Non-invasive biopsy diagnosis of diabetic kidney disease via deep learning applied to retinal images: a population-based study

Ziyao Meng, Zhouyu Guan*, Shujie Yu*, Yilan Wu*, Yaoning Zhao*, Jie Shen*, Cynthia Ciwei Lim, Tingli Chen, Dawei Yang, An Ran Ran, Feng He, Haslina Hamzah, Sarkaaj Singh, Anis Syazwani Abd Raof, Jian Wen Samuel Lee-Boey, Soo-Kun Lim, Xufang Sun, Shuwang Ge, Gang Xu, Hua Su, Yang Cheng, Feng Lu, Xiaofei Liao, Hai Jin, Chenxin Deng, Lei Ruan, Cuntai Zhang, Chan Wu, Rongping Dai, Yixiao Jin, Wenxiao Wang, Tingyao Li, Ruhan Liu, Jiajia Li, Jia Shu, Yuwei Lu, Xiangning Wang, Qiang Wu, Yiming Qin, Jin Tang, Xiaohua Sheng, Qiong Jiao, Xiaokang Yang, Minyi Guo, Gareth J McKay, Ruth E Hogg, Gerald Liew, Evelyn Yi Lyn Chee, Wynne Hsu, Mong Li Lee, Simon Szeto, Andrea O Y Luk, Juliana C N Chan, Carol Y Cheung*, Gavin Siew Wei Tan*, Yih-Chung Tham*, Ching-Yu Cheng*, Charumathi Sabanayagam*†, Lee-Ling Lim*†, Weiping Jia*†, Huating Li†, Bin Sheng†, Tien Yin Wong†*

Summary

Background Improving the accessibility of screening diabetic kidney disease (DKD) and differentiating isolated diabetic nephropathy from non-diabetic kidney disease (NDKD) are two major challenges in the field of diabetes care. We aimed to develop and validate an artificial intelligence (AI) deep learning system to detect DKD and isolated diabetic nephropathy from retinal fundus images.

Methods In this population-based study, we developed a retinal image-based AI-deep learning system, DeepDKD, pretrained using 734 084 retinal fundus images. First, for DKD detection, we used 486 312 retinal images from 121 578 participants in the Shanghai Integrated Diabetes Prevention and Care System for development and internal validation, and ten multi-ethnic datasets from China, Singapore, Malaysia, Australia, and the UK (65 406 participants) for external validation. Second, to differentiate isolated diabetic nephropathy from NDKD, we used 1068 retinal images from 267 participants for development and internal validation, and three multi-ethnic datasets from China, Malaysia, and the UK (244 participants) for external validation. Finally, we conducted two proof-of-concept studies: a prospective real-world study with 3 months' follow-up to evaluate the effectiveness of DeepDKD in screening DKD; and a longitudinal analysis of the effectiveness of DeepDKD in differentiating isolated diabetic nephropathy from NDKD on renal function changes with 4·6 years' follow-up.

Findings For detecting DKD, DeepDKD achieved an area under the receiver operating characteristic curve (AUC) of 0·842 (95% CI 0·838–0·846) on the internal validation dataset and AUCs of 0·791–0·826 across external validation datasets. For differentiating isolated diabetic nephropathy from NDKD, DeepDKD achieved an AUC of 0·906 (0·825–0·966) on the internal validation dataset and AUCs of 0·733–0·844 across external validation datasets. In the prospective study, compared with the metadata model, DeepDKD could detect DKD with higher sensitivity (89·8% *vs* 66·3%, p<0·0001). In the longitudinal study, participants with isolated diabetic nephropathy and participants with NDKD identified by DeepDKD had a significant difference in renal function outcomes (proportion of estimated glomerular filtration rate decline: 27·45% *vs* 52·56%, p=0·0010).

Interpretation Among diverse multi-ethnic populations with diabetes, a retinal image-based AI-deep learning system showed its potential for detecting DKD and differentiating isolated diabetic nephropathy from NDKD in clinical practice.

Funding National Key R & D Program of China, National Natural Science Foundation of China, Beijing Natural Science Foundation, Shanghai Municipal Key Clinical Specialty, Shanghai Research Centre for Endocrine and Metabolic Diseases, Innovative research team of high-level local universities in Shanghai, Noncommunicable Chronic Diseases-National Science and Technology Major Project, Clinical Special Program of Shanghai Municipal Health Commission, and the three-year action plan to strengthen the construction of public health system in Shanghai.

Introduction

Diabetic kidney disease (DKD) occurs in around 40% of patients with diabetes[1,2] and is associated with significantly increased morbidity and mortality.[3,4] Early detection and intervention of DKD via regular screening is a well established clinical pathway.[5] Currently, screening is based on collecting blood and urine samples, by measuring estimated glomerular filtration rate (eGFR) from blood and

Lancet Digit Health 2025; 7: 100868

Published Online April 30, 2025 https://doi.org/10.1016/j.landig.2025.02.008

*These authors contributed equally (co-first authors). †These authors jointly supervised this work (co-last authors).

Shanghai Belt and Road International Joint Laboratory of Intelligent Prevention and Treatment for Metabolic Diseases, Department of Computer Science and Engineering, School of Electronic, Information, and Electrical Engineering, Institute for Proactive Healthcare, Shanghai Jiao Tong University, Department of Endocrinology and Metabolism, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai Diabetes Institute, Shanghai Clinical Centre for Diabetes, Shanghai Key Laboratory of Diabetes Mellitus, Shanghai, China (Z Meng MEng, Z Guan MD, S Yu MBBS, T Li BSc, R Liu PhD, J Li PhD, J Shu BEng, Y Lu MBBS, Y Qin BEng, Prof M Guo PhD, Prof W Jia MD PhD, Prof H Li MD PhD, Prof B Sheng PhD); **MOE Key Laboratory of AI, School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China** (Z Meng MEng, T Li BSc, R Liu PhD, J Li PhD, J Shu BEng, Y Qin BEng, Prof X Yang PhD, Prof M Guo PhD, Prof B Sheng PhD); **Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital Eye Center, School of Clinical Medicine, Tsinghua Medicine,**

Tsinghua University, Beijing, China (Y Wu BSc, Y Zhao BSc, Y Jin MSc, Prof T Y Wong MD PhD); **Medical Records and Statistics Office, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China** (J Shen PhD); **Department of Renal Medicine, Singapore General Hospital, SingHealth-Duke Academic Medical Centre, Singapore, Singapore** (C C Lim MMed); **Department of Ophthalmology, Shanghai Health and Medical Center, Wuxi, China** (T Chen MD); **Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China** (D Yang PhD, A R Ran PhD, S Szeto MBChB, C Y Cheung PhD); **Singapore Eye Research Institute, Singapore National Eye Centre, Singapore** (F He MSc, H Hamzah BSc, G S W Tan FRCSEd, Y-C Tham PhD, Prof C-Y Cheng MD PhD, C Sabanayagam MD PhD, Prof T Y Wong MD PhD); **Department of Medicine, Faculty of Medicine, Universiti Malaya, Kuala Lumpur, Malaysia** (S Singh MSc, A S A Raof MSc, J W S Lee-Boey MBBS, Prof S-K Lim MBBS, Prof L-L Lim PhD); **Department of Ophthalmology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China** (Prof X Sun MD); **Department of Nephrology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China** (S Ge MD, Prof G Xu MD); **Department of Nephrology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China** (Prof H Su); **Department of Ophthalmology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China** (Y Cheng MD); **National Engineering Research Centre for Big Data Technology and System, Services Computing Technology and System Laboratory, Cluster and Grid Computing Laboratory, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China** (F Lu PhD, Prof X Liao PhD,

Research in context

Evidence before this study

We searched PubMed for studies of deep learning-based chronic kidney disease (CKD), diabetic kidney disease (DKD), diabetic nephropathy, or non-diabetic kidney disease (NDKD) detection from retinal images published from database inception to March 10, 2024, using the search terms "detection of chronic kidney disease using retinal images", "detection of diabetic kidney disease using retinal images", "detection of diabetic nephropathy using retinal images", "detection of non-diabetic kidney disease using retinal images", "artificial intelligence", "machine learning", and "deep learning". The search was limited to English language publications. Although few studies have investigated the detection of CKD (diagnosed by reduced estimated glomerular filtration rate only) via deep learning applied to retinal images, to our knowledge no study has reported the detection of DKD (diagnosed by the reduced estimated glomerular filtration rate or the increased albumin to creatinine ratio) via deep learning applied to retinal images among multi-ethnic populations with diabetes. Furthermore, no study has explored the use of retinal images solely to differentiate isolated diabetic nephropathy from NDKD via deep learning.

Added value of this study

We developed and validated a retinal fundus image-based AI-deep learning system (termed DeepDKD, a single AI system using retinal fundus images solely as inputs) for detecting DKD; and among people with DKD, differentiating isolated diabetic nephropathy from NDKD, in multi-ethnic populations with diabetes (from China, the UK, Australia, Malaysia, and Singapore). For detecting DKD, DeepDKD achieved an area under the receiver operating characteristic curve (AUC) of 0·842 on the internal validation dataset and AUCs of 0·791–0·826 across ten external validation datasets. For differentiating isolated diabetic nephropathy from NDKD, DeepDKD achieved an AUC of 0·906 on the internal validation dataset and AUCs of 0·733–0·844 across three external validation datasets. Furthermore, we conducted two proof-of-concept studies to demonstrate the effectiveness of the DeepDKD system in clinical settings: a prospective real-world study in a primary care setting to evaluate the performance of the DeepDKD system in DKD screening among 325 participants with 3 months' follow-up; and a longitudinal analysis of the effectiveness of the DeepDKD system in differentiating isolated diabetic nephropathy from NDKD on renal function changes over time among 207 participants with 4·6 years of follow-up data.

Implications of all the available evidence

We established a single, non-invasive retinal image-based DeepDKD system which could improve access to screening and detection of DKD, prioritising patients who will need confirmatory laboratory tests in low-resource primary care settings. The ability of DeepDKD to differentiate isolated diabetic nephropathy from NDKD could also be used as an ancillary tool to minimise the need for invasive kidney biopsy. With the increasing availability of digital retinal fundus photography, and the future development of potentially even cheaper mobile phone-based fundus cameras, our DeepDKD system could be adopted as a relatively simple, real-time tool for screening of both eye (diabetic retinopathy) and kidney (DKD) complications in people with diabetes in primary care settings. DeepDKD also has promising applications in the clinical workflow of differentiating isolated diabetic nephropathy from NDKD in more specialised care settings to assist diabetes physicians and nephrologists to enable more precise and personalised diagnostic and therapeutic clinical decision making.

albumin-to-creatinine ratio (ACR) from urine.[6] However, measuring eGFR and ACR poses substantial challenges in primary care and low-resource settings, due to the limited awareness[7] and accessibility to conduct and analyse these tests.[8] The urine dipstick test to assess albuminuria is an alternative screening test because of its lower cost and simpler use.[9,10] However, the sensitivity of the urine dipstick test in detecting albuminuria varies from 43·6% to 69·4%,[10–12] which could lead to missed diagnosis and delayed treatment. Thus, substantial efforts are devoted to developing rapid, simple, point-of-care tests to improve DKD screening in primary care settings.

In patients with DKD, another important clinical process is differentiating kidney damage specifically due to diabetes, referred to as diabetic nephropathy, from other causes, referred to as non-diabetic kidney disease (NDKD). This differentiation is important because many forms of NDKD can be successfully treated, in contrast to isolated diabetic nephropathy, which has a more progressive course leading to irreversible kidney damage.[13] The prevalence of NDKD is high among people with diabetes, ranging from 33% to 72·5%.[13] However, the pathological diagnosis of NDKD typically involves an invasive kidney biopsy.[13] This procedure is not commonly undertaken even if needed, due to potential risks and contraindications.[14] Having access to simpler, non-invasive tests to distinguish isolated diabetic nephropathy from NDKD is useful for endocrinologists and nephrologists to narrow a differential diagnosis.

The retina provides a non-invasive way to evaluate diabetes complications. It is well known that diabetic retinopathy is closely linked to risk, occurrence, and progression of DKD.[15] Emerging artificial intelligence (AI) deep learning technologies[16] applied to retinal fundus images have been developed for automated diabetic retinopathy screening[17–19] as well as chronic kidney disease (CKD).[20,21] In people with diabetes, however, there have been no studies to our knowledge using retinal-image-based AI-deep learning algorithms for detecting DKD or differentiating isolated diabetic nephropathy from NDKD.

To address these two important gaps in DKD management, we aimed to develop and validate a single retinal

image-based AI-deep learning system (termed DeepDKD) to detect DKD and differentiate isolated diabetic nephropathy from NDKD. Furthermore, we aimed to conduct proof-of-concept clinical studies to evaluate the effectiveness of DeepDKD in clinical settings.

Methods

Study design, datasets, and diagnosis criteria

We conducted a multistage population-based study in three parts as described later, using retrospective and prospective datasets. To pretrain the DeepDKD system, we used 734 084 retinal fundus images from 90 067 participants in the Shanghai Diabetes Prevention Program cohort. To develop and internally validate DeepDKD for DKD detection, 486 312 retinal fundus images and clinical metadata (including age, sex [self-reported], BMI, current smoking status [yes or no], duration of diabetes, systolic blood pressure, diastolic blood pressure, HbA_{1c}, and diabetic retinopathy status [yes or no]) from 121 578 participants with diabetes in the Shanghai Integrated Diabetes Prevention and Care system were used. Ten external validation cohorts (the China National Diabetic Chronic Complications Study cohort, the Nicheng Diabetes Screening Project [NDSP] cohort, the Peking Union Diabetes Management [PUDM] cohort, the Eastern China Health Management cohort, the Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy cohort, the Singapore Epidemiology of Eye Diseases study cohort, the UK Biobank [UKB] cohort, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia [MeLODY] cohort, the Longitudinal Study of Ageing cohort, and the Australian Eye and Heart Study cohort) were collected (65 406 participants). To develop and internally validate DeepDKD for distinguishing isolated diabetic nephropathy from NDKD, 267 patients with diabetes who had undergone kidney biopsies in the Diabetic Retinopathy Progression Study cohort were used. Three external validation datasets (the Wuhan Kidney Disease Screening [WHKDS] cohort, the UKB cohort, and the MeLODY cohort) were collected (244 participants). The details of these included datasets are described in the table and appendix (pp 4–5, 15–16). For these datasets, we included participants with diabetes and gradable fundus images. We did not apply any other inclusion or exclusion criteria to the participants or images included in the present study. The retinal image information in these included datasets is described in the appendix (pp 17–18).

The detailed diagnosis criteria of DKD, isolated diabetic nephropathy, and NDKD used in this study are in the appendix (p 6). Diabetic retinopathy is diagnosed and staged according to the International Clinical Diabetic Retinopathy Severity Scale.[22]

This study was approved by the Ethics Committee of Shanghai Sixth People's Hospital (2019-087, approved Aug 29, 2019; 2023-KY-023(K), approved March 7, 2023) and Huadong Sanatorium (2022-14, approved Nov 11, 2022). Informed consent was obtained from participants in this study, which was conducted in accordance with the Declaration of Helsinki.

Model development of the DeepDKD system, metadata model, and combined model

We developed our system (DeepDKD) to detect DKD (non-DKD group, DKD-moderately increased risk group, DKD-high risk group, or DKD-very high risk group) and to distinguish isolated diabetic nephropathy from NDKD, using retinal fundus images only as inputs. DeepDKD used large-scale weakly supervised momentum contrastive learning as a method for extracting transferable visual representations of retinal fundus images with further supervised training in two tasks (appendix p 43). In the pretraining process, we employed the layer-wise adaptive rate scaling optimiser with a default weight decay value of 1×10^{-6}, training the model for 800 epochs with a learning rate of 0·01, a temperature τ of 0·07, and a batch size of 256. In the supervised training process, there were two classifiers. For the DKD classifier (task 1: detecting DKD), we used ResNet-50[23] as the backbone, and set its initial weights from the above well-trained weakly supervised momentum contrastive learning model. The DKD classifier was trained for a total of 100 epochs, and model selection was conducted based on metrics such as the area under the receiver operating characteristic curve (AUC), and Cohen's kappa coefficient on the internal validation set. For the diabetic nephropathy classifier (task 2: differentiating isolated diabetic nephropathy from NDKD), given the insufficient training data to mitigate model overfitting, we directly froze the encoder of the DKD classifier and trained a new linear classifier to differentiate isolated diabetic nephropathy from NDKD. For the classification of an individual participant, the predicted probability scores of available retinal fundus images were averaged.

In addition, we developed a metadata model and a combined model. Both the predicted probability scores from retinal fundus images (derived from DeepDKD) and clinical metadata were input into the combined model. We used the XGBoost classifier[24] in the metadata model and combined model. Details of the model development are in the appendix (pp 7–10).

Explainability analysis of the DeepDKD system

We utilised Grad-CAM[25] as a method for visualising the interpretability of the output predictions from the DeepDKD system. Grad-CAM provided insights into which regions of the input image were crucial for the model's prediction. For the output results of Grad-CAM, we selectively retained the retinal region and normalised pixel values to the range of 0–255.

Proof-of-concept clinical studies of the DeepDKD system

We conducted two proof-of-concept studies to demonstrate the clinical effectiveness of DeepDKD. First, we conducted a prospective real-world study on patients with type 2 diabetes

Prof H Jin PhD); **Department of Geriatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China** (C Deng MMed, L Ruan MD, Prof C Zhang MD); **Key Laboratory of Vascular Aging, Ministry of Education, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China** (C Deng MMed, L Ruan MD, Prof C Zhang MD); **Department of Ophthalmology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China** (C Wu MD, Prof R Dai MD); **Key Laboratory of River Basic Digital Twinning of Ministry of Water Resources, Macau University of Science and Technology, Macau Special Administrative Region, China** (W Wang PhD); **Department of Ophthalmology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China** (X Wang MD, Q Wu MD); **Department of Clinical Laboratory, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China** (J Tang MD); **Department of Nephrology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China** (X Sheng MD); **Department of Pathology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China** (Q Jiao MD); **Centre for Public Health, School of Medicine, Dentistry, and Biomedical Sciences, Queen's University Belfast, Belfast, UK** (G J McKay PhD, Prof R E Hogg PhD); **Westmead Institute for Medical Research, University of Sydney, NSW, Australia** (G Liew MBBS); **School of Computing, National University of Singapore, Singapore** (E Y L Chee PhD, Prof W Hsu PhD, Prof M L Lee PhD); **Department of Medicine and Therapeutics, Hong Kong Institute of Diabetes and Obesity, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong Special Administrative Region, China** (Prof A O Y Luk MD,

Prof J C N Chan MD, Prof L-L Lim PhD); **Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore** (Y-C Tham PhD, Prof C-Y Cheng MD PhD); **Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore** (Y-C Tham PhD, Prof C-Y Cheng MD PhD); **Ophthalmology and Visual Science Academic Clinical Program, Duke-NUS Medical School, Singapore** (Y-C Tham PhD, Prof C-Y Cheng MD PhD, C Sabanayagam MD PhD); **Beijing Key Laboratory of Intelligent Diagnostic Technology and Devices for Major Blinding Eye Diseases, Tsinghua Medicine, Tsinghua University, Beijing, China** (Prof T Y Wong MD PhD)

Correspondence to:
Prof Tien Yin Wong, Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital Eye Center, School of Clinical Medicine, Tsinghua Medicine, Tsinghua University, Beijing 100084, China
wongtienyin@tsinghua.edu.cn

or

Prof Bin Sheng, Department of Computer Science and Engineering, School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
shengbin@sjtu.edu.cn

or

Prof Huating Li, Department of Endocrinology and Metabolism, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai Diabetes Institute, Shanghai Clinical Centre for Diabetes, Shanghai 200233, China
huarting99@sjtu.edu.cn

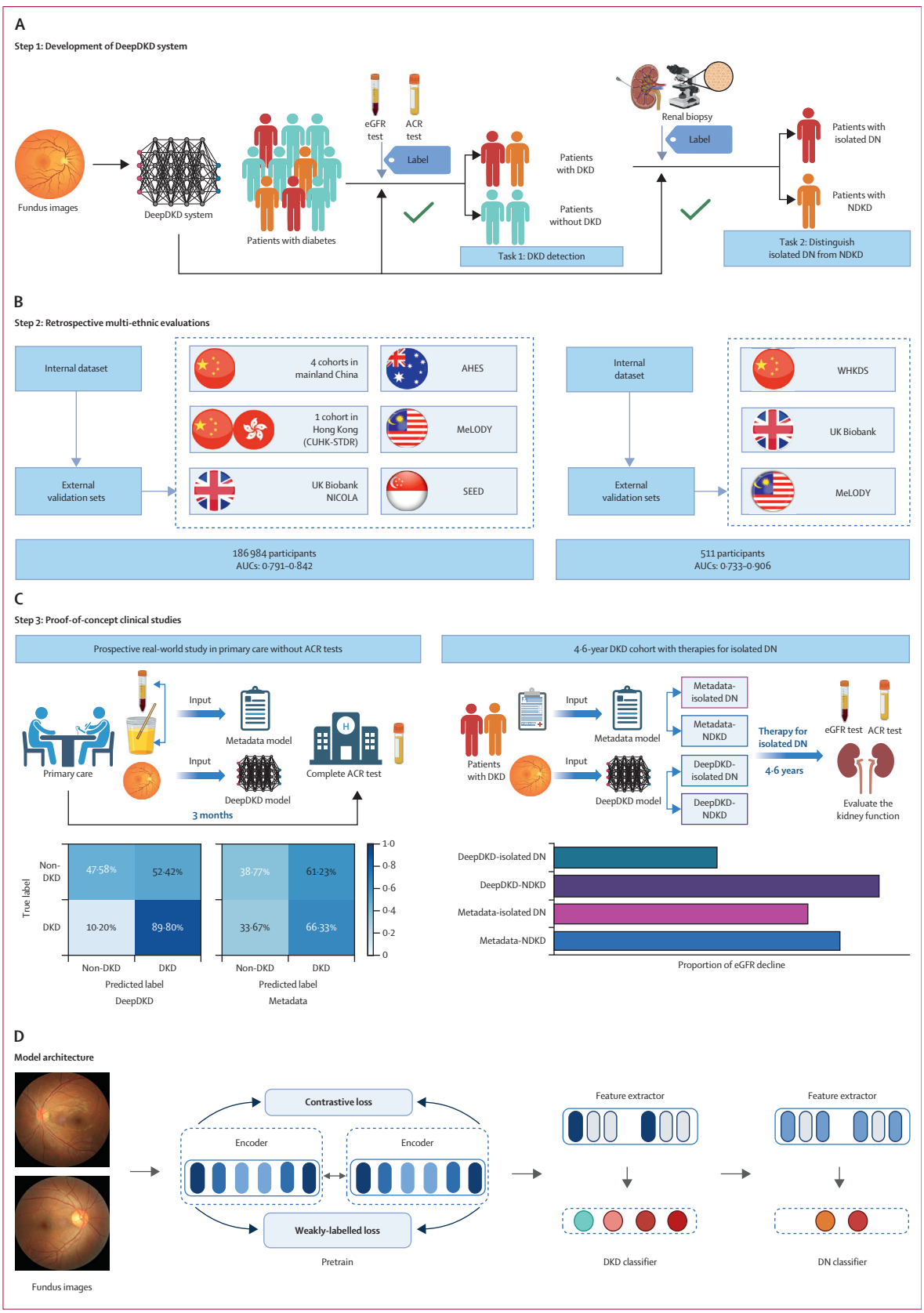
	Developmental dataset	Internal validation dataset	External validation datasets										
			CNDCS	NDSP	PUDM	ECHM	CUHK-STDR	SEED	UKB	MeLODY	NICOLA	AHES
Number of retinal fundus images	340 452	145 860	206 540	6788	1328	25 012	2015	6697	3709	2042	939	2056
Number of participants with diabetes	85 113	36 465	51 635	1697	332	6253	505	1684	1860	686	240	514
Race												
Chinese	85 113 (100%)	36 465 (100%)	51 635 (100%)	1697 (100%)	322 (100%)	6253 (100%)	505 (100%)	451 (26·78%)	6 (0·32%)	144 (20·99%)	..	..
Malay	..	..	..	..	..	..	..	201 (11·94%)		345 (50·29%)	..	..
Indian	..	..	..	..	..	..	..	1032 (61·28%)		186 (27·11%)	..	..
White	..	..	..	..	..	..	..	..	1460 (78·49%)	..	..	..
Mixed	..	..	..	..	..	..	..	..	16 (0·86%)	..	..	..
Asian or Asian British	..	..	..	..	..	..	..	..	167 (8·98%)	..	..	..
Black or Black British	..	..	..	..	..	..	..	..	117 (6·29%)	..	..	..
Caucasian	..	..	..	..	..	..	..	..	..	..	240 (100%)	313 (60·89%)
East Asian	..	..	..	..	..	..	..	..	..	..	..	34 (6·61%)
Middle Eastern	..	..	..	..	..	..	..	..	..	..	..	73 (14·20%)
South Asian	..	..	..	..	..	..	..	..	..	..	..	54 (10·51%)
Other	..	..	..	..	..	..	..	..	94 (5·05%)	11 (1·60%)	..	40 (7·78%)
Sex												
Women	46 521 (54·66%)	19 918 (54·62%)	25 955 (50·27%)	1048 (61·76%)	130 (39·16%)	2791 (44·63%)	236 (46·73%)	776 (46·08%)	595 (31·99%)	346 (50·44%)	81 (33·75%)	139 (27·04%)
Men	38 592 (45·34%)	16 547 (45·38%)	25 680 (49·73%)	649 (38·24%)	202 (60·84%)	3462 (55·37%)	269 (53·27%)	908 (53·92%)	1265 (68·01%)	340 (49·56%)	159 (66·25%)	375 (72·96%)
Age, years	66·05 (8·08)	66·19 (8·09)	57·11 (10·03)	62·37 (3·97)	51·53 (11·07)	58·57 (13·43)	60·93 (13·27)	61·18 (9·69)	59·76 (7·39)	60·94 (11·63)	65·44 (8·64)	63·42 (10·97)
Current smoker	8085 (9·50%)	3465 (9·50%)	11893 (23·03%)	284 (16·74%)	95 (28·61%)	1517 (24·26%)	57 (11·29%)	240 (14·25%)	229 (12·31%)	NA	57 (23·75%)	NA
BMI, kg/m^2	25·07 (3·09)	25·04 (3·09)	25·68 (3·64)	25·96 (3·36)	25·17 (3·61)	24·20 (3·22)	25·83 (4·50)	26·57 (4·78)	31·69 (5·62)	29·57 (6·77)	32·06 (5·98)	NA

(Table continues on next page)

	Developmental dataset	Internal validation dataset	External validation datasets									
			CNDCS	NDSP	PUDM	ECHM	CUHK-STDR	SEED	UKB	MeLODY	NICOLA	AHES
(Continued from previous page)												
Systolic blood pressure, mm Hg	142·47 (19·03)	142·41 (19·03)	137·18 (19·94)	138·17 (15·65)	122·60 (15·45)	117·91 (19·24)	121·94 (19·39)	142·46 (20·79)	145·53 (18·46)	141·56 (16·69)	134·22 (18·65)	NA
Diastolic blood pressure, mm Hg	80·66 (10·14)	80·59 (10·08)	80·11 (11·09)	84·10 (7·66)	79·61 (10·34)	74·49 (14·81)	77·15 (14·94)	77·71 (10·25)	82·95 (10·66)	76·30 (9·40)	79·94 (11·39)	NA
Cholesterol, mmol/L	4·98 (1·13)	4·98 (1·13)	4·98 (1·17)	5·34 (1·05)	4·70 (2·48)	4·91 (1·23)	4·27 (0·91)	5·00 (1·21)	4·64 (1·19)	4·51 (1·12)	NA	NA
HDL-cholesterol, mmol/L	1·36 (0·44)	1·35 (0·42)	1·27 (0·37)	1·29 (0·34)	1·22 (0·39)	1·20 (0·35)	1·36 (0·48)	1·10 (0·33)	1·20 (0·34)	1·21 (0·31)	NA	NA
LDL-cholesterol, mmol/L	2·99 (0·96)	3·00 (0·95)	3·01 (0·99)	3·23 (0·86)	2·68 (0·94)	2·87 (0·96)	2·27 (0·76)	3·03 (0·97)	2·81 (0·88)	2·49 (0·92)	NA	NA
Triglyceride, mmol/L	1·84 (1·47)	1·83 (1·41)	2·41 (2·72)	2·03 (2·02)	1·83 (2·08)	1·95 (2·17)	1·46 (0·93)	NA	2·12 (1·25)	1·83 (1·20)	NA	NA
HbA_{1c}, %	7·22 (1·24)	7·22 (1·25)	7·66 (1·92)	6·75 (1·21)	7·55 (3·46)	6·73 (2·26)	6·97 (1·81)	7·56 (1·57)	7·13 (1·51)	7·95 (1·96)	7·58 (1·72)	7·97 (1·94)
Fasting plasma glucose, mmol/L	7·97 (2·69)	7·96 (2·72)	9·02 (3·53)	7·76 (1·98)	8·47 (2·63)	8·17 (3·32)	7·78 (3·00)	NA	7·22 (3·09)	8·10 (3·94)	NA	NA
Duration of diabetes, years	7·77 (5·44)	7·77 (5·45)	6·98 (5·87)	4·95 (5·22)	5·22 (2·84)	5·23 (2·94)	12·04 (9·59)	7·72 (9·04)	8·56 (10·22)	8·67 (3·81)	6·17 (2·09)	NA
eGFR, mL min^{-1} per 1·73 m^2	85·72 (16·40)	85·57 (16·31)	90·62 (18·08)	94·28 (11·74)	96·10 (16·68)	91·76 (22·73)	81·12 (26·23)	81·72 (21·41)	88·32 (17·05)	77·43 (31·00)	75·80 (19·44)	78·17 (24·05)
ACR, mg/g	16·09 (7·60–41·98)	15·61 (7·42–39·96)	13·48 (5·34–41·69)	8·17 (5·31–15·42)	1·18 (0·68–3·04)	20·35 (9·41–68·48)	NA	17·68 (9·26–42·17)	15·15 (8·01–36·51)	38·30 (14·16–306·42)	NA	NA
Diabetic retinopathy	10337 (12·15%)	4589 (12·58%)	8821 (17·08%)	189 (11·14%)	37 (11·15%)	672 (10·78%)	331 (65·54%)	400 (23·75%)	94 (5·05%)	218 (31·78%)	28 (11·67%)	66 (12·84%)
DKD status												
Non-DKD group	54310 (63·81%)	23808 (65·29%)	34438 (66·70%)	1490 (87·80%)	314 (94·58%)	3618 (57·86%)	423 (83·76%)*	1032 (61·28%)	1244 (66·88%)	262 (38·19%)	187 (77·92%)*	398 (77·43%)*
Moderately increased risk group	23674 (27·81%)	9636 (26·43%)	12668 (24·53%)	178 (10·49%)	11 (3·31%)	1710 (27·35%)	NA	429 (25·48%)	477 (25·65%)	166 (24·20%)	NA	NA
High risk group	5248 (6·17%)	2209 (6·06%)	3186 (6·17%)	20 (1·18%)	5 (1·51%)	607 (9·71%)	NA	133 (7·90%)	93 (5·00%)	106 (15·45%)	NA	NA
Very high risk group	1881 (2·21%)	812 (2·23%)	1343 (2·60%)	9 (0·53%)	2 (0·60%)	318 (5·09%)	NA	90 (5·34%)	46 (2·47%)	152 (22·16%)	NA	NA

Data are n, %, n (%), mean (SD), or median (IQR). Data are presented as mean (SD) or median (IQR) for continuous variables, or n (%) for categorical variables. ACR=albumin-to-creatinine ratio. AHES=Australian Eye and Heart Study. CNDCS=China National Diabetic Chronic Complications Study. CUHK-STDR=Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy. DKD=diabetic kidney disease. ECHM=Eastern China Health Management. eGFR=estimated glomerular filtration rate. HbA_{1c}=glycated haemoglobin. MeLODY=Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia. NA=not available. NDSP=Nicheng Diabetes Screening Project. NICOLA=Northern Ireland Cohort for the Longitudinal Study of Ageing. PUDM=Peking Union Diabetes Management. SEED=Singapore Epidemiology of Eye Diseases study. UKB=UK Biobank. *DKD was diagnosed by a decreased eGFR of less than 60 mL/min^{-1} per 1·73 m^2 in CUHK-STDR, NICOLA, and AHES cohorts because ACR is not available for these participants.

Table: **Characteristics of the datasets used in the development and validation of detecting DKD**


A
Step 1: Development of DeepDKD system
Fundus images
DeepDKD system
Patients with diabetes
eGFR test
ACR test
Label
Patients with DKD
Patients without DKD
Task 1: DKD detection
Renal biopsy
Label
Patients with isolated DN
Patients with NDKD
Task 2: Distinguish isolated DN from NDKD
B
Step 2: Retrospective multi-ethnic evaluations
Internal dataset
External validation sets
4 cohorts in mainland China
AHES
1 cohort in Hong Kong (CUHK-STDR)
MeLODY
UK Biobank NICOLA
SEED
186 984 participants
AUCs: 0·791–0·842
Internal dataset
External validation sets
WHKDS
UK Biobank
MeLODY
511 participants
AUCs: 0·733–0·906
C
Step 3: Proof-of-concept clinical studies
Prospective real-world study in primary care without ACR tests
Primary care
Input
Metadata model
Input
DeepDKD model
Complete ACR test
3 months
Non-DKD
DKD
True label
47·58%
52·42%
10·20%
89·80%
Non-DKD
DKD
Predicted label
DeepDKD
38·77%
61·23%
33·67%
66·33%
Non-DKD
DKD
Predicted label
Metadata
1·0
0·8
0·6
0·4
0·2
0
4·6-year DKD cohort with therapies for isolated DN
Patients with DKD
Input
Metadata model
Input
DeepDKD model
Metadata-isolated DN
Metadata-NDKD
DeepDKD-isolated DN
DeepDKD-NDKD
Therapy for isolated DN
4·6 years
eGFR test
ACR test
Evaluate the kidney function
DeepDKD-isolated DN
DeepDKD-NDKD
Metadata-isolated DN
Metadata-NDKD
Proportion of eGFR decline
D
Model architecture
Fundus images
Contrastive loss
Encoder
Encoder
Weakly-labelled loss
Pretrain
Feature extractor
DKD classifier
Feature extractor
DN classifier

from a primary care centre in Shanghai, China (without ACR tests) to evaluate the effectiveness of integrating DeepDKD into the routine clinical practice for screening and detecting DKD. The sensitivity and specificity of the metadata model, DeepDKD system, and combined model were compared using the high-sensitivity operating points (preset at 95%) selected from the internal validation dataset. Second, we conducted an analysis of retinal fundus images and clinical metadata in a community-based cohort with longitudinal data to demonstrate the effectiveness of DeepDKD in distinguishing isolated diabetic nephropathy from NDKD. We aimed to compare the proportion of eGFR decline between the isolated diabetic nephropathy group and NDKD group identified by the metadata model, DeepDKD system, and combined model, to assess the discriminative ability of these three models. Details of the methodology of these two clinical studies are in the appendix (pp 11–12).

Statistical analysis

We developed three different AI models for detecting DKD and differentiating isolated diabetic nephropathy from NDKD: DeepDKD system, metadata model, and combined model. The performance of the three models was evaluated in the validation datasets using the AUC generated by plotting sensitivity versus 1–specificity. We originally measured sensitivity and specificity at an optimal threshold balancing the two measures and calculated positive predictive value and negative predictive value for all the validation datasets. For analysis of the model performance using different operating points, we selected a high-sensitivity operating point (preset at 95%), a high-specificity point (preset at 95%), and a balanced operating point (based on maximising Youden's index) from the internal validation dataset. Cluster-bootstrap, biased-corrected, asymptotic two-sided 95% CIs adjusted for clustering by patients were calculated and presented for sensitivities, specificities, and AUCs. All hypotheses tested were two-sided, and a p of less than 0·05 was considered statistically significant. Analyses were performed using SPSS (version 25.0) and Python (version 3.8.16).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

See Online for appendix

We conducted a three-part population-based study (figure 1). The characteristics of participants included in this study are summarised in the table and appendix (pp 15–16).

We evaluated the performance of the DeepDKD system, metadata model, and combined model in detecting the presence of DKD. In the internal validation dataset, the metadata model achieved an AUC of 0·680 (95% CI 0·674–0·685); the DeepDKD system achieved an AUC of 0·842 (0·838–0·846), which was superior to the metadata model (p<0·0001); and the combined model achieved an AUC of 0·847 (0·843–0·851; figure 2; appendix pp 21–24). The performance of the combined model was only slightly better than that of DeepDKD (p<0·0001), demonstrating the accurate detection performance of DeepDKD. On the ten external validation datasets, the AUCs were 0·573–0·721 for the metadata model, 0·791–0·826 for the DeepDKD system, and 0·767–0·838 for the combined model. Moreover, DeepDKD could accurately stratify participants with DKD into three stages (moderately increased risk group or worse [AUCs 0·791–0·842], high risk group or worse [AUCs 0·792–0·906], or very high risk group [AUCs 0·573–0·967]) across the internal and external validation datasets.

Furthermore, DeepDKD showed comparable performance of DKD detection in different subgroups of participants stratified by diabetic retinopathy status and glycaemic control (appendix pp 27–28). We also compared the performance of the DeepDKD system and urine dipstick protein test in DKD screening (appendix p 29). Among two external validation datasets (NDSP and PUDM), the sensitivity of DeepDKD (77·8% in both datasets) was substantially higher than that of the urine dipstick protein test (19·8% vs 0%).

For real-world clinical applications, the operating point could be set to confidently differentiate healthy controls from DKD cases, to reduce missed DKD cases. Performance metrics for DKD detection using the DeepDKD system, metadata model, and combined model were determined by the high-sensitivity operating points selected from the internal validation dataset (appendix pp 30–31). For external validation datasets, the DeepDKD system achieved sensitivities from 79·3% to 98·8%, and maintained specificities ranging from 31·7% to 66·6%. The combined model also achieved sensitivities from 81·9% to 98·8% and specificities from 28·6% to 64·1%. The metadata model

Figure 1: **Overall study design of the development and validation of the DeepDKD system**

(A) Step 1: study design of the development of the DeepDKD system. There are two tasks of the DeepDKD system, using retinal fundus images as inputs. The first task is to detect and stage DKD among patients with diabetes. The second task is to differentiate isolated DN NDKD among patients with DKD. (B) Step 2: retrospective multi-ethnic evaluations of the DeepDKD system in the internal and external validation datasets. For detecting DKD, ten external validation datasets from multiple countries were included. For differentiating isolated DN from NDKD, three external validation datasets from three countries were included. (C) Step 3: two proof-of-concept clinical studies of the DeepDKD system. In the first real-world prospective study, we aimed to compare the DKD screening accuracy of the DeepDKD system and that of the metadata model, in primary care setting without ACR tests. In the second longitudinal study, we aimed to compare the effectiveness of the DeepDKD system and that of the metadata model in differentiating isolated DN from NDKD. (D) Model architecture of the DeepDKD system. The DeepDKD system was first pretrained using weakly supervised momentum contrastive learning to extract transferable visual representations from retinal fundus images. Then, the system was trained to detect DKD (four categories: Non-DKD group, DKD-moderately increased risk group, DKD-high risk group, or DKD-very high risk group) and distinguish isolated DN from NDKD, using retinal fundus images only as inputs. The two fundus images shown are examples of inputs to the DeepDKD system. The blue cells represent distinct network layers, where varying shades of blue indicate weight differences across each layer. ACR=albumin-to-creatinine ratio. AHES=Australian Eye and Heart Study. AUC=area under the receiver operating characteristic curve. CUHK-STDR=Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy. DKD=diabetic kidney disease. DN=diabetic nephropathy. eGFR=estimated glomerular filtration rate. MeLODY=Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia. NDKD=non-diabetic kidney disease. NICOLA=Northern Ireland Cohort for the Longitudinal Study of Ageing. SEED=Singapore Epidemiology of Eye Diseases study. WHKDS=Wuhan Kidney Disease Screening.

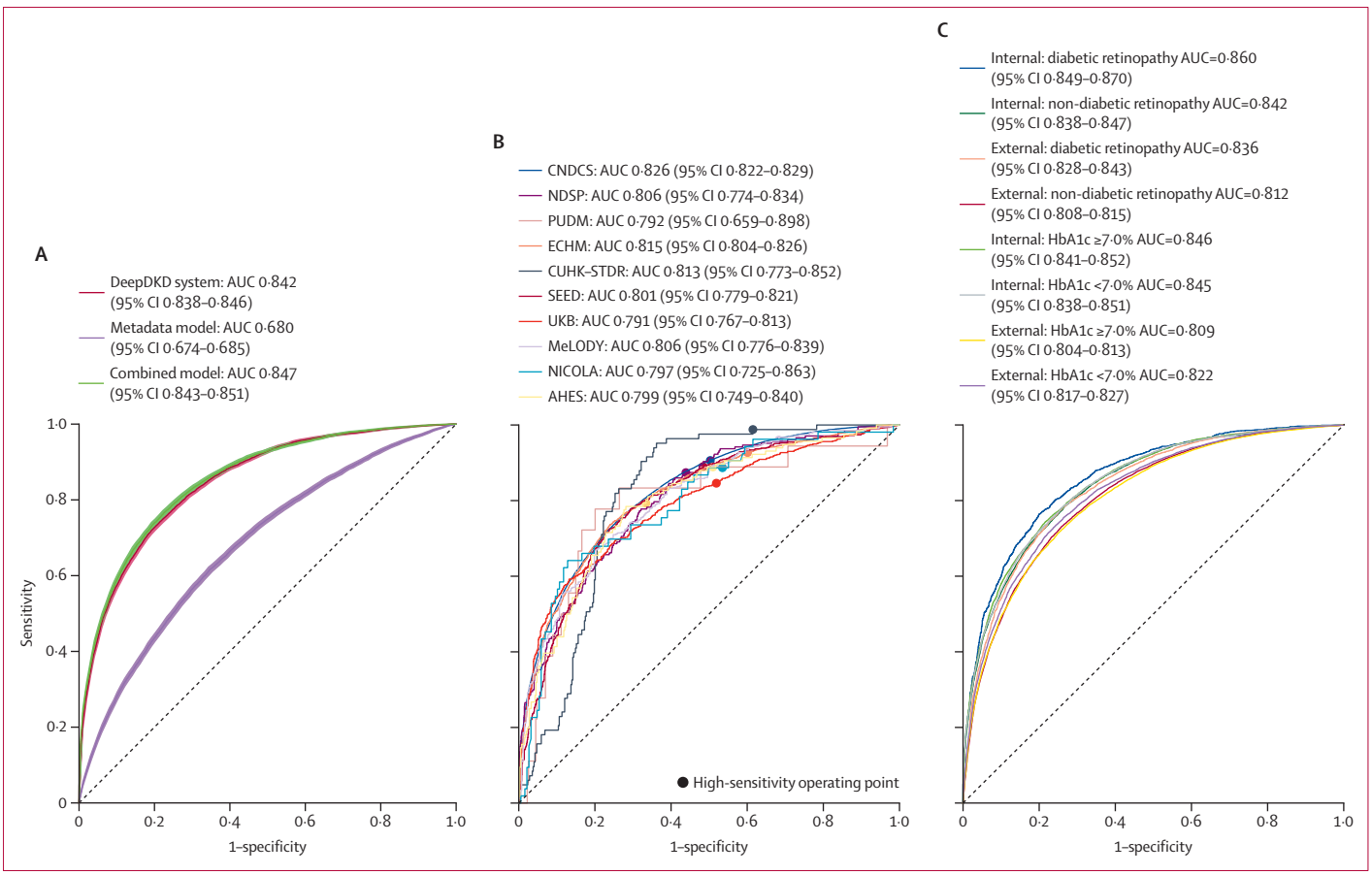


Figure 2: **Receiver operating characteristic curve and AUC of the DeepDKD system for detecting diabetic kidney disease in the internal validation dataset, external validation datasets, and different subgroups**

(A) Performance of the DeepDKD system, metadata model, and combined model in the internal validation dataset. (B) Performance of the DeepDKD system in the external validation datasets. Notably, the performance of the DeepDKD system in the external validation datasets using the high-sensitivity operating point selected from the internal validation dataset is shown. (C) Performance of the DeepDKD system in different subgroups in the internal and external validation datasets. Two subgroup analyses were conducted: (1) participants with different statuses of retinopathy (participants with diabetic retinopathy *vs* those without diabetic retinopathy), and (2) participants with different statuses of glycaemic control (participants with HbA_{1c} <7% *vs* participants with HbA_{1c} ≥7%). AHES=Australian Eye and Heart Study. AUC=area under the receiver operating characteristic curve. CNDCS=China National Diabetic Chronic Complications Study. CUHK-STDR=The Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy. ECHM=The Eastern China Health Management. HbA_{1c}=glycated haemoglobin. MeLODY=The Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia. NDSP=Nicheng Diabetes Screening Project. NICOLA=The Northern Ireland Cohort for the Longitudinal Study of Ageing. PUDM=Peking Union Diabetes Management. SEED=The Singapore Epidemiology of Eye Diseases study. UKB=UK Biobank.

achieved sensitives from 77·6% to 98·3%, but specificities only ranged from 4·2% to 31·5%.

We also developed the DeepDKD system, metadata model, and combined model for differentiating isolated diabetic nephropathy from NDKD. In the internal validation dataset, the DeepDKD system achieved an AUC of 0·906 (95% CI 0·825–0·966), which was significantly higher than that of the metadata model (0·772 [0·656–0·868], p=0·029) and comparable to that of the combined model (0·949 [0·895–0·988], p=0.23; figure 3; appendix p 32). In three external multi-ethnic cohorts (WHKDS, MeLODY, and UKB), the AUCs of the metadata model were 0·633–0·718, the DeepDKD system were 0·733–0·844, and the combined model were 0·767–0·852. Furthermore, DeepDKD could accurately differentiate isolated diabetic nephropathy from DKD in different subgroups of participants stratified by diabetic retinopathy status, proteinuria, and haematuria (appendix pp 34–36). Performance metrics for differentiating isolated diabetic nephropathy from NDKD using the DeepDKD system were also determined by the operating points selected from the internal validation dataset (appendix p 37). To identify NDKD from isolated diabetic nephropathy with high confidence, we used a high-sensitivity operating point. For external validation datasets, we achieved sensitivities from 80·0% to 98·4%. To identify isolated diabetic nephropathy with high confidence, we used a high-specificity operating point. For external validation datasets, we achieved specificities from 80·0% to 100·0%.

To better understand how DeepDKD could detect DKD or isolated diabetic nephropathy using retinal fundus images, we conducted explainability analyses using the Grad-CAM method to provide insights into the regions that could influence the predictions of DeepDKD. Several

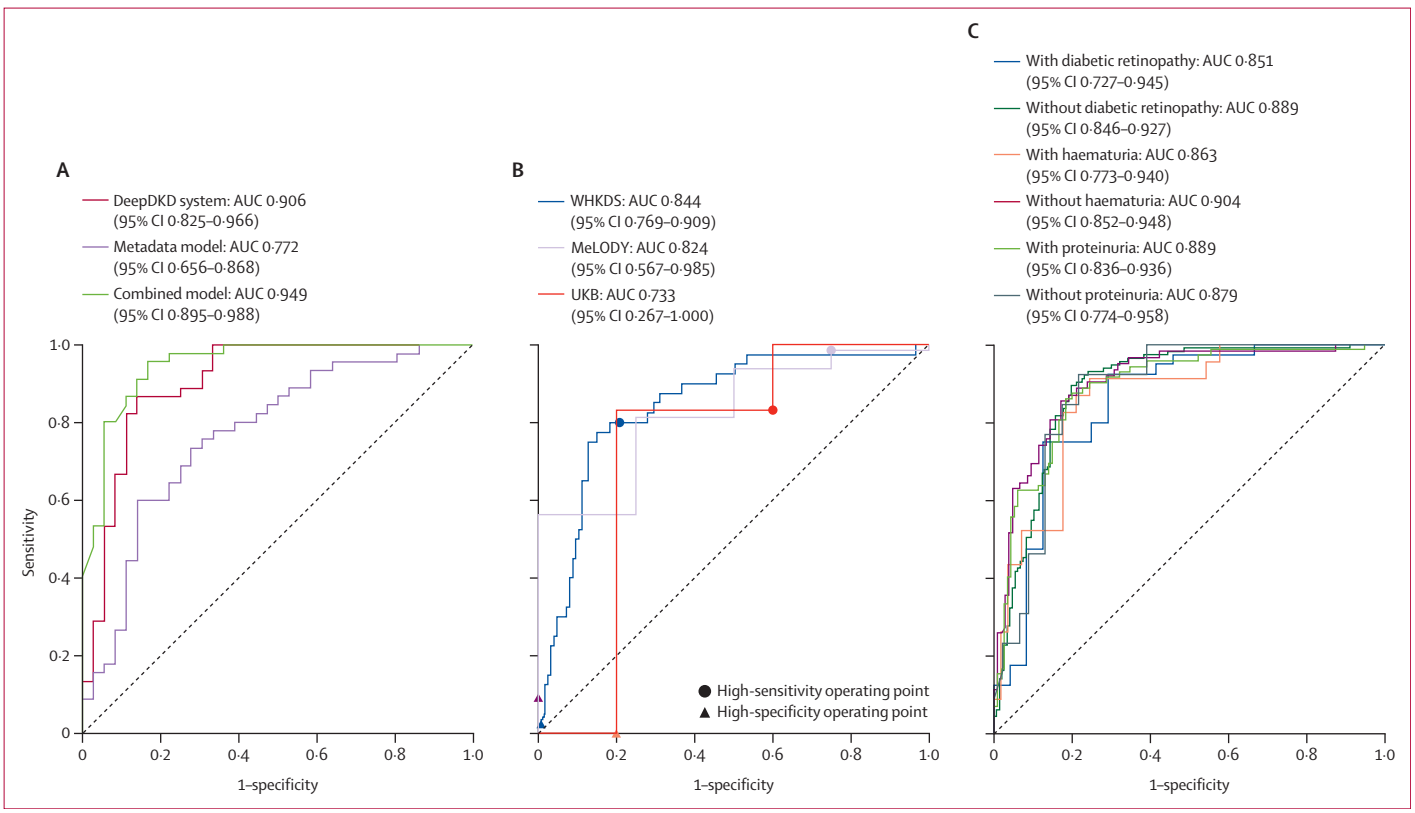


Figure 3: **Receiver operating characteristic curve and AUC of the DeepDKD system for differentiating isolated diabetic nephropathy and non-diabetic kidney disease in the internal validation dataset, external validation datasets, and different subgroups**

(A) Performance of the DeepDKD system, metadata model, and combined model in the internal validation dataset. (B) Performance of the DeepDKD system in the external validation datasets. Notably, the performance of the DeepDKD system in the external validation datasets using the high-sensitivity and high-specificity operating points selected from the internal validation dataset are shown. (C) Performance of the DeepDKD system in different subgroups in the internal and external validation datasets. Three subgroup analyses were conducted: (1) participants with different statuses of retinopathy (participants with diabetic retinopathy *vs* those without diabetic retinopathy), (2) participants with different statuses of proteinuria (participants with proteinuria *vs* those without proteinuria), and (3) participants with different statuses of haematuria (participants with haematuria *vs* those without haematuria). AUC=area under the receiver operating characteristic curve. MeLODY=The Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia. UKB=UK Biobank. WHKDS=Wuhan Kidney Disease Screening.

representative examples of original retinal fundus images and corresponding saliency maps are in the appendix (pp 44–45). Corresponding kidney biopsy images are also in the appendix (p 45). The results of explainability analyses indicated that our DeepDKD system detected DKD or isolated diabetic nephropathy based on changes related to the dilatation of venules, rarefaction of vessels, and diabetic retinopathy changes.

In the first real-world prospective study (figure 4), among 325 patients with eGFR of 60 mL/min per 1·73m^2 or more and negative urine dipstick protein at baseline, 98 patients could be diagnosed with DKD, supported by elevated ACR in further examinations. Using the high-sensitivity operating point selected from the internal validation dataset, the DeepDKD system could achieve both higher sensitivity (89·8% *vs* 66·3%, p<0·0001) and higher specificity (47·6% *vs* 38·8%, p=0·072) compared with the metadata model (appendix p 38). The sensitivity (86·7%) and specificity (50·7%) of the combined model were comparable to the DeepDKD system (sensitivity p=0·65 and specificity p=0·48). Furthermore, the DeepDKD system showed comparable performance of DKD detection in different subgroups of participants stratified by diabetic retinopathy status and glycaemic control (appendix p 39).

In the second longitudinal study (figure 5), there was a significant difference (p=0·0010) in the proportion of eGFR decline between participants of isolated diabetic nephropathy (27·45%) and those of NDKD (52·56%) identified by the DeepDKD system (appendix p 40). Similarly, there was a significant difference (p=0·0084) in the proportion of eGFR decline between participants of isolated diabetic nephropathy (35·71%) and those of NDKD (51·82%) identified by the combined model. However, the proportion of eGFR decline between participants of isolated diabetic nephropathy (41·46%) and those of NDKD (47·59%) identified by the metadata model was similar (p=0·20). These results suggested that some individuals with NDKD were misdiagnosed as isolated diabetic nephropathy in clinical practice using the metadata model.

Discussion

Improving the accessibility and ease of routine screening of DKD addresses a major public health need in diabetes management. In some clinical scenarios, differentiating

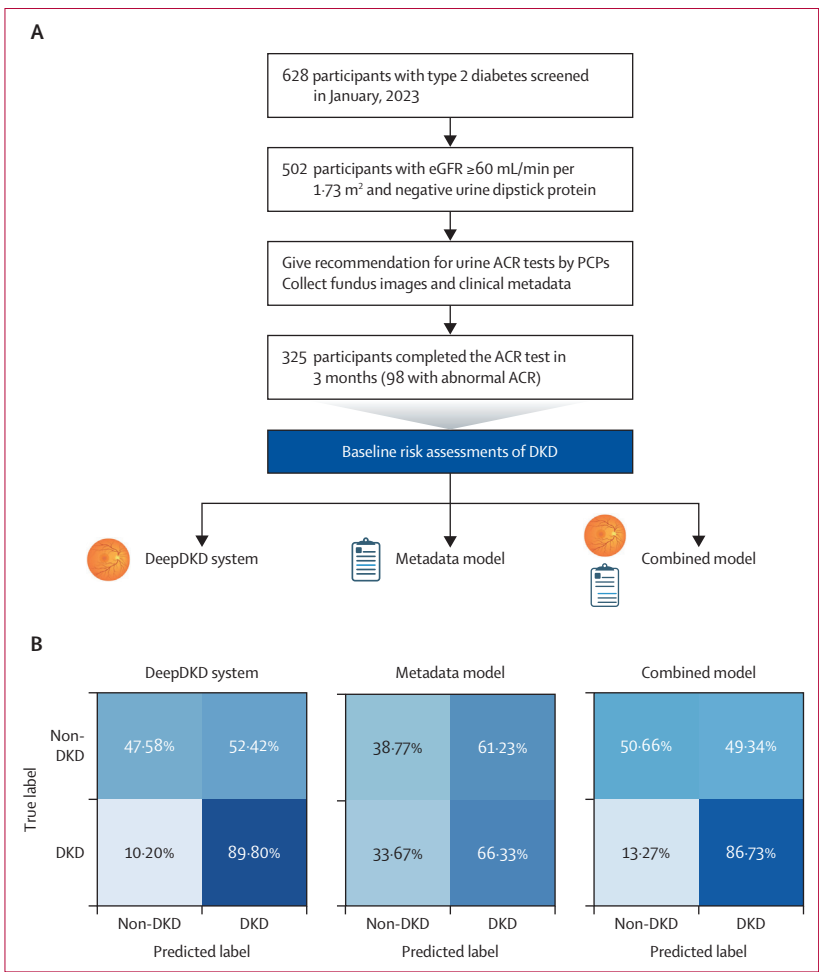


Figure 4: **Real-world prospective study on screening DKD among patients with type 2 diabetes, in a primary care setting without ACR tests**

(A) Flowchart of the real-world prospective study. (B) Comparison of the screening accuracy of the DeepDKD system, metadata model, and combined model on screening diabetic kidney disease, using the high-sensitivity operating points selected from the internal validation dataset. ACR=albumin-to-creatine ratio. DKD=diabetic kidney disease. eGFR=estimated glomerular filtration rate. PCP=primary care physician.

isolated diabetic nephropathy from NDKD, currently performed via kidney biopsy, is a major challenge among endocrinologists and nephrologists in managing DKD. In this study, we developed, validated and tested a single retinal fundus image-based AI-deep learning system (DeepDKD) for detecting DKD; and among people with DKD, accurately differentiating isolated diabetic nephropathy from NDKD (appendix p 46). We demonstrated that DeepDKD could be integrated into routine clinical workflow in the primary care setting, supporting its use for DKD screening. We also demonstrated that among people with DKD, DeepDKD could accurately differentiate isolated diabetic nephropathy from NDKD, allowing the possibility of personalised management of DKD, as diabetic nephropathy and NDKD have different therapeutic options. This added capability of DeepDKD might serve as an adjunctive tool to assist endocrinologists and nephrologists in more precise clinical decision making.

To the best of our knowledge, this study is the first work to establish an AI-deep learning system for DKD detection based on retinal fundus images alone. Previous work focused on developing AI-deep learning algorithms to detect moderate or advanced generalised kidney damage (ie, CKD) among people in the general community.[20,21] Although conceptually useful, screening for general CKD in community settings is currently not widely accepted, nor shown to be cost-effective.[26] Our current study is more specifically targeted at screening for DKD among patients with diabetes, which is universally recommended in clinical guidelines, because early detection of DKD facilitates timely intervention to prevent kidney disease progression.[27] DKD screening using blood or urine tests is often infrequently conducted (eg, studies suggest that only 32·2–38·7% of individuals with diabetes have routine DKD screening[28,29]), resulting in missed diagnosis and delayed intervention.[8,30] The low screening coverage is partly due to the lack of awareness by physicians and people with diabetes, and insufficient financial resources to conduct and analyse blood or urine tests, particularly in primary care and resource-limited community settings. Because digital retinal fundus photography for routine diabetic retinopathy screening is recommended,[31] DeepDKD could be integrated into routine diabetic retinopathy screening to conduct cost-effective DKD screening without additional examinations. Furthermore, there are already several deep learning algorithms to detect diabetic retinopathy from retinal images.[17,18] The development of a single deep learning algorithm to detect both diabetic retinopathy and DKD is a promising future research direction, to facilitate the screening of multiple diabetes complications using retinal images.

Our study addresses a major public health problem in current DKD screening. Currently, instead of measurement of eGFR and ACR, the simpler urine dipstick protein tests are often used to assess albuminuria in primary care and low-resource settings because of lower costs, simplicity, and fast results.[32,33] However, the low sensitivity of urine dipstick protein tests has been a major hurdle and might lead to missed diagnoses. Our DeepDKD system shows robust screening performance across multiple datasets, demonstrating that it could serve as a simpler, non-invasive tool to detect early stages of DKD for further confirmation by urine or blood tests. Retinal imaging can be performed by minimally trained health-care workers and integrated into a simpler comprehensive model of eye and kidney screening among people with diabetes.[34] Furthermore, we demonstrated that the sensitivity of DeepDKD was substantially higher than that of the urine dipstick protein test. Moreover, in our proof-of-concept prospective real-world study, we showed that DeepDKD could accurately identify DKD patients among participants with eGFR of 60 mL/min per 1·73 m^2 or more and a negative urine dipstick protein. As a result, the integration of DeepDKD into clinical workflows of DKD screening could not only improve the screening sensitivity but also

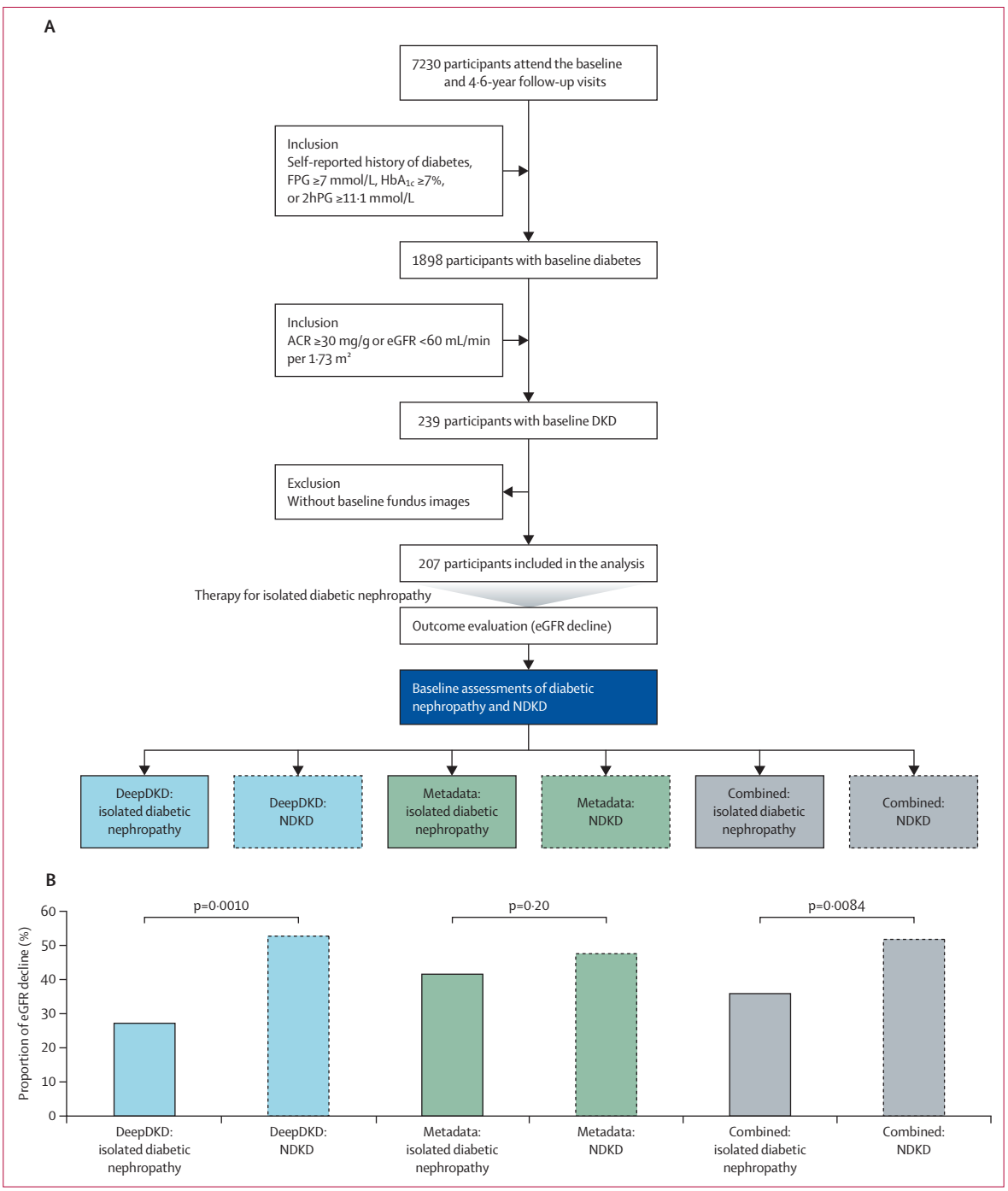


Figure 5: **Longitudinal study on differentiating isolated diabetic nephropathy from NDKD among patients with DKD**
(A) Flowchart of the longitudinal study. (B) Comparison of the proportion of estimated glomerular filtration rate decline between isolated diabetic nephropathy group and NDKD group identified by the DeepDKD system, metadata model, or combined model. We used logistic regression models to compare the proportion of eGFR decline between different groups, adjusting for baseline age, sex, BMI, duration of diabetes, systolic blood pressure, diastolic blood pressure, and glycated haemoglobin. 2hPG=2-h post-load plasma glucose. ACR=albumin-to-creatine ratio. DKD=diabetic kidney disease. eGFR=estimated glomerular filtration rate. FPG=fasting plasma glucose. HbA_{1c}=glycated haemoglobin. NDKD=non-diabetic kidney disease.

provide a fast point-of-care screening method for primary care or low-resource regions with limited laboratory resources.

Our study addresses a second clinical gap. Among patients with DKD, those suspected of NDKD are sometimes recommended to undergo invasive kidney biopsies[35]

given the different treatment approaches for isolated diabetic nephropathy and NDKD. Our deep learning system could differentiate isolated diabetic nephropathy from NDKD based solely on non-invasive retinal images with high accuracy, thus improving the ease and possibility of future clinical adoption. Notably, DeepDKD maintained its discriminative capability among different subgroups stratified by diabetic retinopathy status, proteinuria, and haematuria. It is well known that diabetic nephropathy and diabetic retinopathy are closely correlated, and diabetic retinopathy is often used in clinical practice to differentiate isolated diabetic nephropathy from NDKD.[36] However, some patients with isolated diabetic nephropathy are without diabetic retinopathy. For these patients, DeepDKD could also be utilised to differentiate isolated diabetic nephropathy from NDKD accurately, using retinal images alone. In the longitudinal study, we showed that participants who were identified as isolated diabetic nephropathy by the DeepDKD system had a better kidney function prognosis than participants identified by the metadata model, indicating that some participants with NDKD were misdiagnosed as isolated diabetic nephropathy in clinical practice using the metadata model. These findings highlight the potential benefit of integrating DeepDKD within the clinical workflow to assist physicians in differentiating isolated diabetic nephropathy from NDKD for improved risk stratification and tailored management.

Our study had some limitations. First, the DeepDKD system was trained in a Chinese population followed by validation in multi-ethnic cohorts from Singapore, Malaysia, Australia, and the UK. Further testing in more diverse multi-ethnic populations could further demonstrate its generalisability and robustness. Second, some potential biases cannot be eliminated in the present retrospective study such as data distribution, selection bias, and unknown confounders. Third, the performance of DeepDKD showed variations across the validation datasets. The variation in imaging quality and characteristics between different types of fundus cameras could influence the performance. Further studies are needed to evaluate the generalisability and interoperability of DeepDKD across different types of fundus cameras. Fourth, although two proof-of-concept clinical studies were conducted to evaluate the effectiveness of DeepDKD, further randomised controlled trials are needed to evaluate the impact of real-time system integration in routine workflow on patient outcomes.

In conclusion, we established a single, non-invasive retinal image-based DeepDKD system which could improve access to DKD screening, prioritising patients who will need confirmatory laboratory tests in low-resource primary care settings. The ability of DeepDKD to differentiate isolated diabetic nephropathy from NDKD can also be used as an ancillary tool to minimise the need for invasive kidney biopsy. With the increasing availability of digital retinal fundus photography, and the future development of potentially even cheaper mobile phone-based fundus cameras, DeepDKD could be adopted as a relatively simple, real-time tool for screening of both eye and kidney complications in people with diabetes in primary care settings. DeepDKD also has promising applications in the clinical workflow differentiating isolated diabetic nephropathy from NDKD in more specialised care settings to assist specialists in enabling more precise and personalised diagnostic and therapeutic clinical decision making.

Contributors

ZM, ZG, SY, YW, JShe, CYC, GSWT, Y-CT, C-YC, CS, L-LL, WJ, HL, BS, and TYW conceptually designed the study. ZM, ZG, SY, YW, YZ, JShe, TC, DY, ARR, FH, HH, SS, ASAR, JWSL-B, S-KL, XSu, SG, GX, HS, YC, FL, XL, HJ, CD, LR, CZ, CW, RD, YJ, TL, RL, JL, JShu, YL, XW, YQ, JT, XSh, QJ, GJM, REH, GL, EYLC, WH, MLL, CYC, GSWT, Y-CT, C-YC, CS, L-LL, WJ, HL, BS, and TYW collected and analysed the data. ZM, ZG, SY, YZ, and TL performed the statistical analysis. ZM, ZG, and SY drafted the manuscript. ZM, ZG, SY, YZ, HL, BS, and TYW discussed and reviewed the manuscript. All authors critically edited, read, and approved the final manuscript. ZM and TYW had full access to all the data in the study and had responsibility for the integrity of the data and the accuracy of the data analysis. ZM and TYW accessed and verified the data. All authors had final responsibility for the decision to submit for publication. ZM, ZG, SY, YW, YZ, JShe, CYC, GSWT, Y-CT, C-YC, CS, L-LL, and WJ contributed equally and shared co-first authorship. CS, L-LL, WJ, HL, BS, and TYW jointly supervised this work and shared the co-last authorship.

Declaration of interests

CCL declares honoraria paid to their institution from Boehringer Ingelheim, AstraZeneca, and Sebia. TC is supported by the Clinical Special Program of Shanghai Municipal Health Commission (20224044) and the three-year action plan to strengthen the construction of public health system in Shanghai (2023-2025 GWVI-11.1-28). XY is supported by the Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102). REH is supported by Optos and Novartis. GL is supported by the Australian National Health and Medical Research Council (571012). SS received consulting fees from Bayer, Astella, and Roche; payment or honoraria from Bayer and Roche; and support for attending meetings and travel from Bayer and Roche. AOYL received research grants from Amgen, Bayer, Biogen, Boehringer Ingelheim, MSD, Novo Nordisk, Roche, Health Bureau, and Hong Kong Research Grants Council outside of the submitted work; travelling support from the International Diabetes Federation and the Asian Association for the Study of Diabetes for attending conferences; serves as the President of the Hong Kong Association for the Study of Obesity; and is a council member of Hong Kong Society of Endocrinology, Diabetes and Metabolism. GSWT received research grants from Artificial Intelligence Singapore, the Health Services Research Grant from the National Medical Research Council, the Clinician Scientist Award-Investigator Grant from the National Medical Research Council, and the Large Collaborative Grant from the National Medical Research Council; consulting fees from AbbVie-Allergan, Bayer, Haag-Steit, Leica, Novartis, Roche, Rxilient, and Zeiss; payment or honoraria from Bayer, Haag-Steit, Leica, Novartis, Roche, Rxilient, and Zeiss; and stock from Eyris. C-YC received consulting fees from Medi-Whale. L-LL declares grants paid to institution from AstraZeneca and Abbott; payment or honoraria from AstraZeneca, Boehringer Ingelheim, Novo Nordisk, ZP Therapeutics, Sanofi, Viatris, and Novartis. WJ is supported by the Shanghai Municipal Key Clinical Specialty, Shanghai Research Centre for Endocrine and Metabolic Diseases (2022ZZ01002), and the Chinese Academy of Engineering (2022-XY-08). HL is supported by the Excellent Young Scientists Fund of the National Natural Science Foundation of China (82022012) and the General Fund of the National Natural Science Foundation of China (81870598), and the Innovative research team of high-level local universities in Shanghai (SHSMU-ZDCX20212700). BS is supported by the General Program of the National Natural Science Foundation of China (62272298) and the National Key R & D Program of China (2022YFC2407000). TYW is supported by the National Key R & D

Program of China (2022YFC2502800) and the National Natural Science Foundation of China (8238810007); has received consulting fees from AbbVie, Aldropika Therapeutics, Bayer, Boehringer Ingelheim, Carl Zeiss, Genentech, Iveric Bio, Novartis, Opthea, Quaerite Bipharm Research, Plano, Roche, Sanofi, and Shanghai Henlius; and is a co-founder of companies EyRiS and VISRE. All other authors declare no competing interests.

Data sharing

The de-identified individual-participant data are available on request from Tien Yin Wong. The code used in the current study for developing the algorithm is available at https://github.com/DeepDKD/deepdkd.

Acknowledgments

This study was supported by the National Key R & D Program of China (2022YFC2502800), the National Natural Science Foundation of China (82388101), and the Beijing Natural Science Foundation (IS23096) to TYW; the Shanghai Municipal Key Clinical Specialty, Shanghai Research Centre for Endocrine and Metabolic Diseases (2022ZZ01002), the Excellent Young Scientists Fund of the National Natural Science Foundation of China (82022012), and the Innovative research team of high-level local universities in Shanghai (SHSMU-ZDCX20212700) to HL; the National Key R & D Program of China (2022YFC2407000), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0509200), the General Program of the National Natural Science Foundation of China (62272298) to BS; the Clinical Special Program of Shanghai Municipal Health Commission (20224044), and the three-year action plan to strengthen the construction of public health system in Shanghai (2023-2025 GWVI-11.1-28) to TC. The Atlantic Philanthropies, the Economic and Social Research Council, the UKCRC Centre of Excellence for Public Health Northern Ireland, the Centre for Ageing Research and Development in Ireland, the Office of the First Minister and Deputy First Minister, the Health and Social Care Research and Development Division of the Public Health Agency, the Wellcome Trust Wolfson Foundation and Queen's University Belfast provided core financial support for NICOLA Waves 1 and 2. Additionally, we are grateful to all the participants of the NICOLA Study, and the whole NICOLA team, which includes nursing staff, research scientists, clerical staff, computer and laboratory technicians, managers, and receptionists. This research was conducted using the UK Biobank Resource (application number 104443). The computation resources in this paper were provided by the Center for High Performance Computing at Shanghai Jiao Tong University, and the AI for Science Platform supported by the Artificial Intelligence Institute at Shanghai Jiao Tong University.

References

1 Bikbov B, Purcell CA, Levey AS, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2020; **395:** 709–33.

2 Xie Y, Bowe B, Mokdad AH, et al. Analysis of the Global Burden of Disease study highlights the global, regional, and national trends of chronic kidney disease epidemiology from 1990 to 2016. *Kidney Int* 2018; **94:** 567–81.

3 Thomas MC, Brownlee M, Susztak K, et al. Diabetic kidney disease. *Nat Rev Dis Primers* 2015; **1:** 15018.

4 Wu H, Lau ESH, Yang A, et al. Trends in kidney failure and kidney replacement therapy in people with diabetes in Hong Kong, 2002–2015: a retrospective cohort study. *Lancet Reg Health West Pac* 2021; **11:** 100165.

5 Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V. Chronic kidney disease. *Lancet* 2021; **398:** 786–802.

6 de Boer IH, Khunti K, Sadusky T, et al. Diabetes management in chronic kidney disease: a consensus report by the American Diabetes Association (ADA) and Kidney Disease: Improving Global Outcomes (KDIGO). *Diabetes Care* 2022; **45:** 3075–90.

7 Duru OK, Middleton T, Tewari MK, Norris K. the landscape of diabetic kidney disease in the United States. *Curr Diab Rep* 2018; **18:** 14.

8 Wen CP, Chang CH, Tsai MK, et al. Diabetes with early kidney involvement may shorten life expectancy by 16 years. *Kidney Int* 2017; **92:** 388–96.

9 Sumida K, Nadkarni GN, Grams ME, et al. conversion of urine protein-creatinine ratio or urine dipstick protein to urine albumin-creatinine ratio for use in chronic kidney disease screening and prognosis: an individual participant-based meta-analysis. *Ann Intern Med* 2020; **173:** 426–35.

10 White SL, Yu R, Craig JC, Polkinghorne KR, Atkins RC, Chadban SJ. Diagnostic accuracy of urine dipsticks for detection of albuminuria in the general community. *Am J Kidney Dis* 2011; **58:** 19–28.

11 Park JI, Baek H, Kim BR, Jung HH. Comparison of urine dipstick and albumin:creatinine ratio for chronic kidney disease screening: a population-based study. *PLoS One* 2017; **12:** e0171106.

12 Lim D, Lee DY, Cho SH, et al. Diagnostic accuracy of urine dipstick for proteinuria in older outpatients. *Kidney Res Clin Pract* 2014; **33:** 199–203.

13 Anders HJ, Huber TB, Isermann B, Schiffer M. CKD in diabetes: diabetic kidney disease versus nondiabetic kidney disease. *Nat Rev Nephrol* 2018; **14:** 361–77.

14 Luciano RL, Moeckel GW. Update on the native kidney biopsy: core curriculum 2019. *Am J Kidney Dis* 2019; **73:** 404–15.

15 Wong CW, Wong TY, Cheng CY, Sabanayagam C. Kidney and eye diseases: common risk factors, etiological mechanisms, and pathways. *Kidney Int* 2014; **85:** 1290–302.

16 Guan Z, Li H, Liu R, et al. Artificial intelligence in diabetes management: advancements, opportunities, and challenges. *Cell Rep Med* 2023; **4:** 101213.

17 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; **316:** 2402–10.

18 Ting DSW, Cheung CY-L, Lim G, et al. development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017; **318:** 2211–23.

19 Sheng B, Pushpanathan K, Guan Z, et al. Artificial intelligence for diabetes care: current and future prospects. *Lancet Diabetes Endocrinol* 2024; **12:** 569–95.

20 Zhang K, Liu X, Xu J, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat Biomed Eng* 2021; **5:** 533–45.

21 Sabanayagam C, Xu D, Ting DSW, 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.

22 Wilkinson CP, Ferris FL 3rd, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003; **110:** 1677–82.

23 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition 2016 IEEE Conference on Computer Vision and Pattern Recognition; June 27–30, 2016 (abstr https://10.1109/CVPR.2016.90).

24 Chen T, Guestrin C. Xgboost: a scalable tree boosting system. KDD '16: the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Aug 13–17, 2016.

25 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis* 2017; **2017:** 618–26.

26 Levin A, Okpechi IG, Caskey FJ, Yang C-W, Tonelli M, Jha V. Perspectives on early detection of chronic kidney disease: the facts, the questions, and a proposed framework for 2023 and beyond. *Kidney Int* 2023; **103:** 1004–08.

27 Lim L-L, Chow E, Chan JCN. Cardiorenal diseases in type 2 diabetes mellitus: clinical trials and real-world practice. *Nat Rev Endocrinol* 2023; **19:** 151–63.

28 Collins AJ, Foley RN, Chavers B, et al. United States Renal Data System 2011 annual data report: atlas of chronic kidney disease & end-stage renal disease in the United States. *Am J Kidney Dis* 2012; **59** (1 Suppl 1): e1–420.

29 Ferrè S, Storfer-Isser A, Kinderknecht K, et al. Fulfillment and validity of the kidney health evaluation measure for people with diabetes. *Mayo Clin Proc Innov Qual Outcomes* 2023; **7:** 382–91.

30 Chan JCN, Lim L-L, Wareham NJ, et al. The *Lancet* Commission on diabetes: using data to transform diabetes care and patient lives. *Lancet* 2021; **396:** 2019–82.

31 Wong TY, Sun J, Kawasaki R, et al. Guidelines on diabetic eye care: The International Council of Ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. *Ophthalmology* 2018; **125:** 1608–22.

32 Levin A, Stevens PE, Bilous RW, et al. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl* 2013; **3:** 1–150.

33 Jeon J, Kim J. Dipstick proteinuria and risk of type 2 diabetes mellitus: a nationwide population-based cohort study. *J Transl Med* 2021; **19:** 271.

34 Navaneethan SD, Zoungas S, Caramori ML, et al. Diabetes management in chronic kidney disease: synopsis of the KDIGO 2022 clinical practice guideline update. *Ann Intern Med* 2023; **176:** 381–87.

35 Espinel E, Agraz I, Ibernon M, Ramos N, Fort J, Serón D. Renal biopsy in type 2 diabetic patients. *J Clin Med* 2015; **4:** 998–1009.

36 Liang S, Zhang X-G, Cai G-Y, et al. Identifying parameters to distinguish non-diabetic renal diseases from diabetic nephropathy in patients with type 2 diabetes mellitus: a meta-analysis. *PLoS One* 2013; **8:** e64184.