Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms

Tyler Hyungtaek Rim, Geunyoung Lee*, Youngnam Kim*, Yih-Chung Tham*, Chan Joo Lee, Su Jung Baik, Young Ah Kim, Marco Yu, Mihir Deshmukh, Byoung Kwon Lee, Sungha Park, Hyeon Chang Kim, Charumathi Sabayanagam, Daniel S W Ting, Ya Xing Wang, Jost B Jonas, Sung Soo Kim†, Tien Yin Wong†, Ching-Yu Cheng†*

Summary

Background The application of deep learning to retinal photographs has yielded promising results in predicting age, sex, blood pressure, and haematological parameters. However, the broader applicability of retinal photograph-based deep learning for predicting other systemic biomarkers and the generalisability of this approach to various populations remains unexplored.

Methods With use of 236 257 retinal photographs from seven diverse Asian and European cohorts (two health screening centres in South Korea, the Beijing Eye Study, three cohorts in the Singapore Epidemiology of Eye Diseases study, and the UK Biobank), we evaluated the capacities of 47 deep-learning algorithms to predict 47 systemic biomarkers as outcome variables, including demographic factors (age and sex); body composition measurements; blood pressure; haematological parameters; lipid profiles; biochemical measures; biomarkers related to liver function, thyroid function, kidney function, and inflammation; and diabetes. The standard neural network architecture of VGG16 was adopted for model development.

Findings In addition to previously reported systemic biomarkers, we showed quantification of body composition indices (muscle mass, height, and bodyweight) and creatinine from retinal photographs. Body muscle mass could be predicted with an R^2 of 0·52 (95% CI 0·51–0·53) in the internal test set, and of 0·33 (0·30–0·35) in one external test set with muscle mass measurement available. The R^2 value for the prediction of height was 0·42 (0·40–0·43), of bodyweight was 0·36 (0·34–0·37), and of creatinine was 0·38 (0·37–0·40) in the internal test set. However, the performances were poorer in external test sets (with the lowest performance in the European cohort), with R^2 values ranging between 0·08 and 0·28 for height, 0·04 and 0·19 for bodyweight, and 0·01 and 0·26 for creatinine. Of the 47 systemic biomarkers, 37 could not be predicted well from retinal photographs via deep learning ($R^2 \leq 0·14$ across all external test sets).

Interpretation Our work provides new insights into the potential use of retinal photographs to predict systemic biomarkers, including body composition indices and serum creatinine, using deep learning in populations with a similar ethnic background. Further evaluations are warranted to validate these findings and evaluate the clinical utility of these algorithms.

Funding Agency for Science, Technology, and Research and National Medical Research Council, Singapore; Korea Institute for Advancement of Technology.

Lancet Digital Health 2020; 2: e526–36

*Co-first authors

†Contributed equally

Singapore Eye Research Institute, Singapore National Eye Centre, Singapore (T H Rim MD, Y-C Tham PhD, M Yu PhD, M Deshmukh MSc, C Sabayanagam PhD, D S W Ting MD, Prof T Y Wong MD, Prof C-Y Cheng MD); **Ophthalmology and Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore** (T H Rim, Y-C Tham, C Sabayanagam, D S W Ting, Prof T Y Wong, Prof C-Y Cheng); **Department of Ophthalmology, Institute of Vision Research, Severance Hospital** (T H Rim, Prof S S Kim MD), **Division of Cardiology, Severance Cardiovascular Hospital** (C J Lee MD, Prof S Park MD), **Healthcare Research Team, Health Promotion Center, Severance Gangnam Hospital** (S J Baik MD), **Division of Medical Information and Technology** (Y A Kim PhD) **and Division of Cardiology, Severance Gangnam Hospital** (Prof B K Lee MD), **Integrated Research Center for Cerebrovascular and Cardiovascular Disease** (Prof S Park), **and Department of Preventive Medicine** (Prof H C Kim MD), **Yonsei University College of Medicine, Seoul, South Korea; Medi Whale, Seoul, South Korea** (G Lee BS, Y Kim MSc); **Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China** (Y X Wang MD); **and Department of Ophthalmology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany** (Prof J B Jonas MD)

Introduction

The retina is the only organ that allows direct, non-invasive, in-vivo visualisation of the microvasculature and neural tissues. It thus affords a unique opportunity for the non-invasive detection of systemic vascular and neurological diseases.[1] In recent decades, our understanding of retina–systemic disease relationships has relied on classic epidemiological studies based on observable, human-defined retinal features (eg, retinopathy or retinal vascular calibre).[2] The potential discovery of unobservable retinal features associated with systemic diseases has been enhanced by advances in artificial intelligence technology, specifically deep learning.[3] Deep learning can be used to predict many systemic biomarkers using retinal photographs, obviating the need for observable, precharacterised retinal features.[4] Using UK Biobank data, Poplin and colleagues showed that deep-learning models could predict six cardiovascular risk factors using only retinal photographs with reasonable accuracy.[5] Using these photographs, deep learning could accurately predict features such as age and sex, which otherwise cannot be similarly identified by human eyes alone. Additionally, deep learning was reported to predict serum haemoglobin concentrations from retinal photographs, suggesting a potential new model for automated anaemia screening.[6] More recently, several studies on deep learning-predicted systemic conditions from retinal photographs have been published.[7–11]

Correspondence to:
Prof Sung Soo Kim, Department of Ophthalmology, Institute of Vision Research, Severance Hospital, Yonsei University College of Medicine, Seoul 03722, South Korea
semekim@yuhs.ac
or
Prof Ching-Yu Cheng, Singapore Eye Research Institute, Singapore National Eye Centre, Singapore 169856
chingyu.cheng@duke-nus.edu.sg

Research in context

Evidence before this study

We searched PubMed and preprint archives for articles published in English that contained information on applying deep learning to retinal photographs for prediction of systemic biomarkers, published up to Aug 14, 2020. Our search terms included "retina", "deep learning", "cardiovascular", "anaemia", and "systemic". Previous studies reported that deep learning could predict cardiovascular risk factors, including age, sex, glycated haemoglobin A_{1c}, blood pressure, smoking status, and body-mass index, and haematological parameters, including serum haemoglobin concentration, haematocrit, and red blood cell count, from retinal photographs. However, application of deep learning in this field is still in the early stages, and the broader capacity of deep learning for predicting other systemic biomarkers based on retinal photographs remains unclear.

Added value of this study

We developed deep-learning algorithms to predict 47 systemic biomarkers from retinal photographs using seven diverse Asian and European cohorts. In addition to previously reported systemic biomarkers, we showed quantification of body composition measures (muscle mass, height, and bodyweight) and serum creatinine concentration from retinal photographs. Nevertheless, the performances were generally lower when validated in external test sets. Our study additionally provided information on systemic biomarkers that could not be predicted from retinal photographs, including those related to thyroid function, biochemical measures, haematological parameters other than haematocrit and haemoglobin, and C-reactive protein.

Implications of all the available evidence

The current findings add to evidence on the applicability of retinal photograph-based deep learning for predicting systemic biomarkers. Overall, deep learning can predict age, sex, and blood pressure well, and shows promising performance in predicting body composition indices and serum creatinine concentration. However, further research is warranted to assess the clinical utility of these predictions. If validated, these algorithms could be implemented as add-on tests in primary-care settings equipped with retinal cameras, such as in diabetic retinopathy screening centres.

Nevertheless, the applications of artificial intelligence and deep learning in this field are in the early stages. The broader capacity of deep learning to predict other systemic biomarkers based on retinal photographs, as well as the generalisation of the trained deep-learning algorithm,[12] remains unexplored. Hence, in this study, we developed 47 deep-learning algorithms to estimate 47 systemic biomarkers and evaluated the algorithms' performance with use of external datasets from diverse populations.

Methods

Study population and datasets

To develop and validate the deep-learning algorithms, we used cross-sectional datasets from 236 257 retinal photographs from 72 890 participants enrolled from two health screening centres in Seoul, South Korea (Severance Main Hospital and Severance Gangnam Hospital), one Beijing Eye Study cohort,[13] three Singapore Epidemiology of Eye Diseases (SEED) study cohorts from Singapore (Chinese, Indian, and Malay participants),[14,15] and the UK Biobank.[16] The sample selection is described in detail in the appendix (pp 6–8).

See Online for appendix

Retinal photograph-based deep-learning algorithms have been shown to have potential use as a screening tool for cardiovascular risk.[5] Therefore, the population of interest was the general population in the communities. We used the dataset from Severance Main Hospital to train the deep-learning algorithms, because this dataset included the most biomarkers.

The dataset from Severance Main Hospital was divided randomly into developmental and internal test sets at an 8:2 ratio based on the individual level. The developmental set was again divided randomly into training (for updating model parameters) and validation sets (for hyperparameter tuning and model selection). To prevent model overfitting, we divided the dataset by individuals rather than by examinations. Therefore, multiple examinations from the same individual on different dates were not distributed across the training, validation, and internal test sets. We tested the algorithms developed from Severance Main Hospital on four external test sets, including those from the Severance Gangnam Hospital, the Beijing Eye Study, the SEED study, and the UK Biobank.

This study was approved by the Institutional Review Board of Severance Hospital, Yonsei University College of Medicine, Seoul, South Korea. Each included epidemiological study obtained ethical approval, and all participants provided written informed consent. The study adhered to the tenets of the Declaration of Helsinki. All datasets and retinal photographs were de-identified according to the Health Insurance Portability and Accountability Act standards.

Retinal photography

Details of retinal photography and quality assessment are provided in the appendix (p 1). Briefly, six different types of retinal camera were used across study sites. We used only gradable retinal photographs. The gradability of the retinal photographs was determined by human graders except for those from the UK Biobank, where retinal photographs were sorted[17] and graded using an in-house deep-learning algorithm for image quality

assessment. Retinal diseases, including age-related macular degeneration, diabetic retinopathy, and referable retinal disease, were defined from retinal photographs except in the UK Biobank, in which the status of retinal diseases was self-reported.

Clinical outcome variables

We included 47 systemic biomarkers as outcome variables. In addition to demographic factors (age and sex), these biomarkers were selected on the basis of: relevance to cardiovascular diseases (blood pressure, body composition, kidney function, lipid profile, diabetes-related measures, and C-reactive protein); evidence of predictability from previous studies (haemoglobin and other haematological parameters); and available data from the blood (biochemical markers and liver and thyroid function markers). We excluded biomarkers that are related to tumours, those measured from urine and stool, or those that reflect acute episodes (eg, recent infection). All clinical examinations, tests, and retinal photography were performed on the same visit day.

Of the body composition measurements, fat mass and body muscle mass were calculated on the basis of bioelectrical impedance analyses performed using an ACCUNIQ BC720 (Selvas Healthcare, Daejeon, South Korea) during physical examinations at both health screening centres. Bioelectrical impedance analysis uses a sudden decrease in voltage due to resistance and reactance (impedance) from body tissues to determine impedance and estimate the fat-free and fat masses. The ACCUNIQ BC720 device was validated using a Hologic QDR-4500W fan-beam DEXA scanner (Hologic, Bedford, MA, USA). The percentage of body fat was calculated as the body fat mass divided by total bodyweight. Measurements of other biomarkers are described in the appendix (p 2). The actual, measured value of each biomarker was used as a reference standard for training the deep-learning algorithms.

Model development

We adopted the typical neural network architecture of VGG16.[18] Retinal photographs were resized to 300×300 pixels before inputting into VGG16. To prevent overfitting, we applied various well known augmentation methods, such as random crop, flip up-down, rotation, brightness, and saturation. Because we did not use a pretrained network, the model parameters were trained from scratch. Because our datasets included retinal photographs from different ethnic groups that have different retinal pigmentation, we applied enhanced contrast techniques on retinal photographs before training and testing our algorithm. Saliency maps were generated using Guided Grad-CAM[19] from the test sets. Details of the model development and visualisation techniques are provided in the appendix (p 3).

Statistical analysis

To evaluate performance, we used the mean absolute error with a coefficient of determination (R^2) and Bland-Altman plots[20] for continuous variables (eg, age), and area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy for binary variables (eg, sex). Sensitivity and specificity were determined from the Youden index. We used a non-parametric bootstrap procedure with 2000 samples to obtain 95% CIs and reported the 2·5th and 97·5th percentiles.

Given the lack of a standard guideline for using R^2 to determine the level of predictive acceptance, we arbitrarily defined predictable biomarkers as those that could be predicted by deep-learning algorithms with an R^2 of greater than 0·15 in both the internal test set and at least one of the external test sets. Because age-related decline of body muscle mass is an important public health issue in older people, we trained and tested a separate deep-learning algorithm to predict muscle mass in this subgroup aged 65 years or older. To identify differences in performance across ethnicities, we used datasets that included more than one ethnic group (ie, SEED and the UK Biobank), because fair comparison would be possible only when the data were collected using the same protocol across multiple ethnic groups. Ethnicity in the UK Biobank was classified according to the category used by the UK Office of National Statistics: "White" (n=22 415) included British, Irish, and any other White background and "non-White" (n=2188) included Asian people (n=360), Black people (n=292), people with a mixed background (n=88), and other ethnic groups (n=1448; appendix p 8). We did not perform analysis for each of the four non-White ethnic groups, because meaningful analysis was not possible due to small sample sizes. We also performed subgroup analyses for biomarkers of age, sex, and systolic blood pressure according to the presence of retinal disease.

Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The development, validation, and external testing of our deep-learning algorithms for the prediction of systemic biomarkers included 236 257 retinal photographs obtained using six different types of retinal camera (72 890 participants) from seven data sources: two health screening centres in South Korea, the Beijing Eye Study, three cohorts in the SEED study, and the UK Biobank (table 1). Similar characteristics between the developmental set and internal test set suggest that the data were randomly well divided. The

	Severance Main Hospital		Severance Gangnam Hospital (external test set 1)	Beijing Eye Study (external test set 2)	SEED study (external test set 3)	UK Biobank (external test set 4)
	Developmental set	Internal test set				
Participants	27 516	6879	4343	1060	7726	25 366
Examinations	43 497	10 849	4662	1060	7726	25 366
Retinal photographs	86 994	21 698	9324	4234	63 275	50 732
Retinal disease*						
Referable retinal disease	2324 (5·3%)	597 (5·5%)	333 (7·7%)	NA	NA	NA
AMD	NA	NA	NA	11 (1·0%)	360 (4·7%)	362 (1·4%)
Diabetic retinopathy	NA	NA	NA	38 (3·6%)	597 (7·7%)	288 (1·1%)
Demographic factors						
Age, years†	52·92 (7·51)	53·00 (7·67)	54·51 (7·06)	58·85 (5·71)	55·12 (7·57)	55·13 (8·16)
Sex						
Female	14 941 (54·3%)	3715 (54·0%)	2680 (61·7%)	660 (62·3%)	3770 (48·8%)	11 009 (43·4%)
Male	12 575 (45·7%)	3164 (46·0%)	1663 (38·3%)	400 (37·7%)	3956 (51·2%)	14 357 (56·6%)
Body composition						
Body muscle mass, kg	45·75 (9·34)	45·61 (9·36)	46·48 (9·32)	NA	NA	NA
Height, cm	166·51 (8·37)	166·45 (8·52)	166·36 (8·68)	161·92 (7·81)	161·71 (8·97)	169·03 (9·16)
Bodyweight, kg	66·95 (12·37)	66·77 (12·32)	68·80 (13·51)	68·90 (11·64)	66·99 (13·28)	77·17 (15·46)
Percentage body fat	27·59 (6·72)	27·60 (6·72)	26·64 (6·05)	NA	NA	NA
Body-mass index, kg/m^2	24·02 (3·23)	23·96 (3·16)	24·72 (3·61)	26·24 (3·85)	25·61 (4·68)	26·93 (4·59)
Body fat mass, kg	18·49 (5·92)	18·43 (5·83)	18·47 (6·15)	NA	NA	NA
Kidney function test						
Creatinine, mg/dL	0·79 (0·23)	0·79 (0·24)	0·80 (0·18)	0·72 (0·15)	0·79 (0·37)	0·82 (0·17)
Blood pressure						
Systolic blood pressure, mm Hg	121·06 (14·17)	121·01 (14·12)	124·62 (14·81)	126·78 (18·70)	136·86 (20·67)	138·18 (19·28)
Diastolic blood pressure, mm Hg	78·74 (11·16)	78·76 (11·26)	76·06 (10·49)	70·79 (11·61)	78·70 (10·53)	81·61 (10·61)
Heart rate, beats per min	73·37 (11·18)	73·56 (11·31)	69·76 (10·71)	NA	NA	NA
Haematological parameters						
Haemoglobin, g/dL	14·51 (1·50)	14·51 (1·50)	14·71 (1·50)	NA	13·67 (1·51)	14·25 (1·22)
Haematocrit, %	42·62 (3·96)	42·60 (3·96)	43·94 (4·13)	NA	NA	41·63 (3·47)
Red blood cell count, 10^{12} per L	4·70 (0·44)	4·69 (0·44)	4·81 (0·47)	NA	4·75 (0·53)	4·53 (0·41)

Data are presented as n, n (% of participants), or mean (SD). 15 selected systemic biomarkers are shown; the other systemic biomarkers evaluated are presented in the appendix (p 12). SEED=Singapore Epidemiology of Eye Diseases. NA=data not available. AMD=age-related macular degeneration. *Referable retinal disease in health screening centres included a wide range of different retinal diseases (such as AMD, diabetic retinopathy, epiretinal membrane, or retinal vein occlusion) that require referral to an ophthalmologist for further management; number by each specific retinal disease is not available. †Participant age range was 40–69 years for all datasets except the Beijing Eye study, for which the range was 50–69 years.

Table 1: **Characteristics of the study populations**

characteristics of other biomarkers are provided in the appendix (p 12).

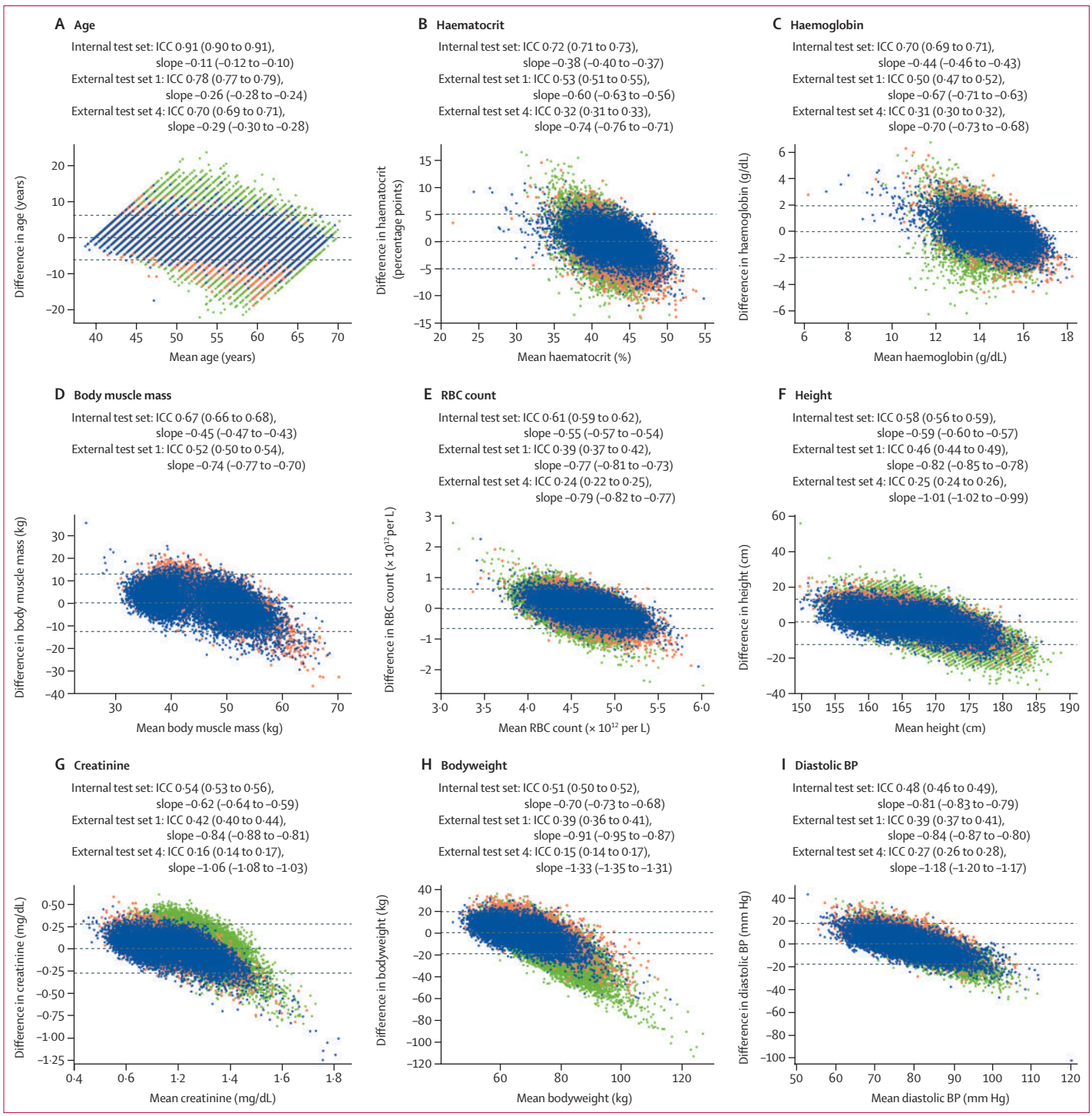
First, we trained our deep learning algorithms based on 86 994 retinal photographs (denoted as the developmental set) from a health screening centre affiliated with Severance Main Hospital, then tested these algorithms using 21 698 retinal photographs (internal test set) from the same centre. Second, we tested the algorithms using 9324 retinal photographs (external test set 1) from a health screening centre affiliated with the Severance Gangnam Hospital. Finally, we further externally tested the algorithms using 4234 retinal photographs from the Beijing Eye Study (external test set 2), 63 275 retinal photographs from the SEED study in Singapore (external test set 3), and 50 732 retinal photographs from the UK Biobank (external test set 4).

Of the 47 deep-learning algorithms we developed, 13 biomarkers showed either an R^2 of greater than 0·15 or an AUC of greater than 0·90 for predicting systemic biomarkers from retinal photographs in the internal test set, and ten biomarkers showed an R^2 of greater than 0·15 or an AUC of greater than 0·90 in at least one external test

	Severance Main Hospital (internal test set)	Severance Gangnam Hospital (external test set 1)	Beijing Eye Study (external test set 2)	SEED study (external test set 3)	UK Biobank (external test set 4)
Demographic factors					
Sex					
AUC	0·96 (0·96 to 0·96)	0·90 (0·89 to 0·91)	0·91 (0·89 to 0·93)	0·90 (0·89 to 0·91)	0·80 (0·79 to 0·80)
Accuracy	0·91 (0·90 to 0·91)	0·82 (0·81 to 0·83)	0·85 (0·82 to 0·87)	0·83 (0·82 to 0·84)	0·70 (0·69 to 0·70)
Sensitivity	0·93 (0·91 to 0·94)	0·90 (0·88 to 0·92)	0·83 (0·77 to 0·89)	0·70 (0·64 to 0·80)	0·70 (0·67 to 0·76)
Specificity	0·82 (0·80 to 0·85)	0·80 (0·79 to 0·81)	0·89 (0·84 to 0·93)	0·84 (0·83 to 0·86)	0·72 (0·71 to 0·73)
Age, years					
MAE	2·43 (2·39 to 2·47)	3·38 (3·30 to 3·46)	3·78 (3·63 to 3·93)	3·77 (3·71 to 3·82)	4·50 (4·46 to 4·55)
R^2	0·83 (0·82 to 0·84)	0·61 (0·59 to 0·64)	0·36 (0·31 to 0·41)	0·63 (0·62 to 0·64)	0·51 (0·50 to 0·52)
Body composition					
Body muscle mass, kg					
MAE	5·11 (5·04 to 5·19)	6·09 (5·96 to 6·23)	NA	NA	NA
R^2	0·52 (0·51 to 0·53)	0·33 (0·30 to 0·35)	NA	NA	NA
Height, cm					
MAE	5·20 (5·13 to 5·28)	5·93 (5·80 to 6·05)	5·48 (5·23 to 5·73)	6·21 (6·10 to 6·31)	7·09 (7·02 to 7·15)
R^2	0·42 (0·40 to 0·43)	0·28 (0·25 to 0·30)	0·23 (0·18 to 0·27)	0·25 (0·24 to 0·27)	0·08 (0·06 to 0·09)
Bodyweight, kg					
MAE	7·69 (7·57 to 7·81)	9·63 (9·42 to 9·84)	8·28 (7·89 to 8·68)	9·69 (9·51 to 9·86)	11·81 (11·69 to 11·92)
R^2	0·36 (0·34 to 0·37)	0·19 (0·16 to 0·22)	0·17 (0·11 to 0·22)	0·11 (0·10 to 0·13)	0·04 (0·03 to 0·05)
Percentage body fat*					
MAE	4·71 (4·64 to 4·78)	4·50 (4·39 to 4·60)	NA	NA	NA
R^2	0·23 (0·21 to 0·24)	0·08 (0·05 to 0·12)	NA	NA	NA
Body-mass index*, kg/m²					
MAE	2·15 (2·12 to 2·19)	2·37 (2·31 to 2·42)	2·90 (2·76 to 3·04)	3·52 (3·45 to 3·58)	3·47 (3·44 to 3·51)
R^2	0·17 (0·16 to 0·18)	0·14 (0·12 to 0·16)	0·06 (0·02 to 0·10)	0·04 (0·03 to 0·05)	0·01 (0·002 to 0·02)
Kidney function test					
Creatinine, mg/dL					
MAE	0·11 (0·11 to 0·11)	0·12 (0·12 to 0·12)	0·11 (0·10 to 0·11)	0·17 (0·16 to 0·18)	0·15 (0·15 to 0·16)
R^2	0·38 (0·37 to 0·40)	0·26 (0·24 to 0·28)	0·12 (0·06 to 0·18)	0·06 (0·04 to 0·09)	0·01 (0·001 to 0·02)
Blood pressure					
Diastolic blood pressure, mm Hg					
MAE	7·20 (7·09 to 7·30)	7·59 (7·26 to 7·91)	8·09 (7·72 to 8·47)	7·14 (7·02 to 7·26)	7·67 (7·59 to 7·74)
R^2	0·35 (0·33 to 0·36)	0·21 (0·18 to 0·24)	0·23 (0·17 to 0·28)	0·27 (0·25 to 0·29)	0·16 (0·15 to 0·17)
Systolic blood pressure, mm Hg					
MAE	9·29 (9·16 to 9·43)	10·55 (10·31 to 10·79)	13·20 (12·58 to 13·83)	13·95 (13·69 to 14·22)	13·57 (13·44 to 13·70)
R^2	0·31 (0·29 to 0·32)	0·17 (0·15 to 0·20)	0·19 (0·15 to 0·24)	0·21 (0·19 to 0·22)	0·20 (0·19 to 0·21)
Haematological parameters					
Haematocrit, %					
MAE	2·03 (2·00 to 2·06)	2·81 (2·75 to 2·88)	NA	NA	2·62 (2·59 to 2·64)
R^2	0·57 (0·56 to 0·59)	0·26 (0·23 to 0·30)	NA	NA	0·09 (0·08 to 0·11)
Haemoglobin, g/dL					
MAE	0·79 (0·78 to 0·80)	0·96 (0·94 to 0·98)	NA	0·98 (0·96 to 1·00)	0·93 (0·92 to 0·94)
R^2	0·56 (0·55 to 0·57)	0·33 (0·30 to 0·36)	NA	0·32 (0·29 to 0·35)	0·06 (0·04 to 0·08)
Red blood cell count*, 10^{12} per L					
MAE	0·26 (0·25 to 0·26)	0·35 (0·34 to 0·35)	NA	0·37 (0·36 to 0·38)	0·33 (0·32 to 0·33)
R^2	0·45 (0·44 to 0·47)	0·14 (0·10 to 0·17)	NA	0·14 (0·11 to 0·17)	–0·02 (–0·04 to –0·01)

Data in parentheses are 95% CIs. SEED=Singapore Epidemiology of Eye Diseases. AUC=area under the receiver operating characteristics curve. MAE=mean absolute error. NA=data not available. *Percentage body fat, body-mass index, and red blood cell count are not predictable biomarkers based on our criteria.

Table 2: **Biomarkers with the highest performance from retinal photographs via deep learning**



(Figure continues on next page)

set (table 2). The algorithms clearly predicted sex (AUC 0·96 [95% CI 0·96–0·96]; accuracy 0·91 [0·90–0·91]) and age (mean absolute error 2·43 years [2·39–2·47]; R^2=0·83 [0·82–0·84]) in the internal test set. The algorithm that predicted sex performed well in the four external test sets (AUC 0·80–0·91). The age prediction algorithm yielded R^2 values of 0·36 to 0·63 in the four external test sets.

Regarding the body composition biomarkers, our deep-learning algorithms achieved an R^2 of 0·52 (95% CI

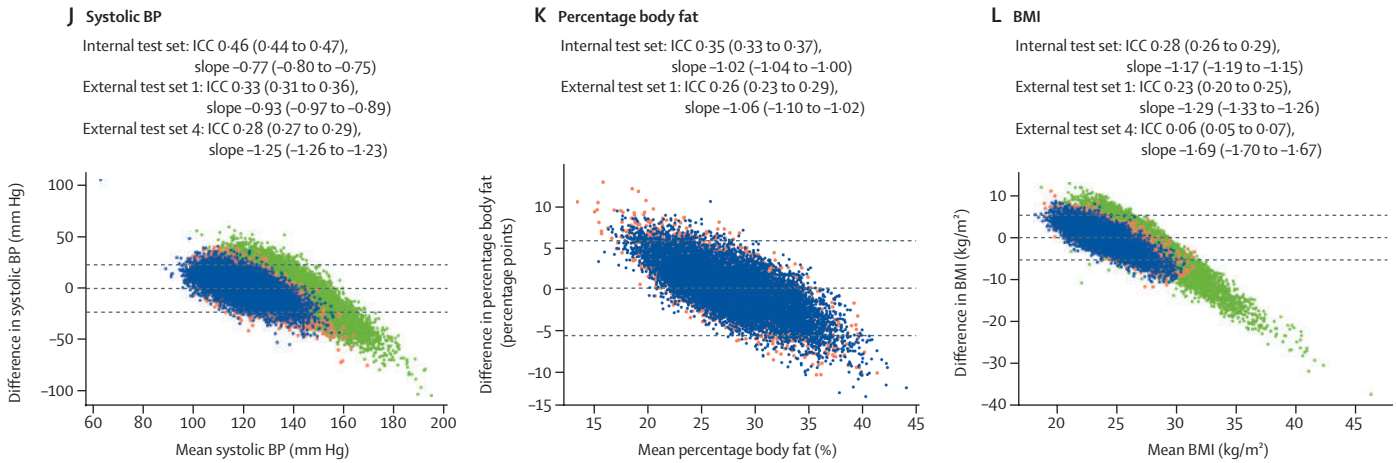


Figure: **Bland-Altman plots for predicting systemic biomarkers**
The x-axes represent the mean of the measured and predicted values, and the y-axes represent the difference between the measured and predicted values. Each dot represents one examination in the internal test set (Severance Main Hospital; blue dots), the external test set 1 (Severance Gangnam Hospital; orange dots), or the external test set 4 (UK Biobank; green dots). The dashed lines represent the mean of the difference with 95% limits of agreement (1·96 × SD). ICC and slope of linear fit are presented with 95% CIs. BMI=body-mass index. BP=blood pressure. ICC=intraclass correlation coefficient. RBC=red blood cell.

0·51–0·53) for predicting body muscle mass, 0·42 (0·40–0·43) for height, and 0·36 (0·34–0·37) for bodyweight in the internal test set. The body muscle mass was also predicted in external test set 1, with an R^2 of 0·33 (0·30–0·35). In subgroup analysis of older people (aged ≥65 years; appendix p 4), the results showed reasonable performance, with a mean absolute error of 5·90 kg (5·58–6·21) and an R^2 of 0·22 (0·15–0·28) in external test set 1. Across external test sets 1–3, the deep-learning algorithms for prediction of height and bodyweight achieved R^2 values of between 0·23 and 0·28 for height and 0·11 and 0·19 for bodyweight. For both height and bodyweight predictions, the validation performance was lowest in external test set 4 (R^2≤0·08). Our algorithm could predict body-mass index (BMI), with R^2 values of 0·17 (0·16–0·18) in the internal test set and 0·14 (0·12–0·16), 0·06 (0·02–0·10), 0·04 (0·03–0·05), and 0·01 (0·002–0·02) in external test sets 1–4, respectively. The percentage body fat was predicted with an R^2 of 0·23 (0·21–0·24) in the internal test set, but the performance was poor, with an R^2 of 0·08 (0·05–0·12) in external test set 1. Percentage body fat was not available for the other datasets.

Our algorithm for predicting serum creatinine concentration achieved an R^2 of 0·38 (95% CI 0·37 to 0·40) in the internal test set, 0·26 (0·24 to 0·28) in external test set 1, and 0·12 (0·06 to 0·18) in external test set 2. However, the predictive performance was relatively poor, with an R^2 of 0·06 (0·04 to 0·09) in external test set 3, and 0·01 (0·001 to 0·02) in external test set 4. Our algorithm for predicting blood pressure yielded an R^2 of 0·35 (0·33 to 0·36) for diastolic blood pressure and 0·31 (0·29 to 0·32) for systolic blood pressure in the internal test set. The predictions of blood pressure yielded R^2 values of 0·16–0·27 for diastolic blood pressure and 0·17–0·21 for systolic blood pressure in all external test sets. Our deep-learning algorithms for haematological parameters could predict the haematocrit level (R^2=0·57 [0·56 to 0·59]), haemoglobin concentration (R^2=0·56 [0·55 to 0·57]), and red blood cell count (R^2=0·45 [0·44 to 0·47]) in the internal test set. However, in external test set 4, R^2 values were 0·09 (0·08 to 0·11) for haematocrit, 0·06 (0·04 to 0·08) for haemoglobin, and –0·02 (–0·04 to –0·01) for red blood cell count.

In general, the algorithms performed poorly in external test set 4. Specifically, the deep-learning algorithms for prediction of height, bodyweight, BMI, creatinine, haematocrit, haemoglobin, and red blood cell counts showed limited generalisability in external test 4 (R^2≤0·09). As shown in the Bland-Altman plots (figure), proportional bias was observed, indicating that the predicted values in the lower range were overestimated and those in the higher range were underestimated.

Other biomarkers that could not be well predicted from retinal photographs via deep learning (R^2 <0·15 in the internal test set and external sets) are described in the appendix (p 13). Our deep-learning algorithms for six biomarkers yielded an R^2 of 0·10–0·15 in the internal test set and of 0·02–0·09 in external test set 1: lipid profile (HDL cholesterol and triglyceride), liver function test (γ-glutamyl transferase and alanine aminotransferase), and diabetes blood test (glycated haemoglobin A_{1c} [HbA_{1c}] and fasting glucose). 28 other biomarkers, such as C-reactive protein and free thyroxine, could not be predicted from retinal photographs via deep learning.

We applied our algorithms to each ethnic group in the SEED study (comprising Chinese, Indian, and Malay populations) and the UK Biobank (comprising White and non-White populations; table 3). In SEED, sex, age,

	SEED study			UK Biobank	
	Chinese (n=2598)	Indian (n=2752)	Malay (n=2376)	White (n=22 415)	Non-White (n=2188)
Demographic factors					
Sex					
AUC	0·95 (0·94 to 0·96)	0·91 (0·90 to 0·92)	0·93 (0·92 to 0·94)	0·78 (0·78 to 0·79)	0·86 (0·84 to 0·89)
Accuracy	0·88 (0·86 to 0·89)	0·76 (0·74 to 0·78)	0·84 (0·83 to 0·86)	0·72 (0·71 to 0·72)	0·76 (0·73 to 0·79)
Sensitivity	0·89 (0·84 to 0·94)	0·84 (0·75 to 0·90)	0·86 (0·83 to 0·89)	0·70 (0·66 to 0·76)	0·85 (0·77 to 0·91)
Specificity	0·88 (0·84 to 0·92)	0·84 (0·80 to 0·89)	0·86 (0·82 to 0·90)	0·92 (0·91 to 0·93)	0·76 (0·71 to 0·81)
Age, years					
MAE	3·71 (3·61 to 3·81)	3·88 (3·78 to 3·98)	3·69 (3·58 to 3·79)	4·52 (4·48 to 4·57)	3·45 (3·28 to 3·62)
R^2	0·57 (0·55 to 0·60)	0·60 (0·57 to 0·62)	0·70 (0·68 to 0·72)	0·51 (0·49 to 0·52)	0·65 (0·61 to 0·70)
Body composition					
Height, cm					
MAE	5·61 (5·45 to 5·77)	6·65 (6·47 to 6·83)	7·65 (7·43 to 7·86)	7·14 (7·07 to 7·20)	6·61 (6·31 to 6·91)
R^2	0·30 (0·27 to 0·32)	0·19 (0·16 to 0·22)	−0·09 (−0·16 to −0·02)	0·06 (0·05 to 0·07)	0·18 (0·13 to 0·22)
Body weight, kg					
MAE	8·41 (8·16 to 8·67)	10·45 (10·10 to 10·80)	9·96 (9·64 to 10·28)	11·87 (11·75 to 11·99)	11·49 (10·95 to 12·04)
R^2	0·19 (0·16 to 0·22)	−0·06 (−0·10 to −0·02)	0·10 (0·06 to 0·13)	0·04 (0·03 to 0·05)	0·06 (0·02 to 0·10)
Body-mass index, kg/m^2					
MAE	2·69 (2·61 to 2·78)	3·91 (3·78 to 4·05)	4·19 (4·04 to 4·34)	3·41 (3·37 to 3·45)	3·42 (3·24 to 3·60)
R^2	0·07 (0·04 to 0·09)	−0·28 (−0·32 to −0·23)	−0·23 (−0·28 to −0·18)	0·01 (0·003 to 0·01)	0·01 (−0·03 to 0·04)
Kidney function test					
Creatinine, mg/dL					
MAE	15·30 (14·17 to 16·42)	15·87 (14·81 to 16·93)	19·42 (17·87 to 20·96)	0·13 (0·13 to 0·13)	0·15 (0·13 to 0·16)
R^2	0·08 (0·05 to 0·15)	0·05 (0·02 to 0·09)	−0·03 (−0·05 to −0·01)	−0·001 (−0·01 to 0·01)	−0·01 (−0·05 to 0·04)
Blood pressure					
Diastolic blood pressure, mm Hg					
MAE	6·72 (6·52 to 6·91)	7·17 (6·97 to 7·38)	7·56 (7·32 to 7·80)	7·69 (7·61 to 7·77)	7·25 (6·91 to 7·59)
R^2	0·28 (0·25 to 0·32)	0·23 (0·19 to 0·26)	0·27 (0·24 to 0·30)	0·16 (0·15 to 0·17)	0·26 (0·21 to 0·31)
Systolic blood pressure, mm Hg					
MAE	12·20 (11·82 to 12·58)	13·22 (12·81 to 13·63)	16·72 (16·14 to 17·30)	13·65 (13·51 to 13·79)	12·52 (11·89 to 13·14)
R^2	0·25 (0·23 to 0·28)	0·21 (0·19 to 0·23)	0·06 (0·03 to 0·10)	0·20 (0·19 to 0·20)	0·27 (0·23 to 0·31)
Haematological parameters					
Haemoglobin, g/dL					
MAE	1·06 (1·05 to 1·07)	1·05 (1·04 to 1·06)	0·97 (0·96 to 0·98)	0·92 (0·91 to 0·93)	1·07 (1·02 to 1·12)
R^2	0·10 (0·07 to 0·12)	0·29 (0·28 to 0·30)	0·32 (0·31 to 0·34)	0·07 (0·05 to 0·08)	−0·01 (−0·16 to 0·13)
Red blood cell count, 10^{12}/L					
MAE	0·35 (0·34 to 0·35)	0·41 (0·40 to 0·41)	0·42 (0·41 to 0·42)	0·32 (0·32 to 0·33)	0·38 (0·36 to 0·40)
R^2	0·19 (0·17 to 0·20)	−0·09 (−0·10 to −0·07)	−0·05 (−0·07 to −0·03)	−0·03 (−0·05 to −0·02)	0·07 (0·01 to 0·14)

Data in parentheses are 95% CIs. SEED=Singapore Epidemiology of Eye Diseases. AUC=area under the receiver operating characteristics curve. MAE=mean absolute error.

Table 3: **Performance by ethnic group in the SEED study and UK Biobank test sets**

and diastolic blood pressure were well predicted across the three ethnic groups. However, for biomarkers such as sex, height, bodyweight, systolic blood pressure, and red blood cell count, the algorithms' performances were better in the Chinese population compared with the Malay or Indian populations. In the UK Biobank, the algorithms' performances for bodyweight, BMI, creatinine, and haematological parameters were similarly poor in White and non-White groups, whereas those for sex, age, height, and diastolic and systolic blood pressure were better in the non-White group compared with the White group.

In subgroup analyses, the presence of retinal disease was not found to substantially influence the algorithms' performance (appendix p 5).

We used saliency maps to identify the area from which the deep-learning algorithms might have predicted systemic biomarkers (appendix pp 9–11). Models trained to predict age and sex primarily highlighted the optic disc and retinal vessels. Models that predicted haematocrit, haemoglobin, systolic blood pressure, and diastolic blood pressure used attention masks focused on the features of retinal vessels. The optic disc area and features were mainly used to predict measures of body composition

(eg, BMI, percentage body fat, and bodyweight). No specific features were identified from the saliency maps used to predict other biomarkers.

Discussion

In this study, we developed deep-learning algorithms to predict 47 systemic biomarkers from retinal photographs. More than 230 000 retinal photographs taken using diverse camera types from approximately 72 000 participants in seven diverse Asian and European cohorts were used. One key finding was the quantification of a cluster of body composition measures (muscle mass, height, and bodyweight), and creatinine from retinal photographs. Our results also confirmed previous studies that used deep learning to predict age, sex, systolic blood pressure, diastolic blood pressure, BMI, haematocrit, serum haemoglobin, and red blood cell counts.[5,6] Nevertheless, the performance of deep-learning algorithms for systemic biomarkers was lower in the external test sets.

Body composition measures, such as muscle mass, have recently gained attention as more reliable biomarkers than BMI for nutritional status as well as for cardiometabolic risk.[21] Sarcopenia is defined as an age-related process of skeletal muscle loss, which can be assessed via changes in body muscle mass. International clinical practice guidelines recommend that adults aged 65 years or older should be screened annually for sarcopenia.[22] Although sarcopenia is an important concern in ageing populations, the lack of suitable tools limits large-scale population screening. In this study, we used bioelectrical impedance analysis as the reference standard, and our deep-learning algorithm could quantify body muscle mass from retinal photographs with a mean absolute error of 6·09 kg (R^2=0·33) in external set 1. Our subgroup analysis in older adults (aged ≥65 years) showed reasonable performance, with a mean absolute error of 5·90 kg (R^2=0·22) in external test set 1. Therefore, further studies are needed to determine whether retinal photography might be a good adjunctive screening tool for sarcopenia.

Chronic kidney disease has been labelled a silent killer by the American Society of Nephrology, which recommends routine kidney function screening in the general population.[23] In a study using retinal photographs,[24] a deep-learning algorithm was able to predict chronic kidney disease (AUC 0·73) with modest generalisability. Chronic kidney disease was defined as a binary outcome on the basis of an estimated glomerular filtration rate (eGFR) of less than 60 units. eGFR was estimated using the Chronic Kidney Disease Epidemiology Collaboration equation, using information on creatinine level, age, sex, and bodyweight.[24] Deep-learning algorithms can predict age and sex with high accuracy, which could explain why deep learning based on eGFR can predict chronic kidney disease with superior performance compared with deep learning based on serum creatinine concentration, as in our study.

Our study supported the previous findings by Poplin and colleagues,[5] who used retinal photographs from the UK Biobank and EyePACs datasets and developed deep-learning algorithms to predict age, sex, HbA_{1c}, systolic blood pressure, diastolic blood pressure, and BMI, and those by Mitani and colleagues,[6] who reported algorithms to predict serum haemoglobin concentrations, haematocrit, and red blood cell count using retinal photographs from the UK Biobank. Compared with these studies, our algorithms generally yielded higher R^2 values in predicting age, HbA_{1c}, diastolic blood pressure, BMI, haemoglobin, haematocrit, and red blood cell count (appendix p 14). In saliency maps shown by Poplin and colleagues,[5] the blood vessels and optic disc were highlighted for age, sex, and systolic blood pressure prediction. Similarly, our saliency maps (internal and external test sets) highlighted retinal vessels and the optic disc as the main features. These observations are biologically plausible given that previous studies have shown that age, sex, and blood pressure are important determinants of retinal vessel calibre.[25]

Previous studies provide little information on the performance of deep-learning algorithms across multiple health-care settings. In the present study, using multiple datasets collected from different sites and settings, only three of the 47 trained deep-learning algorithms (those for age, sex, and diastolic and systolic blood pressure) showed a consistently good performance across all external sets, indicating that generalisation was only applicable to selective biomarkers. Our study also provides information about clusters of systemic biomarkers that could not be predicted from retinal photographs, including those related to thyroid function, most biochemical measures, haematological parameters other than haematocrit and haemoglobin, and C-reactive protein. Retinal microvascular abnormalities reflect cumulative microcirculatory damage from hypertension, ageing, and other processes.[26] Therefore, retinal changes could be more useful for predicting biomarkers related to cardiovascular and chronic disease, compared with other time-sensitive biomarkers.

Our subgroup analysis showed that the predictive performances for systemic biomarkers differed by ethnicity. Although the reason for this ethnic difference is unclear, a possible explanation is different basic profiles of biomarkers across ethnicities. Given that our algorithms were trained on Asian data, the poorer performances in body composition measures observed in the UK Biobank dataset could be due to the body composition profile of the Asian population being generally different to that of the European population.[27] BMI also differs between the three Asian ethnic groups in the SEED study,[28] which partially explains the lower performance for BMI prediction in the Malay and Indian populations. Given the apparent impact of ethnicity on the performance of deep-learning algorithms based on retinal photographs, we recommend that similar studies in the future also present findings by ethnicity.

Retinal photography is non-invasive and increasingly available in primary-care settings and screening programmes (eg, diabetic retinopathy screening[29]). Our algorithms could potentially be used as add-on tests to identify individuals who require referral for confirmatory tests (eg, dual-energy x-ray absorptiometry for sarcopenia, venous blood puncture for serum creatinine or haemoglobin). Because the performance of age and sex prediction is particularly good, sophisticated combination of these deep-learning algorithms could eventually be used to predict cardiovascular risk.[5] However, the current evaluation does not support firm conclusions regarding the value of deep learning for disease screening in new populations.

Our study had several limitations. First, because images with artifacts or of poor quality were excluded, our results were unlikely to be confounded by image quality issues, but the performance of our algorithm needs to be further evaluated in real-world datasets, which could contain images of varying quality. Second, the biomarker of body muscle mass was only available in the two Korean datasets. Therefore, further replication is needed to confirm it as a predictable biomarker and its external validity. Third, although retinal photography and all other tests were done on the same visit day, some biomarker values (eg, systolic blood pressure, bodyweight) could still fluctuate over time. Although this effect cannot be quantified, these fluctuations could cause information bias. However, the relatively large scale of our data might neutralise this bias. Lastly, we acknowledge that our deep-learning algorithms might have been overtrained because they did not perform well in the external test sets. Additionally, the fundamental differences among the internal and external test sets (eg, different fundus cameras, brightness and fields of photographs, and ethnicities) could have resulted in poorer performance in the external test sets.

In conclusion, we developed deep-learning algorithms to predict systemic biomarkers from retinal photographs. In addition to previously reported biomarkers (age, sex, blood pressure, and haematological parameters), we identified novel predictable biomarkers, including body composition measurements (height, bodyweight, and body muscle mass) and kidney function (creatinine). However, our findings show that optimal prediction in external test sets was applicable to only some biomarkers, and we found that the retina provides little information about many other systemic biomarkers. Given the challenge of generalisation, further evaluation of clinical utility is needed in future studies.

For the **full list of gradable photographs and code** see https://github.com/medi-whale/UKBIOBANK_FUNDUS_Classifier

For the **prediction models and tutorial** see https://github.com/medi-whale/systemic_biomarkers_via_deeplearning

Contributors

THR, GL, SSK, TYW, and C-YC conceptualised the study. THR, GL, YK, CJL, SJB, YAK, MD, TYW, and C-YC reviewed the literature. THR, GL, YK, SSK, TYW, and C-YC designed the study. THR, GL, YK, Y-CT, YAK, SJB, MY, BKL, HCK, YXW, JBJ, SSK, TYW, and C-YC collected the data. GL and YK developed the algorithm. THR, GL, and YK analysed the data. THR, GL, Y-CT, HCK, CS, DSWT, YXW, JBJ, SSK, TYW, and C-YC drafted the manuscript. BKL, SP, HCK, CS, DSWT, YXW, JBJ, SSK, TYW, and C-YC did the critical revision. All authors contributed to data interpretation and read and approved the final manuscript.

Declaration of interests

THR was a scientific adviser to the start-up company Medi Whale. He received stock as a part of the standard compensation package. DSWT and TYW hold a patent on a deep-learning system for the detection of retinal diseases and this patent is not directly related to this study. DSWT is a co-founder of EyRiS. TYW has received consulting fees from Allergan, Bayer, Boehringer-Ingelheim, Genentech, Merck, Novartis, Oxurion, Roche, and Samsung Bioepis. TYW is a co-founder of Plano and EyRiS. Potential conflicts of interest are managed according to the institutional policies of the Singapore Health System (SingHealth) and the National University of Singapore. GL and YK are employees of Medi Whale and GL owns stock in Medi Whale. THR and GL hold the following patents that may be affected by this study: 10–2018–0166720(KR), 10–2018–0166721(KR), 10–2018–0166722(KR), and PCT/KR2018/016388, cardiovascular disease diagnostic assistance method and apparatus; 10–2018–0157559(KR), 10–2018–0157560(KR), and 10–2018–0157561(KR), diagnostic assistance system; 62/694,901(US) and 62/776,345(US), diagnostic technology using artificial intelligence; 62/715,729(US), method for controlling the portable fundus camera and diagnosing disease using the portable fundus camera; and 10–2017–0175865(KR), method for predicting cardio-cerebrovascular disease using an eye image. All other authors declare no competing interests.

Data sharing

The UK Biobank test dataset was obtained from the UK Biobank (application number 45925) and a full list of the identification numbers of gradable photographs and code are available online. Data cannot be shared publicly due to the violation of patient privacy and the absence of informed consent for data sharing. Data are available to researchers who meet the criteria for access to confidential data; requests should be made to Sung Soo Kim, Department of Ophthalmology, Yonsei University (semekim@yuhs.ac). The prediction models with Python implementation and a detailed tutorial are available online.

Acknowledgments

This work was supported by grants from the Agency for Science, Technology, and Research (grant number A19D1b0095) and the National Medical Research Council, Singapore (grant numbers NMRC/CIRG/1417/2015, NMRC/CIRG/1488/2018), and by the Ministry of Trade, Industry and Energy, South Korea, and Korea Institute for Advancement of Technology through the International Cooperative Research & Development programme (project number P0011929).

References

1 Wong TY, Klein R, Couper DJ, et al. Retinal microvascular abnormalities and incident stroke: the Atherosclerosis Risk in Communities Study. *Lancet* 2001; **358:** 1134–40.

2 Rim TH, Teo AWJ, Yang HHS, Cheung CY, Wong TY. Retinal vascular signs and cerebrovascular diseases. *J Neuroophthalmol* 2020; **40:** 44–59.

3 Wagner SK, Fu DJ, Faes L, et al. Insights into systemic disease through retinal imaging-based oculomics. *Transl Vis Sci Technol* 2020; **9:** 6.

4 Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019; **103:** 167–75.

5 Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018; **2:** 158–64.

6 Mitani A, Huang A, Venugopalan S, et al. Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng* 2020; **4:** 18–27.

7 Kim YD, Noh KJ, Byun SJ, et al. Effects of hypertension, diabetes, and smoking on age and sex prediction from retinal fundus images. *Sci Rep* 2020; **10:** 4623.

8 Chang J, Ko A, Park SM, et al. Association of cardiovascular mortality and deep learning-funduscopic atherosclerosis score derived from retinal fundus images. *Am J Ophthalmol* 2020; published online March 25. https://doi.org/S0002-9394(20)30127-6.

9 Yamashita T, Asaoka R, Terasaki H, et al. Factors in color fundus photographs that can be used by humans to determine sex of individuals. *Transl Vis Sci Technol* 2020; **9:** 4.

10 Vaghefi E, Yang S, Hill S, Humphrey G, Walker N, Squirrell D. Detection of smoking status from retinal images; a Convolutional Neural Network study. *Sci Rep* 2019; **9:** 7180.

11 Zhang L, Yuan M, An Z, et al. Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: a cross-sectional study of chronic diseases in central China. *PLoS One* 2020; **15:** e0233166.

12 Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; **17:** 195.

13 Jonas JB, Xu L, Wang YX. The Beijing eye study. *Acta Ophthalmol* 2009; **87:** 247–61.

14 Foong AW, Saw SM, Loo JL, et al. Rationale and methodology for a population-based study of eye diseases in Malay people: the Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol* 2007; **14:** 25–35.

15 Lavanya R, Jeganathan VS, Zheng Y, 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* 2009; **16:** 325–36.

16 Elliott P, Peakman TC, UK Biobank. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008; **37:** 234–44.

17 Rim TH, Soh ZD, Tham YC, et al. Deep learning for automated sorting of retinal photographs. *Ophthalmol Retina* 2020; **4:** 793–800.

18 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2015; published online April 10. https://arxiv.org/abs/1409.1556 (preprint).

19 Xu K, Ba J, Kiros R, et al. Show, attend and tell: neural image caption generation with visual attention. *arXiv* 2016; published online April 19. https://arxiv.org/abs/1502.03044 (preprint).

20 Yu M, Tham Y-C, Rim TH, Ting DS, Wong TY, Cheng C-Y. Reporting on deep learning algorithms in health care. *Lancet Digital Health* 2019; **1:** e328–29.

21 Atkins JL. Effects of sarcopenic obesity on cardiovascular disease and all-cause mortality. In: Walrand S, ed. Nutrition and skeletal muscle. Cambridge, MA: Elsevier, 2019: 93–103.

22 Dent E, Morley JE, Cruz-Jentoft AJ, et al. International clinical practice guidelines for sarcopenia (ICFSR): screening, diagnosis and management. *J Nutr Health Aging* 2018; **22:** 1148–61.

23 Lewington AJ, Cerdá J, Mehta RL. Raising awareness of acute kidney injury: a global perspective of a silent killer. *Kidney Int* 2013; **84:** 457–67.

24 Sabanayagam C, Xu D, Ting DS, et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health* 2020; **2:** e295–302.

25 Cheung CY, Zheng Y, Hsu W, et al. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology* 2011; **118:** 812–18.

26 Wong TY, Klein R, Klein BE, Tielsch JM, Hubbard L, Nieto FJ. Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality. *Surv Ophthalmol* 2001; **46:** 59–80.

27 James WP, Chunming C, Inoue S. Appropriate Asian body mass indices? *Obes Rev* 2002; **3:** 139.

28 Wong TY, Tham YC, Sabanayagam C, Cheng CY. Patterns and risk factor profiles of visual loss in a multiethnic Asian population: the Singapore Epidemiology of Eye Diseases study. *Am J Ophthalmol* 2019; **206:** 48–73.

29 Wong TY, Sabanayagam C. Strategies to tackle the global burden of diabetic retinopathy: from epidemiology to artificial intelligence. *Ophthalmologica* 2020; **243:** 9–20.