ARTICLES
https://doi.org/10.1038/s41551-020-00626-4

nature biomedical engineering

A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre

Carol Y. Cheung[1,2,15], Dejiang Xu[3,15], Ching-Yu Cheng[1,4], Charumathi Sabanayagam[1,4], Yih-Chung Tham[1,4], Marco Yu[1], Tyler Hyungtaek Rim[1,4], Chew Yian Chai[5,6], Bamini Gopinath[7], Paul Mitchell[7], Richie Poulton[8], Terrie E. Moffitt[9], Avshalom Caspi[9], Jason C. Yam[2], Clement C. Tham[2], Jost B. Jonas[10], Ya Xing Wang[11], Su Jeong Song[12], Louise M. Burrell[13], Omar Farouque[13], Ling Jun Li[14], Gavin Tan[1,4], Daniel S. W. Ting[1,4], Wynne Hsu[3,16], Mong Li Lee[3,16] and Tien Y. Wong[1,4,16] ✉

Retinal blood vessels provide information on the risk of cardiovascular disease (CVD). Here, we report the development and validation of deep-learning models for the automated measurement of retinal-vessel calibre in retinal photographs, using diverse multiethnic multicountry datasets that comprise more than 70,000 images. Retinal-vessel calibre measured by the models and by expert human graders showed high agreement, with overall intraclass correlation coefficients of between 0.82 and 0.95. The models performed comparably to or better than expert graders in associations between measurements of retinal-vessel calibre and CVD risk factors, including blood pressure, body-mass index, total cholesterol and glycated-haemoglobin levels. In retrospectively measured prospective datasets from a population-based study, baseline measurements performed by the deep-learning system were associated with incident CVD. Our findings motivate the development of clinically applicable explainable end-to-end deep-learning systems for the prediction of CVD on the basis of the features of retinal vessels in retinal photographs.

For more than a century, physicians have performed a fundus examination in patients with hypertension to determine the presence and severity of retinal vascular damage as a means to estimate cardiovascular disease (CVD) risk[1–4]. This examination is based on the premise that changes observed in the retinal vasculature, such as the degree of retinal arteriolar narrowing, reflect vascular disease in peripheral, cerebral and coronary blood vessels[5–10] and may therefore be a marker of CVD risk.

The development of digital retinal photography and computer software have enabled more-objective estimation of the degree of retinal arteriolar narrowing by measuring the calibre of the retinal vessels[11–14]. Most presently used software is semi-automated, requiring human intervention to adequately measure retinal-vessel calibre on the basis of prespecified protocols[11–14]. Using these approaches, substantial prospective epidemiological data have demonstrated that changes in retinal-vessel calibre (such as narrower retinal arterioles and wider/dilated venules) are associated with classic CVD risk factors (such as blood pressure and diabetes)[9,15–19] and the presence of subclinical vascular diseases (such as carotid atherosclerosis and aortic stiffness)[20,21] and predictive of subsequent clinical CVD events (such as coronary heart disease, stroke and mortality)[22–35]. However, further validation of these highly promising findings in other clinical studies has been limited by the current versions of the semi-automated software, which require human input.

Artificial intelligence, particularly deep learning with convolutional neural networks (CNNs), is being evaluated in medical imaging fields such as radiology[36–38]. Deep-learning CNNs developed for retinal photographs have shown superior performance in the detection of diabetic retinopathy and other retinal diseases compared with human assessment[39–43]. Deep-learning CNNs developed by Poplin et al. have recently demonstrated that retinal photographs themselves are associated with a range of CVD risk factors, including blood pressure, body mass index (BMI), smoking and glycated-haemoglobin level[44]. However, their study did not

[1]Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore. [2]Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, China. [3]School of Computing, National University of Singapore, Singapore, Singapore. [4]Ophthalmology and Visual Sciences Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore. [5]Emergency Medicine Department, National University Hospital, Singapore, Singapore. [6]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [7]Centre for Vision Research, Westmead Institute for Medical Research, Westmead, New South Wales, Australia. [8]Dunedin Multidisciplinary Health and Development Research Unit, Department of Psychology, University of Otago, Dunedin, New Zealand. [9]Department of Psychology and Neuroscience, Duke University, Durham, NC, USA. [10]Department of Ophthalmology, Medical Faculty Mannheim, Ruprecht-Karls-University, Heidelberg, Germany. [11]Beijing Ophthalmology & Visual Science Key Lab, Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [12]Department of Ophthalmology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [13]Department of Cardiology, Austin Health, Austin Hospital, and Department of Medicine, University of Melbourne, Heidelberg, Victoria, Australia. [14]Division of Obstetrics and Gynaecology, KK Women's and Children's Hospital, Singapore, Singapore. [15]These authors contributed equally: Carol Y. Cheung, Dejiang Xu. [16]These authors jointly supervised this work: Wynne Hsu, Mong Li Lee, Tien Y. Wong. ✉e-mail: wong.tien.yin@snec.com.sg

Table 1 | Description and characteristics of the training, validation and external testing datasets

Study and outcome	Number of individuals	Number of eyes	Age (years)	Gender (male)	Ethnicity	MABP (mm Hg)	BMI (kg m^{-2})	Total cholesterol level (mmol l^{-1})	HbA_{1c} (%)	Currently smoking	Retinal camera
Training and validation for SIVA-DLS											
[a]	2,267	2,267	58.4 (9.19)	1,101 (48.6%)	Chinese (100%)	96.9 (10.9)	23.7 (3.69)	5.47 (1.06)	5.96 (0.72)	318 (14.0%)	Canon CR-DGi 10D
[b]	2,528	2,528	57.6 (10.2)	1,287 (50.9%)	Indian (100%)	97.3 (12.2)	26.6 (5.03)	5.10 (1.10)	6.59 (1.49)	472 (18.7%)	Canon CR-DGi 10D
[c]	514	514	59.9 (8.59)	212 (41.2%)	Malay (100%)	104.2 (13.2)	28.2 (4.85)	5.40 (1.11)	8.21 (1.89)	65 (12.5%)	Canon CR-DGi 10D
Total:	5,309	5,309									
External Testing for SIVA-DLS											
Agreement between SIVA-human and SIVA-DLS											
[d]	875	875	44.6 (12.2)	421 (48.1%)	Chinese (64.34%); Indian (23.43%); Malay (12.23%)	91.5 (11.1)	22.6 (3.16)	5.10 (0.90)	5.69 (0.59)	0 (0%)	Canon CR-DGi 10D
[e]	792	792	38.3 (0.43)	405 (51.1%)	Primarily Caucasian	91.9 (9.73)	27.1 (5.21)	5.23 (1.03)	5.41 (0.56)	168 (21.2%)	Canon NMR-45 20D
[f]	672	672	7.66 (1.04)	337 (50.1%)	Chinese (100%)	78.52 (9.53)	16.15 (2.85)	NA	NA	NA	Topcon TRC 50DX
[g]	1,103	1,103	61.2 (11.6)	842 (76.3%)	Caucasian (70.2%); east Asian (5.3%); south Asian (7.6%); mid Asian (9.6%); others (7.3%)	91.4 (13.1)	29.5 (5.54)	4.57 (1.13)	7.11 (1.86)	324 (28.9)	Canon CR-DGI
[h]	614	614	48.5 (10.7)	451 (73.5%)	Chinese (56.35%); Malay (17.75%); Indian (14.50%); others (11.40%)	100.7 (12.3)	26.1 (4.50)	NA	NA	158 (25.7%)	Canon CRII
[i]	626	626	61.9 (10.7)	388 (62.0%)	Chinese (67.9%); Malay (16.5%); Indian (7.0%); others (8.6%)	97.2 (13.0)	27.2 (5.26)	4.64 (1.20)	6.74 (1.53)	78 (12.5%)	Canon CRII
[j]	269	269	64.3 (11.4)	133 (49.4%)	Chinese (100%)	99.9 (11.7)	25.9 (4.91)	4.27 (0.94)	7.34 (1.41)	21 (7.8%)	Topcon TRC 50DX
[k]	272	272	33.0 (4.90)	All female	Chinese (55.1%); Malay (21.7%); Indian (23.2%)	83.7 (10.6)	25.6 (5.49)	5.24 (1.06)	NA	4 (1.5%)	Canon CR-1, 40D SLR
[l]	238	238	60.1 (11.3)	121 (50.8%)	Chinese (75.2%); Malay (5.5%); Indian (19.3%)	91.0 (9.36)	27.3 (5.01)	4.50 (0.86)	7.54 (1.46)	NA	Topcon NW8
[m]	175	175	53.6 (6.41)	58 (33.1%)	Chinese (90.86%); Malay (6.29%); Indian (2.85%)	92.8 (12.01)	24.0 (4.14)	NA	NA	NA	Canon CR-DGi 10D
Total:	5,636	5,636									

Continued

Table 1 | Description and characteristics of the training, validation and external testing datasets (continued)

Study and outcome	Number of individuals	Number of eyes	Age (years)	Gender (male)	Ethnicity	MABP (mm Hg)	BMI (kg m^{-2})	Total cholesterol level (mmol l^{-1})	HbA_{1c} (%)	Currently smoking	Retinal camera
Association of SIVA-DLS with CVD risk factors, CVD outcomes and mortality											
[n,s]	8,980	8,980	58.1 (10.1)	4,262 (47.5%)	Chinese (34.6); Malay (32.3%); Indian (33.1%)	98.8 (12.8)	25.4 (4.69)	5.49 (1.10)	6.28 (1.31)	1432 (16.0%)	Canon CR-DGi 10D
[o,t]	3,697	3,697	55.6 (10.2)	1,616 (43.7%)	Chinese (100%)	NA	NA	NA	NA	NA	Canon CR6-45NM
[p,u]	45,644	45,644	55.7 (8.13)	20,315 (44.5%)	British (86.6%); other white background (4.2%); Irish (3.6%); Indian (1%); others (4.6%)	100.5 (11.4)	27.1 (4.67)	NA	NA	3,897 (8.56%)	Topcon 3D OCT-1000 Mark II
[q,u]	627	627	41.1 (4.25)	565 (90.1%)	Korean (100%)	89.76 (9.78)	24.30 (3.04)	5.29 (0.90)	5.68 (0.57)	158 (25.2%)	Canon CR6-45NM
[r,u]	243	243	58.2 (10.8)	166 (68.3%)	Primarily Caucasian	96.32 (11.9)	31.71 (6.60)	4.48 (1.27)	NA	66 (27.2%)	Canon EOS 40D
Total:	59,191	59,191									

Data are mean (s.d.) or number of individuals (%). HbA1c, glycated-haemoglobin level. NA, not available. [a]Singapore Chinese eye study (SCES), Chinese adults were selected using an age-stratified random sampling method. [b]Singapore Indian eye study (SINDI), Indian adults were selected using an age-stratified random sampling method. [c]Singapore Malay eye study (SiMES), Malay adults were selected using an age-stratified random sampling method. Only patients with diabetes were included. [d]Singapore prospective study program (SP2), a subgroup of control healthy individuals were randomly selected from the SP2 in Singapore. [e]The Dunedin multidisciplinary health and development study (Dunedin study), individuals aged 38 years from a longitudinal investigation of health and behaviour in a complete birth cohort born between April 1972 and March 1973 in Dunedin, New Zealand. [f]Hong Kong children eye study (HKCES), children of grade 1 to grade 3 (aged about 6–8 years) from primary schools in Hong Kong. [g]Australian heart eye study (AHES), patients who had symptoms of suspected coronary artery disease presented to a tertiary referral hospital for evaluation of potential coronary artery disease by coronary angiography in Australia. [h]Retinal imaging in chest pain study (RICP study), low-risk patients with acute coronary syndrome who had chest pain were recruited from the emergency department in a general hospital in Singapore. [i] Retinal imaging in renal disease study (IRED study), patients diagnosed with chronic kidney disease were recruited from nephrology clinics in a general hospital in Singapore. [j]CUHK sight-threatening diabetic retinopathy study (CUHK-STDR study), patients with diabetes were recruited from a tertiary eye hospital in Hong Kong. [k]Growing up in Singapore towards healthy outcomes birth cohort (GUSTO study), pregnant women who intended to give birth and reside in Singapore for the next 5 years. [l]Singapore integrated diabetic retinopathy program (SiDRP), patients with diabetes were recruited from a diabetic retinopathy screening program in Singapore. [m]Cardiovascular disease screening using retinal vascular imaging study (CVD screening study), participants without a history of stroke and heart disease were recruited from a community-based clinic in Singapore. [n]The Singapore Epidemiology of Eye Diseases Study (SEED study), adults residing in Singapore aged 40–80 years at baseline were selected using an age-stratified random sampling method. [o]The Beijing Eye Study (BES), Chinese adults aged ≥40 years who were initially examined in 2001 in northern China. [p]The UK Biobank. [q]The Kangbuk Samsung health study (KSH study), participants were recruited from a medical health checkup program. [r]The Austin health study, participants with risk factors of atherosclerosis and coronary artery disease from cardiology outpatient clinics in Australia. [s]Outcome is incidence of a CVD event. [t]Outcome is ten-year all-cause mortality. [u]Outcome is CVD risk factors.

evaluate whether specific retinal-vessel features (such as retinal arteriolar calibre narrowing or venular calibre widening) were correlated with CVD risk.

To address these gaps, we developed and tested a deep-learning system (DLS) to specifically measure retinal-vessel calibre from retinal photographs. We assessed the agreement of the retinal-vessel calibre measurement between the DLS and humans, and compared their associations with classic CVD risk factors (such as blood pressure). Finally, we examined the relationship between retinal-vessel calibre measured by DLS and incident CVD using a prospective dataset. Our DLS may provide an additional model for the application of retinal photography for CVD assessment.

Results

Table 1 describes the study cohorts, demographic data and retinal camera information. The mean absolute error of retinal-vessel calibre measurement by human graders using the Singapore I vessel assessment (SIVA) software (SIVA-human) and retinal-vessel calibre measurement using a DLS (SIVA-DLS) is shown (Supplementary Table 1). Figure 1 shows an example of retinal-vessel calibre prediction by SIVA-DLS.

Table 2 shows the intraclass correlation coefficients (ICCs) of retinal-vessel calibre between SIVA-DLS and SIVA-human. Agreement of calibre measurement in the validation and external testing datasets were good to excellent, with ICCs ranging from 0.82 to 0.95; agreement in individual external datasets was also good (all ICCs were above 0.7). Bland–Altman plots show the agreement of retinal-vessel calibre between SIVA-DLS and SIVA-human in the validation dataset and total external dataset (Supplementary Figs. 3 and 4, respectively). The Bland–Altman plots suggested there might be proportional biases on central retinal artery equivalent (CRAE), with poorer agreement between SIVA-DLS and SIVA-human at large central retinal vein equivalent (CRVE) values.

Multivariable linear regression analysis of CVD risk factors and retinal-vessel calibre measured by SIVA-DLS and SIVA-human are shown in Table 3. The associations of CRAE with age, gender, mean arterial blood pressure (MABP), body-mass index (BMI) and total cholesterol, and associations of CRVE with gender, MABP, BMI, glycated-haemoglobin and current smoking were largely identical (that is, similar Nagelkerke's pseudo R^2 (R_N^2) values) between SIVA-DLS and SIVA-human. For example, each s.d. increase in CRAE at zone B ($CRAE_B$) was associated with lower blood pressure (β, −3.67 mm Hg versus −3.28 mm Hg); and each s.d. increase in CRVE at zone B ($CRVE_B$) was associated with higher glycated-haemoglobin level (β, 0.312% versus 0.295%), comparing SIVA-DLS versus SIVA-human, respectively. The regression model fit (that is, R_N^2) for SIVA-DLS was generally higher compared with the R_N^2 of SIVA-human, indicating that SIVA-DLS measurement

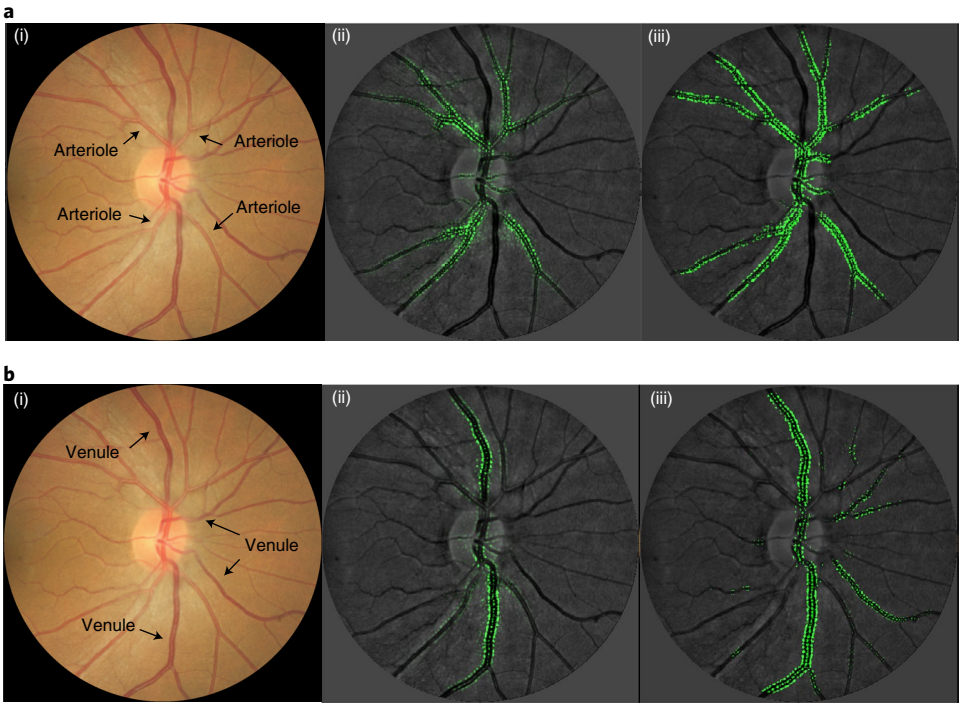

Fig. 1 | An example of CRAE and CRVE prediction using a DLS (SIVA-DLS). a, CRAE. **b**, CRVE. For **a** and **b**, retinal photographs were captured by a 45° digital retinal camera (i) (Canon CR-DGi 10D; Canon). The colour of retinal venules is generally darker and the width of retinal venules is generally larger than that of retinal arterioles. SIVA calculates the calibres at two regions: one from 0.5 to 1.0 disc diameters away from the disc margin (zone B; $CRAE_B$ and $CRVE_B$) and one from 0.5 to 2.0 disc diameters away from the disc margin (zone C; $CRAE_C$ and $CRVE_C$). Integrated gradients at selected high-level (ii) and low-level (iii) layers of the neural network were used to generate a heat map to show the most important and basic features, respectively, by the predicted attributions. The green dots indicate positive attributions. For the actual prediction, the fully connected layer learned proper weights accordingly for different features to obtain retinal-vessel calibre values at different zones. It gives high weight to features within zone B during calculating $CRAE_B$ and $CRVE_B$, and high weight to features within zone C for $CRAE_C$ and $CRVE_C$.

explained more of the variations in the models. The SIVA-DLS models predicted CVD risk factors significantly better than the SIVA-human models ($P<0.05$, bootstrap t-test) or at least comparable with the SIVA-human models ($P>0.05$, bootstrap t-test), except for CRAE at zone C ($CRAE_C$) and CRVE at zone C ($CRVE_C$) with MABP. The multivariable linear regression analysis of CVD risk factors and retinal-vessel calibre in individual external datasets is provided (Supplementary Table 2).

In the Singapore epidemiology of eye diseases (SEED) study, over the follow-up period (median, 4.54 years; range, 0.02–11.8 years), 851 participants developed incident CVD events (174 stroke, 340 myocardial infarction, 181 CVD death and 156 mixed). Table 4 shows that narrower $CRAE_B$ (hazard ratio (HR) per s.d. decrease (95% confidence interval (CI)), 1.12 (1.02–1.24)) and narrower $CRAE_C$ (HR per s.d. decrease, 1.13 (1.02–1.26)) were independently associated with incident CVD events in the SEED study, after adjusting for age, gender, ethnicity, fellow calibre, BMI, MABP, glycated-haemoglobin level, total cholesterol level and smoking at baseline. In the Beijing eye study (BES), narrower $CRAE_B$ (odds ratio (OR), 1.88 (1.34–2.31), first quartile versus fourth quartile; OR per s.d. decrease, 1.27 (1.13–1.41)) and $CRAE_C$ (OR, 1.67 (1.19–2.34), first quartile versus fourth quartile; OR per s.d. decrease, 1.19 (1.06–1.34)) were associated with ten-year all-cause mortality in the crude model (Supplementary Table 3). In the UK Biobank, the Kangbuk Samsung Health (KSH) study and the Austin health study, retinal-vessel calibres measured by SIVA-DLS were associated with several CVD risk factors (Supplementary Table 4). For example, narrower CRAE was associated with higher MABP and greater BMI, whereas wider CRVE was associated with smoking.

Figure 2 shows examples of a comparison between SIVA-DLS and SIVA-human in cases with different CVD risk factors. The heatmaps of SIVA-DLS generally highlighted the boundaries of the arterioles and venules to predict CRAE and CRVE, respectively, in retinal photographs without other features or any retinal pathologies. Furthermore, the values of CRAE and CRVE can enable users to quantitatively assess retinal-vessel narrowing or widening, which has been shown to correlate with CVD risk factors. However, other features or retinal pathologies—such as nerve fibre layer reflection, the presence of laser scars and retinal haemorrhage on the fundus—may affect the prediction of calibre by the SIVA-DLS (representative examples are provided in Supplementary Fig. 5) and may explain the relatively lower agreement in some of the external testing cohorts in Table 2.

Finally, we developed a prototype interface for SIVA-DLS with fully automated measurement of retinal-vessel calibre as an output (Supplementary Fig. 6), and we tested its feasibility prospectively under a telemedicine-based diabetic retinopathy screening program. The high agreement of retinal-vessel calibre between SIVA-DLS and SIVA-human was further demonstrated, with all ICCs above 0.8 (Supplementary Table 4).

Discussion

Deep learning has the potential to transform clinical care in medical imaging fields such as radiology and ophthalmology[36–38,45]. Here we developed and validated a deep-learning CNN (SIVA-DLS) that specifically measured retinal-vessel calibre from retinal photographs. We report a high correlation between the SIVA-DLS and validated human measurements. We demonstrated that the

Table 2 | Agreement estimates of retinal-vessel calibre for SIVA-DLS and SIVA-human

	$CRAE_B$	$CRAE_C$	$CRVE_B$	$CRVE_C$
Validation	ICC (95% CI)	ICC (95% CI)	ICC (95% CI)	ICC (95% CI)
SEED study (n = 1,060)	0.88 (0.866–0.893)	0.917 (0.907–0.926)	0.94 (0.933–0.947)	0.948 (0.941–0.954)
External testing	ICC (95% CI)	ICC (95% CI)	ICC (95% CI)	ICC (95% CI)
Total dataset (n = 5,636)	0.823 (0.814–0.831)	0.864 (0.858–0.871)	0.881 (0.875–0.886)	0.917 (0.913–0.921)
SP2 (n = 875)	0.846 (0.826–0.863)	0.874 (0.857–0.889)	0.87 (0.853–0.885)	0.903 (0.890–0.914)
Dunedin study (n = 792)	0.693 (0.655–0.728)	0.781 (0.753–0.807)	0.876 (0.859–0.892)	0.877 (0.860–0.892)
HKCES (n = 672)	0.689 (0.647–0.727)	0.748 (0.712–0.779)	0.82 (0.794–0.843)	0.808 (0.780–0.833)
AHES (n = 1,103)	0.694 (0.662–0.724)	0.781 (0.757–0.803)	0.785 (0.761–0.806)	0.870 (0.855–0.884)
RICP study (n = 614)	0.812 (0.783–0.837)	0.864 (0.842–0.882)	0.806 (0.777–0.832)	0.901 (0.885–0.915)
IRED study (n = 626)	0.812 (0.784–0.837)	0.857 (0.835–0.877)	0.898 (0.881–0.912)	0.9 (0.884–0.914)
CUHK-STDR study (n = 269)	0.692 (0.624–0.750)	0.708 (0.643–0.763)	0.841 (0.802–0.873)	0.806 (0.760–0.844)
GUSTO study (n = 272)	0.768 (0.714–0.813)	0.819 (0.775–0.854)	0.79 (0.741–0.831)	0.852 (0.816–0.881)
SiDRP (n = 238)	0.822 (0.776–0.860)	0.857 (0.820–0.888)	0.906 (0.880–0.926)	0.92 (0.897–0.937)
CVD screening study (n = 175)	0.791 (0.728–0.841)	0.849 (0.802–0.886)	0.907 (0.877–0.930)	0.901 (0.869–0.926)

Table 3 | Multivariable linear regression analysis of cardiovascular risk factors (dependent variables) and retinal vascular calibre (independent variables) measured by human graders and a DLS, adjusting for age, gender and fellow calibre

	Age		Gender		MABP		BMI		HbA_{1c}		Smoking		Total cholesterol	
	β (s.e.)	R_N^2	OR (s.e.[b])	R_N^2	β (s.e.)	R_N^2	β (s.e.)	R_N^2	β (s.e.)	R_N^2	OR (s.e.[b])	R_N^2	β (s.e.)	R_N^2
SIVA-human $CRAE_B$	7.01 (0.331)**	0.175	1.47 (0.040)**	0.043	−3.28 (0.228)**	0.198	−1.05 (0.105)**	0.242	−0.109 (0.035)*	0.199	0.868 (0.052)*	0.037	−0.104 (0.025)**	0.070
SIVA-DLS $CRAE_B$	7.95 (0.363)**	0.212	1.83 (0.046)**	0.063	−3.67 (0.257)**	0.199	−1.26 (0.118)**	0.244	−0.089 (0.040)*	0.202	0.708 (0.059)**	0.045	−0.089 (0.029)*	0.068
P[a]		<0.001		<0.001		0.896		0.368		0.352		0.014		0.203
SIVA-human $CRAE_C$	−7.00 (0.378)**	0.152	1.66 (0.045)**	0.053	−4.12 (0.253)**	0.210	−1.22 (0.117)**	0.243	−0.160 (0.038)**	0.206	0.938 (0.058)	0.035	−0.109 (0.028)**	0.070
SIVA-DLS $CRAE_C$	−8.85 (0.385)**	0.195	2.25 (0.050)**	0.089	−3.77 (0.270)**	0.201	−1.60 (0.123)**	0.251	−0.195 (0.040)**	0.215	0.700 (0.062)**	0.045	−0.091 (0.030)*	0.069
P[a]		<0.001		<0.001		0.011		0.004		0.051		0.005		0.306
SIVA-human $CRVE_B$	−1.14 (0.328)**	0.175	0.758 (0.038)**	0.043	0.302 (0.218)	0.198	0.421 (0.100)**	0.242	0.295 (0.033)**	0.199	1.03 (0.049)	0.037	0.033 (0.023)	0.070
SIVA-DLS $CRVE_B$	−1.01 (0.359)*	0.212	0.636 (0.043)**	0.063	0.696 (0.244)*	0.199	0.586 (0.112)**	0.244	0.312 (0.037)**	0.202	1.22 (0.055)**	0.045	0.033 (0.027)	0.068
P[a]		<0.001		<0.001		0.896		0.368		0.352		0.014		0.203
SIVA-human $CRVE_C$	−0.298 (0.376)	0.151 (0.151)	0.628 (0.044)**	0.053	1.03 (0.244)**	0.210	0.660 (0.113)**	0.243	0.361 (0.036)**	0.206	1.00 (0.056)	0.035	0.033 (0.026)	0.070
SIVA-DLS $CRVE_C$	0.622 (0.381)	0.195	0.501 (0.047)**	0.089	0.728 (0.255)*	0.201	0.865 (0.117)**	0.251	0.426 (0.039)**	0.215	1.23 (0.058)**	0.045	0.022 (0.028)	0.069
P[a]		<0.001		<0.001		0.011		0.004		0.051		0.005		0.306

Retinal-vessel calibre was analysed as per s.d. increase. Results were adjusted for age (except for models for age), gender (except for models for gender) and fellow calibre (that is, CRVE was included as an independent variable in the model for CRAE, and vice versa). Statistical significance is indicated by asterisks; *P < 0.05, **P < 0.001. β, regression coefficient. [a]Statistical analysis was performed using a bootstrap t-test. [b]Denotes exp[β] and s.e. of β under logistic regression.

relationships between retinal-vessel calibre and classic CVD risk factors measured using SIVA-DLS were better or comparable to humans. Finally, we demonstrated that retinal-vessel calibre measured using SIVA-DLS was associated with incident CVD events.

The rationale of our study is based on a widely accepted concept that retinal vascular health mirrors other vascular beds (for example, cerebral and coronary)[46]. We and other investigators have shown in previous epidemiological and clinical studies that changes in retinal-vessel calibres, measured using semi-automated software, may be markers of CVD risk[6,8,9,15,16,20–33]. However, most widely used software for quantifying retinal-vessel calibre (for example, SIVA, retinal analysis and integrative vessel analysis) requires substantial human assessment and manipulation, usually by trained technicians (graders) following standardized protocols[11–14]. For example, in the current SIVA-human version, several grading steps often require manual correction, such as centreing the grading grid on the optic disc, identifying the vessel type and adjusting to vessel tracing, such that the average grading time per retinal photograph is about 25 min depending on image quality (~10% and ~20–30% manual correction for good and poor quality images, respectively)[13,47]. Furthermore, inter- and intrahuman grader variability is difficult to completely eliminate even when using standardized

Table 4 | Relation of retinal-vessel calibre measured using a DLS to risk of a CVD event in the SEED study

			Model 1		Model 2	
	Number at risk (*n*)	Incident CVD event cases (*n* (%))	HR (95% CI)	*P*	HR (95% CI)	*P*
$CRAE_B$						
First quartile	2,221	246 (11.1)	1.27 (1.02–1.59)	0.035	1.23 (0.98–1.55)	0.079
Second quartile	2,244	189 (8.4)	0.94 (0.76–1.15)	0.534	0.93 (0.75–1.15)	0.491
Third quartile	2,256	202 (9.0)	1.03 (0.85–1.25)	0.792	1.02 (0.83–1.25)	0.854
Fourth quartile	2,259	214 (9.5)	Reference		Reference	
Per s.d. decrease	8,980	851 (9.5)	1.18 (1.08–1.30)	<0.001[a]	1.12 (1.02–1.24)	0.024[a]
$CRAE_C$						
First quartile	2,210	247 (11.2)	1.18 (0.94–1.47)	0.156	1.12 (0.89–1.42)	0.337
Second quartile	2,258	185 (8.2)	0.89 (0.73–1.10)	0.280	0.91 (0.74–1.13)	0.383
Third quartile	2,270	197 (8.7)	0.92 (0.75–1.11)	0.380	0.90 (0.74–1.10)	0.311
Fourth quartile	2,242	222 (9.9)	Reference		Reference	
Per s.d. decrease	8,980	851 (9.5)	1.24 (1.13–1.37)	<0.001[a]	1.13 (1.02–1.26)	0.017[a]
$CRVE_B$						
First quartile	2,190	212 (9.7)	Reference		Reference	
Second quartile	2,258	186 (8.2)	1.19 (0.97–1.47)	0.101	1.14 (0.91–1.41)	0.253
Third quartile	2,266	193 (8.5)	1.31 (1.05–1.64)	0.018	1.17 (0.92–1.47)	0.198
Fourth quartile	2,266	260 (11.5)	1.74 (1.37–2.21)	<0.001	1.30 (1.01–1.67)	0.039
Per s.d. increase	8,980	851 (9.5)	1.14 (1.04–1.25)	0.006[a]	1.08 (0.98–1.18)	0.130[a]
$CRVE_C$						
First quartile	2,209	206 (9.3)	Reference		Reference	
Second quartile	2,252	166 (7.4)	1.12 (0.90–1.40)	0.305	1.07 (0.85–1.34)	0.578
Third quartile	2,262	214 (9.5)	1.37 (1.09–1.71)	0.006	1.24 (0.98–1.57)	0.068
Fourth quartile	2,257	265 (11.7)	1.66 (1.31–2.12)	<0.001	1.23 (0.96–1.59)	0.108
Per s.d. increase	8,980	851 (9.5)	1.21 (1.09–1.33)	<0.001[a]	1.09 (0.99–1.21)	0.091[a]

Mode 1 was adjusted for age, gender, ethnicity and fellow calibre (that is, CRVE was included as an independent variable in the model for CRAE, and vice versa) at baseline. Model 2 was adjusted for age, gender, ethnicity, fellow calibre, BMI, MABP, glycated-haemoglobin level, total cholesterol level and smoking at baseline. Unless otherwise indicated, *P* values were determined on the basis of calibre measures entered as a categorical variable (quartiles) and compared with referent quartile. To account for multiple comparisons in the analysis for the calibre treated as a categorical variable and a continuous variable, *P* values of less than 0.008 (that is, 0.05/6) and 0.025 (that is, 0.05/2) were considered to be statistical significant after Bonferroni correction, respectively. [a]*P* values were determined on the basis of calibre measures entered as a continuous variable.

protocols and training. By contrast, SIVA-DLS takes a few seconds with no human assessment, significantly increasing ease of use and decreasing cost and variability. Although other CNNs have been reported for automated segmentation of retinal vessels[48–50], our current algorithm produces fully automated measurement of retinal-vessel calibre as an output without the need for vessel segmentation. Thus, the predicted values of retinal-vessel calibre from SIVA-DLS can be immediately used in clinical and epidemiological studies, enabling faster, easier, cheaper and more consistent output. The generated heat maps provide a quick and broad visual assessment of the accuracy of the algorithm in measuring retinal vessels, and reveal any gross errors, including mistaken attributions from SIVA-DLS for artifacts and other pathologies. We note that SIVA-DLS is not the only fully automated software to measure retinal-vessel calibre, and future studies could compare SIVA-DLS with other software (such as QUARTZ and ALTAIR)[51,52].

Some key findings in our study should be discussed in the context of existing literature. First, we showed that the pattern of associations between SIVA-DLS retinal-vessel calibre and classic CVD risk factors was largely similar to human measurements. Retinal arteriolar narrowing (smaller arteriolar calibre) has long been known to be an early vascular response in hypertension that is associated with systemic peripheral vasoconstriction and atherosclerosis[5,53]. By contrast, retinal venular dilatation (wider venular calibre) is a clinical sign of diabetic retinopathy and retinal venous occlusion, and is associated with endothelial dysfunction, inflammation and microvascular hypoxia[15,16,54,55]. In large epidemiological studies, such as the atherosclerosis risk in communities (ARIC) study, the cardiovascular health study and the Rotterdam study, among others, narrower retinal arteriolar calibre measured using human-based semi-automated software showed a strong and consistent association with higher blood pressure[17,19,56–62], whereas wider retinal venules were associated with higher glycated-haemoglobin level and diabetes[18,63–67]. We demonstrated similar associations using SIVA-DLS. In the regression models, the CVD risk-factor associations were generally better (at least comparable) for SIVA-DLS, reflecting a higher precision in measurement compared with SIVA-human.

Second, we demonstrated that the narrower CRAE measured by SIVA-DLS was associated with incident CVD and all-cause mortality in two prospective cohorts. These results are consistent with other large prospective studies that used previous human-based semi-automated software[24,30,31,33]. The results of the mortality association in one of our cohorts (BES) should be interpreted with caution as several key risk factors (for example, blood pressure) were not available and there were few all-cause mortality cases overall.

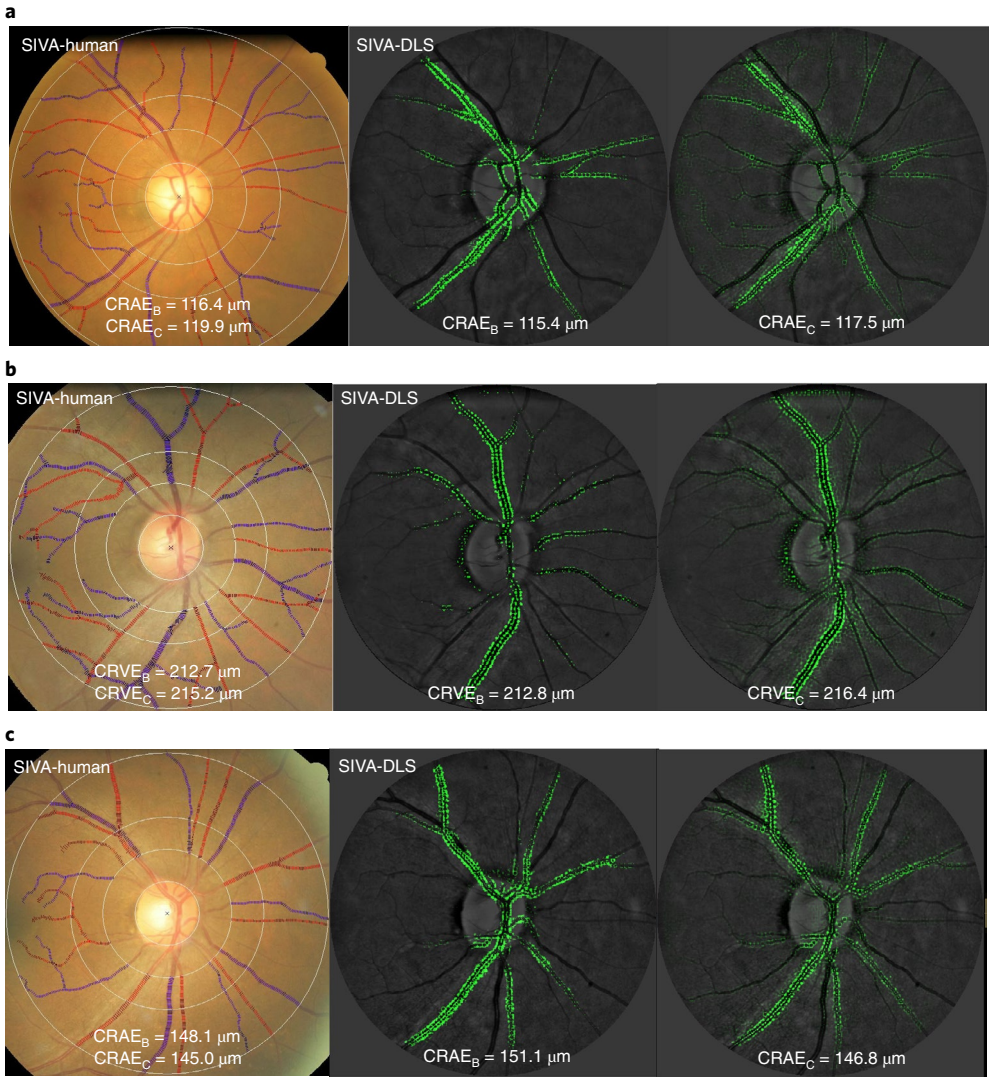

Fig. 2 | Comparison between SIVA-DLS and SIVA-human in cases with different CVD risk factors. a, Hypertension associated with retinal arteriolar narrowing**. b**, Smoking associated with retinal venular widening. **c**, Obesity associated with retinal arteriolar narrowing. In SIVA-human, arterioles and venules are coloured in red and blue, respectively. Integrated gradients at a selected low-level layer of the neural network were used to generate heat maps to show the basic features of SIVA-DLS. Similar to the vessel detection by SIVA-human, the heat maps of SIVA-DLS highlighted the boundaries of the arterioles and venules to predict CRAE and CRVE, respectively. The output values enable users to assess retinal-vessel narrowing or widening quantitatively. Note that, as segments of retinal vessels were not used for training, the highlighted regions in the SIVA-DLS heat maps are mainly used for visualization, and are not directly representative of the vessel edge measured by SIVA-human.

Further testing of SIVA-DLS on prospective cohorts with incident CVD and mortality outcomes will be useful.

Our study used a different approach compared with the study of Poplin et al.[44]. Poplin et al. reported that many CVD risk factors can be predicted from the entire retinal photograph using deep-learning CNNs without specifying which retinal lesion or feature (for example, retinal vessels, optic disc, retinal nerve fibre layer) was responsible for the association. There are concerns regarding the 'black box' phenomenon in such approaches[68–71]. Poplin et al. also did not evaluate the ability of their CNNs to prospectively predict incident CVD from retinal photographs[44]. Our study design was different; we trained our CNNs to specifically measure retinal-vessel calibre (Fig. 1), a feature that most physicians understand is a measure of vascular health and CVD risk[5,72,73]. As discussed above, many large, prospective epidemiological studies have shown that individuals with narrower retinal arterioles are at high risk for CVD[24,28–30,35]. We chose this approach because we believe that using a more-specific vascular marker may enable broader acceptance of retinal photography as an instrument for CVD risk assessment in the clinical community[45,74]. Both approaches (Poplin's CNNs using the entire retinal image and SIVA-DLS specifically focused on retinal-vessel calibre) may have complementary roles in CVD risk assessment.

The strengths of this study include a large and diverse clinical sample, with external testing datasets from 15 multiethnic, multicountry cohorts and >70,000 retinal images in different clinical settings, the availability of prospective datasets to examine incident CVD and mortality, and a DLS without intermediate steps (for example, segmentation of retinal vessel) and with a focused vascular feature (that is, retinal-vessel calibre) for potential clinical application. Furthermore, we demonstrated the robustness of the DLS with a prototype version in a prospective feasibility study. However, our study had limitations. First, we trained and tested our SIVA-DLS on only gradable retinal photographs (proportion of about 93% of all images). Gradability of retinal photographs is typically based on

a protocol that includes quality-control measures so that the largest six retinal vessels (arterioles and venules) are visible for retinal calibre measurement by humans[13]. However, as in other algorithms for retinal photographs, an initial gradability assessment is essential, combining quality checks for image acquisition and inclusion, before calibre measurements. We have started developing a separate CNN for automated filtering of ungradable photographs (which we have tested on the UK Biobank dataset; Methods), which will be incorporated into the next version of SIVA-DLS. Second, in the Bland–Altman plots, there was a suggestion of proportional biases on CRAE, with poorer agreement between SIVA-DLS and SIVA-human for the larger CRAE values; this could imply that there was more variability of SIVA-DLS in younger patients as they tend to have a larger arteriolar calibre (larger CRAE). We are presently working on further testing SIVA-DLS on younger cohorts. Third, as in many deep-learning studies on medical imaging, the labels to train a CNN may not be 'gold standard', and can only be classified as 'current standard'. Here, despite substantial grader training and the use of highly standardized protocols, human measurements were used as ground-truth labels to train the CNN, and the CNN was therefore probably affected by intergrader variability (that is, measurement error from human). One approach to address this is by using the average of multiple human graders for each retinal photograph; however, creating such datasets is costly and is unavailable at present. Fourth, the predictions of retinal-vessel calibre by SIVA-DLS are quantitative values, and inaccurate prediction is not easily noticeable by physicians even if the heat maps of SIVA-DLS highlight the boundaries of the arterioles and venules. Fifth, the R^2 values in the regression models of CVD risk factors explained by retinal-vessel calibre were relatively low, indicating that a large proportion of variability of CVD risk factors cannot be explained by retinal vessels. This is not surprising and applies to many of the current and potential biomarkers of CVD (for example, C-reactive protein[75,76]). Sixth, the presence of subtle retinal pathologies (for example, retinal haemorrhage) or other features (for example, reflection of the nerve fibre layer, laser scars) may introduce additional errors for SIVA-DLS. Finally, although we contacted many collaborative groups globally to test and validate the SIVA-DLS, the current data sharing and medico-legal environment has restricted the ability to share retinal photographs and clinical data.

What are the possible applications of our study? First, an immediate use of SIVA-DLS is its application in clinical and epidemiological studies, replacing the need for the presently used semi-automated systems (SIVA-human). Second, further studies using SIVA-DLS based on the measuring of retinal vessels will enable the generation of larger prospective datasets to test clinical value, potentially enabling clinicians to evaluate patients at risk for CVD. Such studies may include testing in a real-world setting, such as that which was performed in our prospective feasibility study. Third, retinal-vessel calibre measurement has already been incorporated as biomarker measures in clinical trials for retinal (for example, diabetic retinopathy) and systemic (for example, dementia and hypertension) diseases[77–81]. For example, Hughes et al.[80] showed that antihypertensive treatment is associated with improvement in retinal arteriolar narrowing, suggesting that measurement of retinal-vessel calibre can be an additional means to assess the usefulness of new antihypertensive therapies[80].

In conclusion, we developed, validated and externally tested a DLS to measure retinal-vessel calibres from retinal photographs. Retinal-vessel calibre is a specific clinical feature of CVD risk that many physicians may appreciate and accept. We showed that such retinal calibre measurements are correlated with CVD risk factors and are associated with incident CVD events. This will have immediate value for research application in clinical studies and—ultimately, if proven in future studies—for clinical CVD prediction and risk stratification.

Methods

We trained and validated and then externally tested a deep-learning CNN to specifically measure retinal blood vessel (arteriolar and venular) calibres from digital retinal photographs (referred to as SIVA-DLS) using a large and diverse multiethnic multicountry dataset of more than 70,000 retinal photographs from 15 datasets.

This study was approved by the human ethics boards of SingHealth, Singapore; the National Healthcare Group, Singapore; the Western Sydney Local Health Network, Australia; the Lower South Regional Ethics Committee, New Zealand; the Kowloon Central/East Research Ethics Committee, Hong Kong; the Medical Ethics Committee of the Beijing Tongren Hospital, China; the North West Multicentre Research Ethics Committee, United Kingdom; the Institutional Review Board of Kangbuk Samsung Hospital, South Korea; and the Human Research Ethics Committee of Austin Health, Australia. All of the studies were conducted in accordance with the Declaration of Helsinki, and written informed consent was obtained from each participant.

Retinal-vessel calibre. For all of the retinal photographs in this study, trained and experienced human technicians measured retinal arteriolar and venular calibres from optic-disc-centred digital fundus photographs using the SIVA software (SIVA v.4.0, referred to as SIVA-human)[13]. The SIVA software automatically identifies the optic disc, places a grid with reference to the centre of optic disc, identifies the vessel type and calculates the retinal-vessel calibre. All of the technicians were masked to participant characteristics and were responsible for the visual evaluation of the automated measurement and manual intervention, if necessary, according to a standardized protocol. Technicians may have to recentre the grading grid on the optic disc, re-mark vessel type and visually evaluate accuracy of vessel tracing, such that the average grading time per retinal photograph is about 25 min depending on image quality (~10% and ~20–30% manual correction for good and poor quality images, respectively)[13,47]. The protocol was first used in the ARIC study based on the revised Knudtson–Parr–Hubbard formula, and was subsequently modified[11,82]. In brief, individual measurements of the retinal arteriolar and venular calibres of the six largest arterioles and venules were summarized as CRAE and CRVE, respectively[11,13]. SIVA calculates the calibres at two regions: the first region from 0.5 to 1.0 disc diameters away from the disc margin (zone B, $CRAE_B$ and $CRVE_B$) and the second region from 0.5 to 2.0 disc diameters away from the disc margin[13] (zone C, $CRAE_C$ and $CRVE_C$; Supplementary Fig. 1). The intra- and intergrader reliability for the calibre measurements by human technicians using SIVA is high; ICCs of >0.9 were reported previously[13]. There is also a high correlation between SIVA measurement and the commonly used semi-automated retinal-vessel calibre software used in the ARIC study and other cohorts (for example, retinal analysis, integrative vessel analysis); these has been reported previously[83].

Clinical and image datasets. *Training and validation.* We used only gradable retinal photographs with the labelled calibre measurements ($CRAE_B$, $CRAE_C$, $CRVE_B$ and $CRVE_C$ by SIVA-human) for training and validation. Retinal photographs were obtained from the SEED study, which is a population-based study comprising adults residing in Singapore aged 40 to 80 years at baseline, from the following three major ethnic groups: Chinese, Indian and Malay[84–86]. Digital retinal photography was taken using a 45° digital retinal camera (CR-DGi 10D; Canon) after pupil dilation. We divided the SEED dataset into two groups: 80% to train a CNN model (that is SIVA-DLS) to measure CRAE/CRVE; and 20% to validate the performance of CRAE/CRVE measured by SIVA-DLS.

External testing to evaluate agreement between DLS and human measurements and relationship with CVD risk factors. We obtained ten independent retrospectively collected datasets, listed below, that comprise gradable retinal photographs with labelled SIVA-human-measured retinal calibre from different clinical and population settings for assessing the agreement between SIVA-DLS and SIVA-human, and for comparing the relationship with CVD risk factors.

(1) Singapore prospective study program (SP2)
(2) Dunedin multidisciplinary health and development study (Dunedin study)
(3) Hong Kong children eye study (HKCES)
(4) Australian heart eye study (AHES)
(5) Retinal imaging in chest pain study (RICP study)
(6) Retinal imaging in renal disease study (IRED study)
(7) Chinese University of Hong Kong sight-threatening diabetic retinopathy study (CUHK-STDR study)
(8) Growing up in Singapore towards healthy outcomes study (GUSTO study)
(9) Singapore integrated diabetic retinopathy program (SiDRP)
(10) Cardiovascular disease screening using retinal vascular imaging study (CVD screening study)

SP2 was a population-based cohort study of participants aged 24–95 years living in Singapore who attended a detailed health examination[87]. A subgroup of healthy participants from SP2 were randomly selected, and was defined as no history/presence of stroke, heart disease, diabetes mellitus, uncontrolled hypertension, obesity, current smoking, high myopia, glaucoma or any retinal diseases[88]. The Dunedin study was a prospective cohort study based on a population-representative 1972–1973 birth cohort from New Zealand of 93% white European ancestry[89].

HKCES was a population-based cohort study of eye conditions in children of grade 1 to grade 3 (aged about 6–8 years) from primary schools in Hong Kong[90]. AHES was a hospital-based study of patients with symptoms for assessing suspected coronary artery disease by coronary angiography at Westmead hospital in Sydney, Australia[91]. The RICP study was a hospital-based study of retinal abnormalities in patients with low-risk acute coronary syndrome who had chest pain (thrombolysis in myocardial infarction risk score <3) presenting to the Emergency Department, National University Hospital, Singapore. The IRED study was a hospital-based study of retinal abnormalities in patients with chronic kidney disease recruited from nephrology clinics in National University Hospital in Singapore. The CUHK-STDR study was a hospital-based study of diabetic retinopathy in participants with diabetes recruited from the Hong Kong Eye Hospital in Hong Kong[92]. The GUSTO study was an ongoing prospective birth cohort, aiming to study the diet and lifestyle of mothers during pregnancy and the long-term effects of these factors on their childrens development after birth[93]. SiDRP was a community-based diabetic retinopathy screening program for participants with diabetes in public primary care clinics in Singapore[94]. CVD screening study was a community-based study of CVD screening using retinal vascular imaging in participants older than 40 years without a history of stroke and heart disease recruited from a community-based clinic in Singapore[95].

In these datasets, except for the cohorts with diabetes (CUHK-STDR and SiDRP), retinal photograph of the right eye of each participant was selected; if the right eye photograph was ungradable, the measurement was performed on the left eye. In the diabetic cohorts, one eye was randomly selected for the current study.

We obtained individual CVD risk factors from these studies as far as possible. For most, systolic and diastolic blood pressures were measured using digital blood pressure monitors. MABP was calculated as two-thirds of diastolic plus one-third of systolic blood pressure. BMI was calculated from measured weight and height. Smoking status was collected from interviews; current smokers were defined as those who were currently smoking any number of cigarettes (that is, current versus past smoker/never smoked). Glycated-haemoglobin (%) and total cholesterol ($mmol\,l^{-1}$) were measured from venous samples at hospital reference laboratories.

External testing to evaluate the relationship between the DLS measurements and the incident CVD, mortality and CVD risk factors. We applied the SIVA-DLS to retrospectively measure CRAE and CRVE from baseline gradable retinal photographs in the following external test sets, comprising cohorts with prospective data to evaluate incident CVD events, mortality in addition to CVD risk factors.

(1) The SEED study
(2) The BES
(3) The UK Biobank
(4) The KSH study
(5) The Austin health study

In the SEED study, incident CVD event was obtained by linking with the stroke, myocardial infarction and mortality cases registered by National Registry of Diseases Office, Singapore, by record linkage. An incident CVD event was defined as a newly diagnosed clinical stroke or myocardial infarction or CVD mortality[33]. Stroke and myocardial infarction data and mortality data that were recorded during the period between baseline examination and 31 December 2015, and between baseline examination and 31 May 2017, respectively, were used. We examined the associations between SIVA-DLS measurements at baseline with incident CVD events. The BES was a population-based study that included 4,439 of 5,324 participants (aged at least 40 years) who were initially examined in 2001 in northern China[96]. In 2011, the study was repeated by reinviting all of the participants from the survey from 2001. Information about whether the individuals who participated in the baseline survey in 2001 but did not participate in the follow-up examination in 2011 were alive or dead was obtained by house visits, phone, asking neighbours and contacting the local municipal authorities[96]. We examined the associations between SIVA-DLS measurements at baseline with ten-year all-cause mortality. The UK Biobank is a large, multisite study of >500,000 UK residents aged between 40 and 69 years who were registered with the National Health Service to examine genetic and environmental risk factors for complex diseases of middle and old age in the UK population, with a subset of participants who underwent retinal photography and other ocular examination[97]. The KSH study was a medical health check-up program for promoting employee health through regular examinations and to enhance the early detection of existing diseases, with a subgroup of participants who underwent transcranial Doppler ultrasonography and retinal photography[98]. The Austin Study was a hospital-based study of patients with risk factors of atherosclerosis and coronary artery disease presenting to cardiology outpatient clinics in Austin Health in Heidelberg, Australia[99]. We correlated SIVA-DLS measurements with available CVD risk factors in these cohorts.

As there was originally no label on gradability in the UK Biobank dataset, we applied a recently developed deep-learning algorithm to filter out ungradable retinal photographs (T.H.R. et al., unpublished data). Using this algorithm, we were able to use 45,644 gradable retinal photographs from the UK Biobank in our current analysis.

Prospective feasibility study to test the performance of SIVA-DLS in a clinical setting. We developed a prototype interface of SIVA-DLS with fully-automated measurements of $CRAE_B$, $CRAE_C$, $CRVE_B$ and $CRVE_C$ as outputs, and conducted a prospective feasibility study to test the performance of the SIVA-DLS using this interface. Between 1 December 2019 and 31 December 2019, we included 100 participants who were undergoing a telemedicine-based diabetic retinopathy screening program as part of the SiDRP[94]. For each participant, retinal photographs from both eyes were captured using a retinal camera (Topcon NW8, Topcon) and were sent to the SNEC Ocular Reading Center (SORC) for diabetic retinopathy assessment using a secured IT infrastructure, as described previously[94]. The retinal photograph of the right eye was graded by a trained technician using SIVA-human in the SORC. The same retinal photograph was also uploaded to the SIVA-DLS through the interface for fully-automated retinal-vessel calibre measurement.

Development of the DLS CNN. SIVA-DLS uses a propagation mechanism that enables the deepening and widening of the neural architecture[100]. The input to the SIVA-DLS was a normalized cropped retinal image (512 pixels×512 pixels), and the output was a fully connected layer with the predicted calibre measurements ($CRAE_B$, $CRAE_C$, $CRVE_B$ and $CRVE_C$) of the input image. The fully connected layer learned proper weights according to different features for obtaining retinal-vessel calibre values at different zones. It gives high weight to features within zone B when calculating $CRAE_B$ and $CRVE_B$, and high weight to features within zone C when calculating $CRAE_C$ and $CRVE_C$. SIVA-DLS comprises five dense blocks alternating with transition layers to downsample the features. Each dense block is a series of 100 cAdd units packed with two types of convolutions (1×1 and 3×3). Every transition layer is a 1×1 convolution with pooling. Supplementary Figure 2 shows the detailed architecture of the SIVA-DLS. We trained the SIVA-DLS for 300 epochs with a batch size of 80, and used a cosine-shaped learning rate from 0.1 to 0. To reduce overfitting, we performed data augmentation by randomly rotating the training images by 0–180°, and also introduced dropout with a rate of 0.2 as a regularization technique. The heat maps of selected high-level layer and low-level layer of the CNN generated by integrated gradients in Fig. 1 highlights the boundaries of the arterioles and venules that the SIVA-DLS focuses on to calibrate the CRAE and CRVE prediction.

Statistical analysis. We performed the following statistical analyses using standard statistical software (SPSS, v.24.0; STATA, v.13; MedCalc, v.18.10.2).

First, to assess the validity of SIVA-DLS using SIVA-human data, we examined the distribution and agreement in calibre measurements ($CRAE_B$, $CRAE_C$, $CRVE_B$ and $CRVE_C$) using the ICC and the Bland–Altman plots between the two methods.

Second, we constructed multivariable linear and logistic regression models to correlate each CVD risk factor (age, gender, MABP, BMI, glycated-haemoglobin level, smoking and total cholesterol level as dependent variables) and retinal calibre measurements (independent variables and analysed as per each s.d. increase), adjusting for age (except the models of age), gender (except the models of gender) and fellow calibre (that is, CRVE was included as an independent variable in the model for CRAE, and vice versa). Goodness of fit of the regression models was evaluated by R_N^2 (ref. [101]). R_N^2 is defined as:

$$R_N^2 = \frac{1 - \left\{\frac{\text{Likelihood(regression model)}}{\text{Likelihood(null model)}}\right\}^{\frac{2}{n}}}{1 - \text{Likelihood(null model)}^{2/n}}$$

Where the numerator $1 - \left\{\frac{\text{Likelihood(regression model)}}{\text{Likelihood(null model)}}\right\}^{\frac{2}{n}}$ is also known as the Cox and Snell's pseudo R^2, which measures the ratio of improvement in likelihood comparing the regression model over the null model with only the interception term; n represents the number of observations in the dataset such that the nth square root is served as a geometric mean of the likelihood for each observation. As Cox and Snell's pseudo R^2 ranges from 0 to a value of less than 1, R_N^2 introduced the denominator $\frac{1}{1-\text{Likelihood(null model)}^{2/n}}$ to rescale the index from 0 to 1. We then used bootstrap t-test with 5,000 replicates to compare the two regression models (that is, SIVA-DLS and SIVA-human). If the P value of the bootstrap t-test is less than 0.05, this indicates that the model with larger R_N^2 predicts the CVD risk factors significantly better compared with the other model.

Third, we examined SIVA-DLS measurements with incident clinical CVD outcomes and mortality. In the SEED, we performed Cox proportional-hazards regression to estimate the HR between SIVA-DLS measurements with incident CVD. In the BES, we performed logistic regression models to estimate the OR between SIVA-DLS measurements to ten-year all-cause mortality (as data on time-to-event were unavailable). In the KSH study, we performed multiple linear regression models to correlate each of the CVD risk factors and SIVA-DLS measurements.

Finally, to test the outcomes of the prospective feasibility study, we assessed the agreement of calibre measurements between the two methods using the ICC.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. The deidentified individual-participant data and data on the evaluation of retinal photographs used in the SIVA-DLS are available on request from the corresponding author or C.Y. Cheung (e-mail: carolcheung@cuhk.edu.hk).

Code availability

The custom code is currently available only on request because it is under a patent examination process.

Received: 12 September 2019; Accepted: 8 September 2020;
Published online: 12 October 2020

References

1. Williams, B. et al. British Hypertension Society guidelines for hypertension management 2004 (BHS-IV): summary. *BMJ* **328**, 634–640 (2004).
2. Chobanian, A. V. et al. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the JNC 7 report. *JAMA* **289**, 2560–2572 (2003).
3. Mansia, G. et al. 2007 ESH-ESC guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Blood Press.* **16**, 135–232 (2007).
4. *Hypertension in Adults: Diagnosis and Management* Clinical guideline [CG127], vol. 2019 (National Institute of Health and Clinical Excellence, 2011).
5. Wong, T. Y. & Mitchell, P. Hypertensive retinopathy. *N. Engl. J. Med* **351**, 2310–2317 (2004).
6. Liew, G., Wang, J. J., Mitchell, P. & Wong, T. Y. Retinal vascular imaging: a new tool in microvascular disease research. *Circ. Cardiovasc. Imaging* **1**, 156–161 (2008).
7. Wong, T. Y. et al. Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality. *Surv. Ophthalmol.* **46**, 59–80 (2001).
8. Cheung, C. Y., Ikram, M. K., Chen, C. & Wong, T. Y. Imaging retina to study dementia and stroke. *Prog. Retin. Eye Res.* **57**, 89–107 (2017).
9. Cheung, C. Y., Ikram, M. K., Sabanayagam, C. & Wong, T. Y. Retinal microvasculature as a model to study the manifestations of hypertension. *Hypertension* **60**, 1094–1103 (2012).
10. Wang, S. B. et al. A spectrum of retinal vasculature measures and coronary artery disease. *Atherosclerosis* **268**, 215–224 (2018).
11. Hubbard, L. D. et al. Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the Atherosclerosis Risk in Communities Study. *Ophthalmology* **106**, 2269–2280 (1999).
12. Wong, T. Y. et al. Computer-assisted measurement of retinal vessel diameters in the Beaver Dam Eye Study: methodology, correlation between eyes, and effect of refractive errors. *Ophthalmology* **111**, 1183–1190 (2004).
13. Cheung, C. Y. et al. A new method to measure peripheral retinal vascular caliber over an extended area. *Microcirculation* **17**, 495–503 (2010).
14. McGrory, S. et al. Towards standardization of quantitative retinal vascular parameters: comparison of SIVA and VAMPIRE measurements in the Lothian birth cohort 1936. *Transl. Vis. Sci. Technol.* **7**, 12 (2018).
15. Cheung, C. Y., Ikram, M. K., Klein, R. & Wong, T. Y. The clinical implications of recent studies on the structure and function of the retinal microvasculature in diabetes. *Diabetologia* **58**, 871–885 (2015).
16. Sun, C., Wang, J. J., Mackey, D. A. & Wong, T. Y. Retinal vascular caliber: systemic, environmental, and genetic associations. *Surv. Ophthalmol.* **54**, 74–95 (2009).
17. Ding, J. et al. Retinal vascular caliber and the development of hypertension: a meta-analysis of individual participant data. *J. Hypertens.* **32**, 207–215 (2014).
18. Sabanayagam, C. et al. Retinal microvascular calibre and risk of diabetes mellitus: a systematic review and participant-level meta-analysis. *Diabetologia* **58**, 2476–2485 (2015).
19. Wong, T. Y. et al. Retinal vascular caliber, cardiovascular risk factors, and inflammation: the multi-ethnic study of atherosclerosis (MESA). *Invest. Ophthalmol. Vis. Sci.* **47**, 2341–2350 (2006).
20. Cheung, N. et al. Retinal arteriolar narrowing and left ventricular remodeling: the multi-ethnic study of atherosclerosis. *J. Am. Coll. Cardiol.* **50**, 48–55 (2007).
21. De Silva, D. A. et al. The relationship between aortic stiffness and changes in retinal microvessels among Asian ischemic stroke patients. *J. Hum. Hypertens.* **26**, 716–722 (2012).
22. McGeechan, K. et al. Meta-analysis: retinal vessel caliber and risk for coronary heart disease. *Ann. Intern. Med.* **151**, 404–413 (2009).
23. McGeechan, K. et al. Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. *Am. J. Epidemiol.* **170**, 1323–1332 (2009).
24. Seidelmann, S. B. et al. Retinal vessel calibers in predicting long-term cardiovascular outcomes: the atherosclerosis risk in communities study. *Circulation* **134**, 1328–1338 (2016).
25. Wong, T. Y. et al. Retinal arteriolar narrowing and risk of coronary heart disease in men and women. The atherosclerosis risk in communities study. *JAMA* **287**, 1153–1159 (2002).
26. Wong, T. Y. et al. Retinal microvascular abnormalities and incident stroke: the atherosclerosis risk in communities study. *Lancet* **358**, 1134–1140 (2001).
27. Ikram, M. K. et al. Retinal vessel diameters and risk of stroke: the Rotterdam study. *Neurology* **66**, 1339–1343 (2006).
28. Kawasaki, R. et al. Retinal microvascular signs and risk of stroke: the multi-ethnic study of atherosclerosis (MESA). *Stroke* **43**, 3245–3251 (2012).
29. McGeechan, K. et al. Risk prediction of coronary heart disease based on retinal vascular caliber (from the atherosclerosis risk in communities [ARIC] study). *Am. J. Cardiol.* **102**, 58–63 (2008).
30. Wang, J. J. et al. Retinal vessel diameter and cardiovascular mortality: pooled data analysis from two older populations. *Eur. Heart J.* **28**, 1984–1992 (2007).
31. Mutlu, U. et al. Retinal microvasculature is associated with long-term survival in the general adult Dutch population. *Hypertension* **67**, 281–287 (2016).
32. Cheung, C. Y. et al. Retinal microvascular changes and risk of stroke: the Singapore Malay eye study. *Stroke* **44**, 2402–2408 (2013).
33. Ho, H. et al. Retinopathy signs improved prediction and reclassification of cardiovascular disease risk in diabetes: a prospective cohort study. *Sci. Rep.* **7**, 41492 (2017).
34. Wong, T. Y. et al. Quantitative retinal venular caliber and risk of cardiovascular disease in older persons: the cardiovascular health study. *Arch. Intern. Med.* **166**, 2388–2394 (2006).
35. Wong, T. Y. et al. Retinal microvascular abnormalities and 10-year cardiovascular mortality: a population-based case-control study. *Ophthalmology* **110**, 933–940 (2003).
36. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **380**, 1347–1358 (2019).
37. Chen, J. H. & Asch, S. M. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N. Engl. J. Med.* **376**, 2507–2509 (2017).
38. Obermeyer, Z. & Emanuel, E. J. Predicting the future—big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **375**, 1216–1219 (2016).
39. Ting, D. S. W. et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* **318**, 2211–2223 (2017).
40. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).
41. De Fauw, J. et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018).
42. Kermany, D. S. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**, 1122–1131 (2018).
43. Ran, A. R. et al. Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis. *Lancet Digit. Health* **1**, e172–e182 (2019).
44. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
45. He, J. et al. The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* **25**, 30–36 (2019).
46. Sharafi, S. M. et al. Vascular retinal biomarkers improves the detection of the likely cerebral amyloid status from hyperspectral retinal images. *Alzheimers Dement.* **5**, 610–617 (2019).
47. Chan, V. T. T. et al. Using retinal imaging to study dementia. *J. Vis. Exp.* **129**, 56137 (2017).
48. Welikala, R. A. et al. Automated arteriole and venule classification using deep learning for retinal images from the UK Biobank cohort. *Comput. Biol. Med.* **90**, 23–32 (2017).
49. Yan, Z., Yang, X. & Cheng, K. T. A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE J. Biomed. Health Inf.* **23**, 1427–1436 (2019).
50. Oliveira, A., Pereira, S. & Silva, C. A. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Syst. Appl.* **112**, 229–242 (2018).
51. Welikala, R. A. et al. Automated quantification of retinal vessel morphometry in the UK biobank cohort. In *Proc. 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)* 1–6 (IEEE, 2017).
52. Maderuelo-Fernandez, J. A. et al. Automatic image analyser to assess retinal vessel calibre (ALTAIR). A new tool to evaluate the thickness, area and length of the vessels of the retina. *Int. J. Med. Inf.* **136**, 104090 (2020).
53. Gunn, R. M. Ophthalmoscopic evidence of (1) arterial changes associated with chronic renal diseases and (2) of increased arterial tension. *Trans. Ophthalmol. Soc. UK* **12**, 124–125 (1982).
54. Nguyen, T. T. & Wong, T. Y. Retinal vascular manifestations of metabolic disorders. *Trends Endocrinol. Metab.* **17**, 262–268 (2006).
55. Wong, T. Y. & Scott, I. U. Clinical practice. Retinal-vein occlusion. *N. Engl. J. Med.* **363**, 2135–2144 (2010).
56. Sharrett, A. R. et al. Retinal arteriolar diameters and elevated blood pressure: the atherosclerosis risk in communities study. *Am. J. Epidemiol.* **150**, 263–270 (1999).

57. Wong, T. Y., Klein, R., Klein, B. E., Meuer, S. M. & Hubbard, L. D. Retinal vessel diameters and their associations with age and blood pressure. *Invest. Ophthalmol. Vis. Sci.* **44**, 4644–4650 (2003).
58. Leung, H. et al. Relationships between age, blood pressure, and retinal vessel diameters in an older population. *Invest. Ophthalmol. Vis. Sci.* **44**, 2900–2904 (2003).
59. Wong, T. Y. et al. Retinal microvascular abnormalities and blood pressure in older people: the cardiovascular health study. *Br. J. Ophthalmol.* **86**, 1007–1013 (2002).
60. Kawasaki, R. et al. Cardiovascular risk factors and retinal microvascular signs in an adult Japanese population: the Funagata study. *Ophthalmology* **113**, 1378–1384 (2006).
61. Ikram, M. K. et al. Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? The Rotterdam study. *Invest. Ophthalmol. Vis. Sci.* **45**, 2129–2134 (2004).
62. Jeganathan, V. S. et al. Effect of blood pressure on the retinal vasculature in a multi-ethnic Asian population. *Hypertens. Res.* **32**, 975–982 (2009).
63. Ikram, M. K. et al. Retinal vessel diameters and risk of impaired fasting glucose or diabetes: the Rotterdam study. *Diabetes* **55**, 506–510 (2006).
64. Nguyen, T. T., Wang, J. J. & Wong, T. Y. Retinal vascular changes in pre-diabetes and prehypertension: new findings and their research and clinical implications. *Diabetes Care* **30**, 2708–2715 (2007).
65. Islam, F. M. et al. Quantitative retinal vascular calibre changes in diabetes and retinopathy: the Singapore Malay eye study. *Eye* **23**, 1719–1724 (2009).
66. Tsai, A. S. et al. Differential association of retinal arteriolar and venular caliber with diabetes and retinopathy. *Diabetes Res. Clin. Pr.* **94**, 291–298 (2011).
67. Cheung, C. Y. et al. Retinal vascular geometry in Asian persons with diabetes and retinopathy. *J. Diabetes Sci. Technol.* **6**, 595–605 (2012).
68. Ting, D. S. W. et al. Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **103**, 167–175 (2019).
69. Wang, F., Kaushal, R. & Khullar, D. Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Ann. Intern. Med.* **172**, 59–60 (2019).
70. Krittanawong, C. et al. Deep learning for cardiovascular medicine: a practical primer. *Eur. Heart J.* **40**, 2058–2073 (2019).
71. Kuan, R. Adopting AI in health care will be slow and difficult. *Harvard Business Review* (18 October 2019).
72. Rizzoni, D. & Muiesan, M. L. Retinal vascular caliber and the development of hypertension: a meta-analysis of individual participant data. *J. Hypertens.* **32**, 225–227 (2014).
73. Houben, A., Berendschot, T. & Grassi, G. Retinal microvascular diameters: normative data and their use in clinical hypertension. *J. Hypertens.* **35**, 1573–1574 (2017).
74. Lee, H. et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat. Biomed. Eng.* **3**, 173–182 (2019).
75. Wang, T. J. et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N. Engl. J. Med* **355**, 2631–2639 (2006).
76. Koenig, W., Lowel, H., Baumert, J. & Meisinger, C. C-reactive protein modulates risk prediction based on the Framingham score: implications for future risk assessment: results from a large cohort study in southern Germany. *Circulation* **109**, 1349–1353 (2004).
77. Wickremasinghe, S. S. et al. Retinal vascular calibre changes after intravitreal bevacizumab or dexamethasone implant treatment for diabetic macular oedema. *Br. J. Ophthalmol.* **101**, 1329–1333 (2017).
78. Wickremasinghe, S. S., Rogers, S. L., Gillies, M. C., Zhu, M. & Wong, T. Y. Retinal vascular caliber changes after intravitreal triamcinolone treatment for diabetic macular edema. *Invest. Ophthalmol. Vis. Sci.* **49**, 4707–4711 (2008).
79. Reid, C. M. et al. Aspirin for the prevention of cognitive decline in the elderly: rationale and design of a neuro-vascular imaging study (ENVIS-ion). *BMC Neurol.* **12**, 3 (2012).
80. Hughes, A. D. et al. Effect of antihypertensive treatment on retinal microvascular changes in hypertension. *J. Hypertens.* **26**, 1703–1707 (2008).
81. Thom, S. et al. Differential effects of antihypertensive treatment on the retinal microcirculation: an Anglo-Scandinavian cardiac outcomes trial substudy. *Hypertension* **54**, 405–408 (2009).
82. Knudtson, M. D. et al. Revised formulas for summarizing retinal vessel diameters. *Curr. Eye Res.* **27**, 143–149 (2003).
83. Yip, W. et al. Comparison of common retinal vessel caliber measurement software and a conversion algorithm. *Transl. Vis. Sci. Technol.* **5**, 11 (2016).
84. Sabanayagam, C. et al. Singapore Indian eye study-2: methodology and impact of migration on systemic and eye outcomes. *Clin. Exp. Ophthalmol.* **45**, 779–789 (2017).
85. Foong, A. W. et al. Rationale and methodology for a population-based study of eye diseases in Malay people: the Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol.* **14**, 25–35 (2007).
86. Lavanya, R. et al. Methodology of the Singapore Indian Chinese cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiol.* **16**, 325–336 (2009).
87. Cheung, C. Y. et al. C-reactive protein and retinal microvascular caliber in a multiethnic Asian population. *Am. J. Epidemiol.* **171**, 206–213 (2010).
88. Li, X. et al. Racial differences in retinal vessel geometric characteristics: a multiethnic study in healthy Asians. *Invest. Ophthalmol. Vis. Sci.* **54**, 3650–3656 (2013).
89. Shalev, I. et al. Retinal vessel caliber and lifelong neuropsychological functioning: retinal imaging as an investigative tool for cognitive epidemiology. *Psychol. Sci.* **24**, 1198–1207 (2013).
90. Cheung, C. Y. et al. Quantitative retinal microvasculature in children using swept-source optical coherence tomography: the Hong Kong children eye study. *Br. J. Ophthalmol.* **103**, 672–679 (2018).
91. Gopinath, B. et al. Retinal vascular geometry and the prevalence of atrial fibrillation and heart failure in a clinic-based sample. *Heart Lung Circ.* **28**, 1631–1637 (2018).
92. Sun, Z. et al. OCT angiography metrics predict progression of diabetic retinopathy and development of diabetic macular edema: a prospective study. *Ophthalmology* **126**, 1675–1684 (2019).
93. Li, L. J. et al. Gestational retinal microvasculature and the risk of 5 year postpartum abnormal glucose metabolism. *Diabetologia* **60**, 2368–2376 (2017).
94. Nguyen, H. V. et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology* **123**, 2571–2580 (2016).
95. Yip, W. et al. Reliability and determinants of retinal vessel oximetry measurements in healthy eyes. *Invest. Ophthalmol. Vis. Sci.* **55**, 7104–7110 (2014).
96. Wang, Y. X., Zhang, J. S., You, Q. S., Xu, L. & Jonas, J. B. Ocular diseases and 10-year mortality: the Beijing eye study 2001/2011. *Acta Ophthalmol.* **92**, e424–e428 (2014).
97. Chua, S. Y. L. et al. Cohort profile: design and methods in the eye and vision consortium of UK Biobank. *BMJ Open* **9**, e025077 (2019).
98. Rhee, E. J., Chung, P. W., Wong, T. Y. & Song, S. J. Relationship of retinal vascular caliber variation with intracranial arterial stenosis. *Microvasc. Res.* **108**, 64–68 (2016).
99. Al-Fiadh, A. H. et al. Usefulness of retinal microvascular endothelial dysfunction as a predictor of coronary artery disease. *Am. J. Cardiol.* **115**, 609–613 (2015).
100. Xu, D., Lee, M. L. & Hsu, W. Propagation mechanism for deep and wide neural networks. In *Proc. 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 9220–9228 (IEEE/CVF, 2019).
101. Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).

Acknowledgements

We thank all of the staff at the SNEC Ocular Reading Centre (SORC) for their contribution to this study. We acknowledge funding support from Singapore Ministry of Health's National Medical Research Council (NMRC) grants OFLCG/001/2017, NMRC/STaR/003/2008, NMRC/STaR/0016/2013 and NMRC/CIRG/1371/2013. This research is also supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG Award no. AISG-GC-2019-001). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the National Research Foundation, Singapore.

Author contributions

C.Y. Cheung, D.X., W.H., M.L.L. and T.Y.W. contributed to study conception and design. D.X., W.H. and M.L.L. coded and optimized the DLS. C.Y. Cheung and M.Y. analysed the data. C.Y. Cheung and T.Y.W. contributed to the initial draft of the manuscript. C.Y. Cheung and T.Y.W. had full access to the data, vouch for the integrity of the data and the adherence to the study protocol, and were responsible for the decision to submit the manuscript. All of the authors contributed to data collection and interpretation, and revision of the manuscript for important content.

Competing interests

C.Y. Cheung, D.X., W.H., M.L.L. and T.Y.W. are filing a patent (World Intellectual Property Organization International Bureau; International publication no. WO 2020/167251) for the DLS described in this study.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41551-020-00626-4.

Correspondence and requests for materials should be addressed to T.Y.W.

nature research

Corresponding author(s): Tien Y. Wong

Last updated by author(s): Aug 29, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The Singapore I Vessel Assessment software (SIVA version 4.0) was used to measure retinal-vessel calibre.
Data analysis	We performed the statistical analyses using standard statistical software (SPSS, version 24.0, Chicago; STATA, version 13, Texas; MedCalc, version 18.10.2, Belgium).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. The de-identified individual-participant data and data on the evaluation of retinal photographs used in the SIVA-DLS are available on request from the corresponding author or the first author.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	On the basis of our previous experience and published literature, we know that deep learning requires on the order of tens of thousands or hundreds of thousands of examples. As such, we included as much available data as possible from these datasets.
Data exclusions	We excluded any retinal photographs that were of poor quality or with missing data.
Replication	We ensured that our results generalized over external testing datasets that were not used for training.
Randomization	Samples were randomly allocated to the training and validation datasets.
Blinding	Because the study was retrospective, no blinding was necessary. Splits for training and validation were random and automatically generated.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The deep-learning analysis involved retrospective training and validation. For the training and validation of SIVA-DLS, retinal photographs were retrospectively obtained from the Singapore Epidemiology of Eye Diseases (SEED) study, a population-based study comprising of adults residing in Singapore aged 40 to 80 years at baseline, from 3 major ethnic groups: Chinese, Indians and Malays. For the testing of SIVA-DLS, retinal photographs from 10 independent retrospectively collected dataset were used: 1) Singapore Prospective Study Program (SP2); 2) Dunedin Multidisciplinary Health and Development Study (Dunedin study); 3) Hong Kong Children Eye Study (HKCES); 4) Australian Heart Eye Study (AHES); 5) Retinal imaging in Chest Pain study (RICP study); 6) Retinal imaging study in renal patients (IRED study); 7) Chinese University of Hong Kong Sight-Threatening Diabetic Retinopathy study (CUHK-STDR study); 8) Growing Up in Singapore Towards Healthy Outcomes study (GUSTO study); 9) Singapore Integrated Diabetic Retinopathy Program (SiDRP) and 10) Cardiovascular Disease Screening using Retinal Vascular Imaging study (CVD screening study). For external testing to evaluate tthe relationship of DLS with incident CVD, mortality and CVD risk factors, 5 datasets were used: 1) The SEED study; 2) The Beijing Eye Study (BES); 3) The United Kingdom Biobank (UK Biobank); 4) The Kangbuk Samsung Health Study (KSH study) and 5) The Austin Health Study (Austin study).

Recruitment	The SEED study was a population-based study comprising adults residing in Singapore aged 40 to 80 years at baseline, from 3 major ethnic groups: Chinese, Indians and Malays. SP2 was a population-based cohort study of participants aged 24–95 years living in Singapore who attended a detailed health examination. A sub-group of healthy subjects from SP2 were randomly selected, which was defined as no history/presence of stroke, heart disease, diabetes mellitus, uncontrolled hypertension, obesity, current smoking, high myopia, glaucoma or any retinal diseases. Dunedin study was a prospective cohort study based on a population-representative 1972–1973 birth cohort from New Zealand, 93% White European ancestry. HKCES was a population-based cohort study of eye conditions in children of Grade 1 to Grade 3 (aged about 6–8 years) from primary schools in Hong Kong. AHES was a hospital-based study of symptomatic patients for assessment of suspected coronary artery disease by coronary angiography presenting to Westmead hospital in Sydney, Australia. RICP study was a hospital-based study of retinal abnormalities in low-risk acute coronary syndrome patients with chest pain (thrombolysis in myocardial infarction [TIMI] risk score <3) presenting to the Emergency Department, National University Hospital, Singapore. IRED study was a hospital-based study of retinal abnormalities in patients with chronic kidney disease recruited from nephrology clinics in National University Hospital in Singapore. CUHK-STDR study was a hospital-based study of diabetic retinopathy in subjects with diabetes recruited from Hong Kong Eye Hospital in Hong Kong. GUSTO study was an ongoing prospective birth cohort, aiming to study mothers' diet and lifestyle during pregnancy and their long-term effects on their offspring's development after birth. SiDRP was a community-based diabetic retinopathy screening program for subjects with diabetes in public primary care clinics in Singapore. CVD screening study was a community-based study of CVD screening using retinal vascular imaging in subjects older than 40 years without a history of stroke and heart disease recruited from a community-based clinic in Singapore. The BES was a population-based study that included 4,439 of 5,324 subjects (aged ≥40 years) who were initially examined in 2001 in Northern China. In the year 2011, the study was repeated by re-inviting all participants from the survey from 2001. Information about being alive or dead of subjects, who participated in the baseline survey in 2001 but did not participate in the follow-up examination in 2011, was obtained by house visits, by phone, by asking neighbours and by contacting the local municipal authorities. We examined the associations between SIVA-DLS measurement at baseline with 10-year all-cause mortality. The UK Biobank is a large, multisite, study of >500,000 UK residents aged between 40 and 69 years who were registered with the National Health Service to examine genetic and environmental risk factors for complex diseases of middle and old age in the UK population, with a subset of participants underwent retinal photography and other ocular examination. The KSH study was a medical health check-up program for promoting employee health through regular examinations and to enhance early detection of existing diseases, with a subgroup of participants who underwent transcranial Doppler ultrasonography and retinal photography. The Austin Study was a hospital-based study of patients with risk factors of atherosclerosis and coronary artery disease presenting to cardiology outpatient clinics in Austin Health in Heidelberg, Australia. We correlated SIVA-DLS measurement with available CVD risk factors in these cohorts.
Ethics oversight	This study was approved by the human ethics boards of SingHealth, Singapore, the National Healthcare Group, Singapore, the Western Sydney Local Health Network, Australia, the Lower South Regional Ethics Committee, New Zealand, the Kowloon Central/East Research Ethics Committee, Hong Kong, the Medical Ethics Committee of the Beijing Tongren Hospital, China, the North West Multicentre Research Ethics Committee, UK, the Institutional Review Board of Kangbuk Samsung Hospital, South Korea, and the Human Research Ethics Committee of Austin Health, Australia. All studies were conducted in accordance with the Declaration of Helsinki, with written informed consent obtained from each participant.

Note that full information on the approval of the study protocol must also be provided in the manuscript.