó-Servat, A. Planas, et al., Deep learning of retinal imaging: a useful tool for coronary artery calcium score prediction in diabetic patients, *Appl. Sci.* 12 (2022) 1401.

- [11] C.Y. Cheung, D. Xu, C.-Y. Cheng, et al., A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre, *Nat. Biomed. Eng.* 5 (2021) 498–508.
- [12] A.R. Rudnicka, R. Welikala, S. Barman, et al., Artificial intelligence-enabled retinal vasculometry for prediction of circulatory mortality, myocardial infarction and stroke, *Br. J. Ophthalmol.* (2022), <https://doi.org/10.1136/bjo-2022-321842>.
- [13] S.J. McGurnaghan, L.A.K. Blackbourn, T.M. Caparrotta, et al., Cohort profile: the Scottish diabetes research network national diabetes cohort a population-based cohort of people with diabetes in Scotland, *BMJ Open* (2022) 12, <https://doi.org/10.1136/bmjopen-2022-063046>.
- [14] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, et al., Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2, *BMJ* 336 (2008) 1475–1482, <https://doi.org/10.1136/bmj.39609.449676.25>.
- [15] A. Rawshani, A. Rawshani, S. Franzén, et al., Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes. n.d.
- [16] S.J. McGurnaghan, P.M. McKeigue, S.H. Read, et al., Development and validation of a cardiovascular risk prediction model in type 1 diabetes, *Diabetologia* 64 (2021) 2001–2011.
- [17] M. Ilse, J. Tomczak, M. Welling, Attention-based deep multiple instance learning, in: J. Dy, A. Krause (Eds.), *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80, PMLR, 2018, pp. 2127–2136.
- [18] J. Honaker, G. King, M. Blackwell, *Amelia II: a program for missing data*, *J. Stat. Softw.* 45 (2011) 1–47.
- [19] R Core Team, *R: A language and environment for statistical computing*, 2015.
- [20] P. McKeigue, Quantifying performance of a diagnostic test as the expected information for discrimination: relation to the C-statistic, *Stat. Methods Med. Res.* 28 (2019) 1841–1851, doi: 10.1177/0962280218776989.
- [21] M. Stone, An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 39 (1977) 44–47, <https://doi.org/10.1111/j.2517-6161.1977.tb01603.x>.
- [22] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies, *Int. J. Med. Inf.* 153 (2021), 104510, <https://doi.org/10.1016/j.ijmedinf.2021.104510>.
- [23] C. Wang, J. Shao, X. Xu, et al., DeepLN: a multi-task AI tool to predict the imaging characteristics, malignancy and pathological subtypes in CT-detected pulmonary nodules, *Front. Oncol.* (2022) 12, <https://doi.org/10.3389/fonc.2022.683792>.
- [24] R. Caruana, *Multitask learning*, *Mach. Learn.* 28 (1997) 41–75.
- [25] C. Younan, P. Mitchell, R. Cumming, et al., Cardiovascular disease, vascular risk factors and the incidence of cataract and cataract surgery: the Blue Mountains eye study, *Ophthalmic Epidemiol.* 10 (2003) 227–240.
- [26] D.Y.L. Wong, M.C. Lam, A. Ran, et al., Artificial intelligence in retinal imaging for cardiovascular disease prediction: current trends and future directions, *Curr. Opin. Ophthalmol.* 33 (2022) 440–446.
- [27] N. Simon, R. Simon, Adaptive enrichment designs for clinical trials, *Biostatistics* 14 (2013) 613–625, <https://doi.org/10.1093/biostatistics/kxt010>.