


A noninvasive model for chronic kidney disease screening and common pathological type identification from retinal images

Received: 12 October 2024

Accepted: 17 July 2025

Published online: 29 July 2025

 Check for updates

Qianni Wu^{1,22}, Jianbo Li^{2,3,22}, Lanqin Zhao^{1,22}, Dong Liu^{1,22}, Jingyi Wen^{1,22}, Yunuo Wang^{2,3,22}, Yiqin Wang^{2,3,22}, Naya Huang^{2,3}, Lanping Jiang^{2,3}, Qinghua Liu^{2,3}, Hanming Lin⁴, Pengxia Wan⁴, Shicong Yang⁵, Wenfang Chen⁵, Hongjian Ye^{2,3}, Mohammed Haji Rashid Hassan⁶, Ahmed Hassan Nur⁶, Zefang Dai^{2,3}, Jie Guo⁷, Shanshan Zhou⁷, Jianwen Yu^{2,3}, Weixing Zhang¹, Wenben Chen¹, Ruiyang Li¹, Wai Cheng Iao¹, Juan-juan Feng⁸, Yan Wang⁸, Hua Hong⁸, Peihong Yin⁹, Qing Ye⁹, Chao Xie¹⁰, Min Zhu¹⁰, Xiaoyi Liu¹⁰, Yaozhong Kong¹⁰, Jie Wang¹¹, Ruiying Ma¹¹, Yu Xiao¹¹, Guoguang Chen¹², Rongguo Fu¹³, Yuhe Ke^{14,15}, Jasmine Ong Chiat Ling^{15,16}, Charumathi Sabanayagam^{17,18}, Daniel Shu Wei Ting^{15,18}, Kar Keung Cheng¹⁹, Duoru Lin¹✉, Wei Chen^{2,3}✉ & Haotian Lin^{1,20,21}✉

Chronic kidney disease (CKD) is a global health challenge, but invasive renal biopsies, the gold standard for diagnosis and prognosis, are often clinically constrained. To address this, we developed the kidney intelligent diagnosis system (KIDS), a noninvasive model for renal biopsy prediction using 13,144 retinal images from 6773 participants. The KIDS achieves an area under the receiver operating characteristic curve (AUC) of 0.839–0.993 for CKD screening and accurately identifies the five most common pathological types (AUC: 0.790–0.932) in a multicenter and multi-ethnic validation, outperforming nephrologists by 26.98% in accuracy. Additionally, the KIDS further predicts disease progression based on pathological classification. Given its flexible strategy, the KIDS can be adapted to local conditions to provide a tailored tool for patients. This noninvasive model has the potential to improve CKD clinical management, particularly for those who are ineligible for biopsies.

Chronic kidney disease (CKD) is a substantial worldwide public health problem affecting ~850 million people, and will be life-threatening in the end stage^{1–3}. As the gold standard for the pathological diagnosis and prognosis of CKD, renal biopsy provides significant benefits for both patients and physicians by improving understanding of the disease state and indicating the extent of active and chronic lesions⁴, enabling physicians to make targeted treatment plans to preserve kidney function and reduce the incidence of kidney failure and death⁵. However, renal biopsy is often limited by high-risk anatomical

characteristics, systemic diseases with high mortality, and the lack of biopsy techniques and nephrology services in less developed regions^{6,7}. The most common complication after percutaneous renal biopsy is bleeding, with hematoma occurring in 11% of cases and transfusion required in 1.6% of cases^{8,9}. Additionally, there is an immense shortage of nephrologists in low-income regions, and significant gaps and variability exist in the availability and accessibility of kidney care compared with other regions¹⁰. When patients are unable to undergo renal biopsy or lack access to pathological diagnoses,

A full list of affiliations appears at the end of the paper. ✉ e-mail: lindr5@mail.sysu.edu.cn; chenwei99@mail.sysu.edu.cn; linht5@mail.sysu.edu.cn

treatment plans are always reliant on the subjective judgment of nephrologists, and the prognoses are difficult to determine.

Several efforts, such as those utilizing the proteome¹¹, microRNA¹², and molecular imaging¹³, have been investigated for noninvasive diagnosis and prognostic prediction for CKD. A recent breakthrough study revealed the significant associations between plasma proteins and kidney pathological lesions, which is expected to provide novel biomarkers for the noninvasive diagnosis of CKD¹⁴. However, the marker specificity and causal relationships require further clarification, and the associations have not been independently validated. Additionally, medical imaging has become an important tool for CKD diagnosis through the noninvasive observation of renal structural abnormalities. Artificial intelligence (AI)-assisted medical image analysis has been studied extensively, which can help physicians identify imaging features that cannot be easily discerned by the naked eye, providing valuable information for the early detection, pathological, and prognostic assessment of CKD^{15,16}. Despite the positive advances in AI-assisted medical image analysis for CKD-related detection, a noninvasive model for renal biopsy prediction has not yet been achieved.

Alterations in microvascular structure and function often precede end-organ damage, and microvascular dysfunction in peripheral beds can mirror dysfunction in visceral beds¹⁷. As a convenient window to observe microvascular and neural tissues, the retina can be used for noninvasive detection, diagnosis, staging, and management of systemic diseases^{18–20}. Based on deep learning (DL) algorithms, qualitative associations have been established between ocular features and systemic diseases, such as hepatobiliary diseases²¹, diabetes²², and cardiovascular diseases²³, providing rapid and complementary methods for the screening and identification of systemic diseases²⁴. Specifically, there is a graded relationship between retinopathy severity and declining renal function^{25,26} that can guide CKD detection and onset prediction via retinal image-based DL models^{27,28}. However, primary detection cannot meet further clinical management demands for CKD patients, and accurate treatment decision-making and prognostic evaluation still rely on renal biopsies²⁹. Therefore, it remains urgent to develop an AI model that is as effective as invasive renal biopsy for noninvasive pathological and prognostic prediction of CKD.

In this study, we developed the kidney intelligent diagnosis system (KIDS), a noninvasive model for early detection, pathological diagnosis, and long-term prognosis prediction of CKD using retinal images or in combination with clinical data. We trained the KIDS using retinal images from CKD patients who underwent renal biopsy, **or in combination with clinical data**. This will result in the accurate and effective prediction of pathological types of CKD. The generalizability of the models was validated on an external multicenter and multi-ethnicity dataset from South China, Kashi in the Xinjiang Autonomous Region, and Somalia. Moreover, the performance of the KIDS was compared with that of 12 nephrologists for pathological diagnosis in a prospective dataset. The KIDS was capable of achieving the diagnostic ability of invasive renal biopsy. We also demonstrated that the KIDS has the ability to stratify CKD patients according to prognosis and risk for 5 years. In general, by providing noninvasive, objective pathological and prognostic predictions for CKD patients, this noninvasive model has the potential to improve kidney care quality and reduce the incidence of end-stage renal disease (Fig. 1), particularly in less developed regions.

Results

Characteristics of the datasets

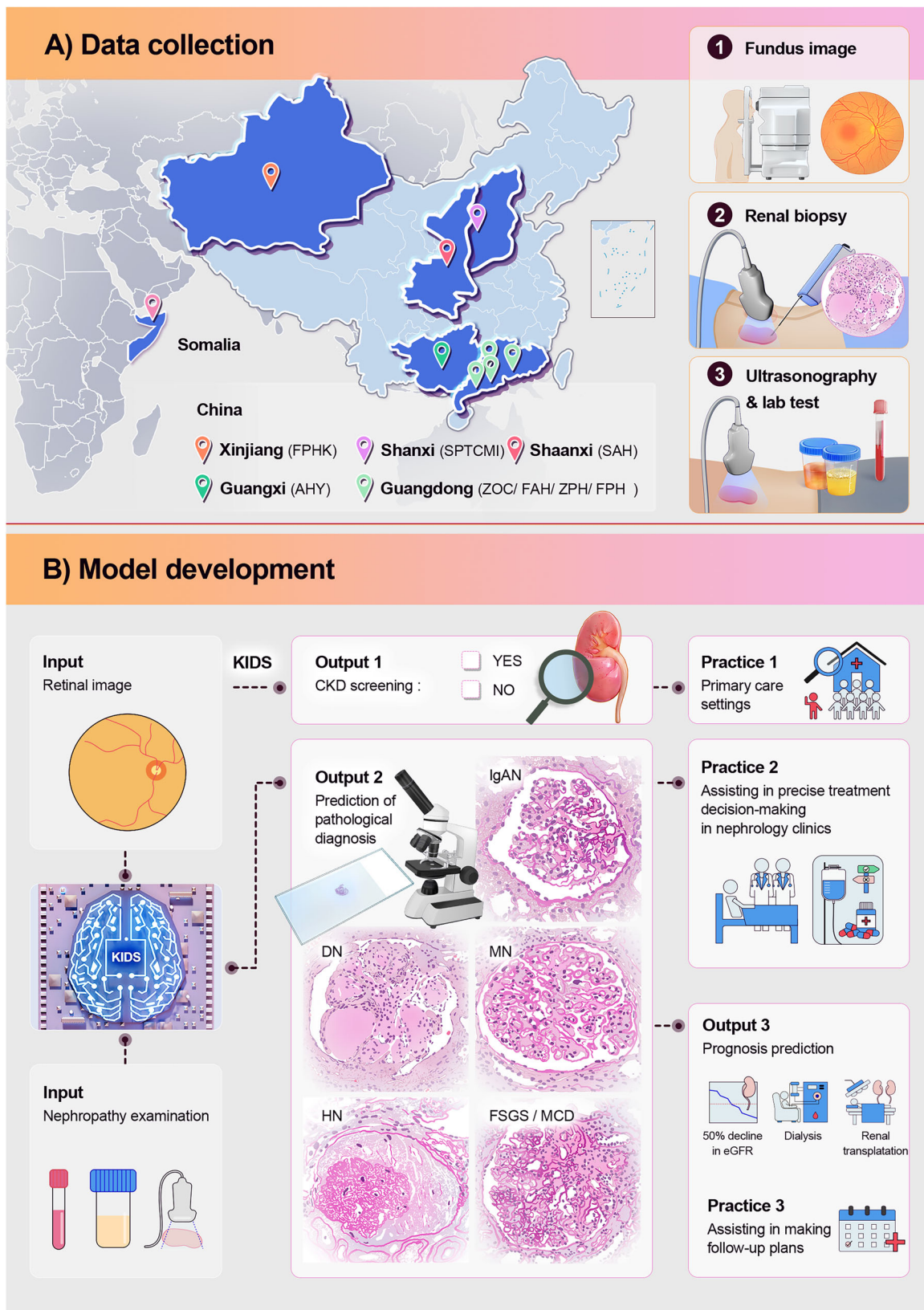
The participant characteristics are summarized in Table 1. For the development and internal testing of the CKD screening DL model, a total of 7609 retinal images from 3965 participants (1731 patients, 2234 controls) were included. Almost half of the patients (46.2%) were in the early stage of CKD, with 17.6% in the moderate stage and 36.3% in the advanced stage. For external validation, we used an external test set consisting of 1461 retinal images from 743 participants with annual health check results.

To predict the pathological diagnosis and the percentage of sclerotic glomeruli (PSG) stage of CKD, 4252 retinal images from 2139 CKD patients who underwent renal biopsy were included from the First Affiliated Hospital of Sun Yat-sen University [FAH, Guangzhou, China] retrospective cohort for the training, validation, and internal testing of the pathological prediction model. The median eGFR was 52.0 mL/min/1.73 m² (interquartile range [IQR]: 21.4–95.2). IgA nephropathy (IgAN, $n=1059$) was the most common pathological type in the datasets, followed by idiopathic membranous nephropathy (MN, $n=547$), arterionephrosclerosis (ANS, $n=382$), diabetic nephropathy (DN, $n=342$), idiopathic minimal change disease (MCD), and focal segmental glomerulosclerosis (FSGS) (MCD/FSGS, $n=261$). A total of 179 cases diagnosed with two or more pathological types from the above five types were individually counted for each type. Patients with MCD/FSGS (34 [IQR 19–53]) were younger than those with MN (54 [IQR 45–63]; $p<0.001$). A total of 7.5% of patients who underwent biopsy presented high PSG (>75%), which was higher in DN patients (13.4%). The demographics and clinical information of the participants with different pathological types are listed in Supplementary Table 1. To assess the generalizability of the pathological diagnosis and PSG staging AI models, an external test dataset containing 358 individuals with 708 images was collected from three other independent centers: Zhongshan City People's Hospital [ZPH, Zhongshan, China], First People's Hospital of Foshan [FPH, Foshan, China] and the Affiliated Hospital of Youjiang Medical University for Nationalities [AHY, Youjiang, Guangxi Zhuang Autonomous Region, China]. Furthermore, real-world test datasets were used to further assess the generalizability of the models. These datasets included 272 CKD patients with 34 retinal images from the First People's Hospital of Kashi [FPHK, Kashi, Xinjiang Uygur Autonomous Region, China], 215 individuals with 73 retinal images from Shanxi Provincial Traditional Chinese Medicine Institute [SPTCMI, Taiyuan, China] and The Second Affiliated Hospital of Xi'an Jiaotong University [SAH, Xi'an, China], and 99 CKD patients from Banadir Hospital [BH, Mogadishu, Somalia] in East Africa. In addition, a total of 574 retinal images prospectively collected from 301 participants with renal biopsy results from FAH were ultimately used as a prospective dataset to test the model's performance.

To compare the ability of the KIDS with that of human nephrologists in diagnosing pathological conditions, we established a prospective, multicenter test dataset comprising 256 patients with renal biopsy results from four independent centers, FAH, ZPH, FPH, and AHY, collected between May 2023 and November 2023. Patients with MCD/FSGS were generally younger (32.0 years [IQR 9.50–46.50]) and predominantly male (70.4%). In contrast, DN patients were significantly older (52.0 years [IQR 44.5–59.0]). Consistent with the derivation cohort, a greater percentage of DN and ANS patients had elevated PSG levels (>75%), at 13.3% and 19.4%, respectively. (Supplementary Table 5).

CKD screening DL model using retinal images

We developed KIDS to distinguish CKD patients from non-CKD controls (Fig. 2a) and achieved an AUC of 0.959 (95% CI: 0.943, 0.975) for CKD screening in the internal test dataset. Furthermore, since distinguishing early CKD patients from non-CKD controls is essential for prompt disease intervention, we performed a subgroup analysis to assess the ability of the KIDS to distinguish early CKD patients, moderate CKD patients, and advanced CKD patients from controls, which yielded AUCs of 0.936 (95% CI: 0.909, 0.962), 0.947 (95% CI: 0.913, 0.982) and 0.993 (95% CI: 0.987, 0.999), respectively (Fig. 2b). In the external test dataset, the KIDS achieved AUCs of 0.858 (95% CI: 0.831, 0.884) for CKD screening and similar performances of 0.839 (95% CI: 0.806, 0.871), 0.897 (95% CI: 0.852, 0.943) and 0.889 (95% CI: 0.837, 0.942) for early, moderate and advanced CKD detection, respectively, in the subgroup analysis (Fig. 2c, d). The model successfully identified CKD patients at different simulated prevalences, with AUCs stabilized at -0.960 on the internal test and 0.860 on the external test (Supplementary Table 4).



Pathological diagnosis and staging of CKD via a noninvasive AI model

Pathological diagnosis multimodal model. We developed a non-invasive model consisting of three submodels for the KIDS (using image-only, clinical data-only, and hybrid data) to identify different pathological diagnoses in CKD patients. In the image-only model, the

AUCs ranged from 0.703 to 0.902 for IgAN, MN, DN, ANS, and MCD/ FSGS in the internal test set, whereas the clinical data-only model achieved AUCs between 0.858 and 0.970. Then, we trained a multimodal AI model using retinal images and clinical data combined, which achieved better performance. Compared with those of the clinical data-only model, the AUCs for the hybrid models were greater, but this

Fig. 1 | A kidney intelligent diagnosis system (KIDS) based on retinal images and clinical data for noninvasive pathological diagnosis and prognosis prediction of chronic kidney disease. **A** Illustration of the geographical distribution map of data sources used in the Kidney Intelligent Diagnosis System (KIDS). The development set, external and multicenter, and multi-ethnic real-world test datasets were collected from 9 hospitals. Data collection included retinal images, renal biopsy, and laboratory and ultrasound examinations. **B** We used retinal images to develop deep learning models for CKD screening. Combined with nephropathy examinations, a multimodal model was developed to predict 5 renal pathology classifications. Then, based on the predicted renal pathology, prognosis

predictions were made for CKD patients. AI artificial intelligence, ZOC Zhongshan Ophthalmic Center of Sun Yat-sen University, FAH First Affiliated Hospital of Sun Yat-sen University, ZPH Zhongshan City People's Hospital, FPH First People's Hospital of Foshan, AHY Affiliated Hospital of Youjiang Medical University for Nationalities, FPHK First People's Hospital of Kashi, SPTCMI Shanxi Provincial Traditional Chinese Medicine Institute, SAH The Second Affiliated Hospital of Xi'an Jiaotong University, BH Banadir Hospital in Somalia. IgAN IgA nephropathy, MN idiopathic membranous nephropathy, DN diabetic nephropathy, ANS arterionephrosclerosis, MCD/FSGS idiopathic minimal change disease and focal segmental glomerulosclerosis.

Table 1 | Baseline characteristics of the participants

	Screening model	Pathological diagnosis model				Progression prediction model
		Development set	Internal test set	External test set	Prospective test set	
Participants	3965	1818	321	358	301	999
Retinal images	7609	3617	635	708	574	1989
Demographic factors						
Age, years	39 [31–50]	43 [32–54]	44 [33–56]	47 [36–54]	37 [30–49]	46 [34–56]
Sex, male	1808 (45.6)	1145 (63.0)	189 (58.9)	215 (60.1)	170 (56.5)	654 (65.5)
Systolic BP, mmHg	125 [114–138]	136 [122–153]	137 [123–153]	147 [130–163]	130 [117–146]	138 [124–155]
Diastolic BP, mmHg	78 [70–86]	85 [76–95]	85 [76–95]	90 [80–102]	83 [76–92]	85 [77–96]
BMI, kg/m ²	23.0 [20.8–25.5]	23.7 [21.4–26.2]	24.1 [21.5–26.7]	24.5 [22.4–27.0]	23.5 [21.2–25.8]	23.99 [21.7–26.3]
CKD	1731 (43.7)					
eGFR, mL/min per 1.73 m ²	99.8 [64.5–114.3]	51.7 [21.7–95.1]	54.9 [20.2–95.2]	74.6 [44.8–105.1]	67.7 [35.5–102.3]	52.23 [24.8–92.8]
CKD stages						
Early CKD	900 (46.2)	820 (45.1)	149 (46.4)	221 (61.7)	177 (58.8)	443 (44.3)
Moderate CKD	304 (17.6)	383 (21.1)	66 (20.6)	79 (22.1)	56 (18.6)	249 (24.9)
Advanced CKD	626 (36.3)	615 (33.8)	106 (33.0)	58 (16.2)	68 (22.6)	307 (30.7)
Albuminuria	818 (20.6)	1711 (93.8)	301 (93.8)	272 (90.4)	338 (94.4)	
Pathological diagnosis						
IgA nephropathy	397 (44.2)	674 (37.1)	112 (34.9)	156 (43.6)	117 (38.9)	390 (39.0)
Diabetic nephropathy	70 (7.8)	224 (12.3)	37 (11.5)	64 (17.9)	17 (5.6)	140 (14.0)
Membranous nephropathy	158 (17.6)	335 (18.4)	57 (17.8)	101 (28.2)	54 (17.9)	248 (24.8)
Arterionephrosclerosis	107 (11.9)	282 (15.5)	61 (19.0)	10 (2.8)	29 (9.6)	203 (20.3)
MCD/FSGS	77 (8.6)	173 (9.5)	32 (10.0)	28 (7.8)	28 (9.3)	127 (12.7)
High PSG	66 (7.3)	140 (7.7)	21 (6.5)	8 (2.2)	21 (7.0)	66 (6.6)

Data are *n* (%), mean \pm SD or median [IQR]. In the high PSG group, the percentage of sclerotic glomeruli was $\geq 75\%$. Early CKD, eGFR ≥ 60 mL/min/1.73 m² with albuminuria; moderate CKD, eGFR 30–59 mL/min/1.73 m²; advanced CKD, eGFR < 30 mL/min/1.73 m².

BP blood pressure, BMI body mass index, CKD chronic kidney disease; eGFR estimated glomerular filtration rate, MCD/FSGS idiopathic minimal change disease and focal segmental glomerulosclerosis; patients with two or more types of pathological diagnoses were counted separately for each type.

difference was not statistically significant after applying the Bonferroni correction. The calibration plots and Hosmer-Lemeshow goodness-of-fit tests indicated a good agreement between the observed probabilities and those predicted by both the clinical data-only and hybrid models, with the exception of the hybrid model for ANS (Supplementary Fig. 3). The continuous net reclassification indexes (NRIs) were 0.491–1.134 and the integrated discrimination improvements (IDIs) were 0.029 to 0.063, all of which were statistically significant. The AUCs, continuous NRIs, and IDIs of the KIDS for making pathological diagnoses in the image-only, clinical data-only, and hybrid models are shown in Fig. 3 and Table 2.

To validate the findings in the retrospective cohort, the KIDS was tested on a prospective cohort dataset, and it achieved AUCs of 0.891–0.967 for IgAN, MN, DN, ANS, and MCD/FSGS in the hybrid model. It also achieved similar performances of 0.953 (95% CI: 0.882, 1.000) for the DN and 0.846 (95% CI: 0.754, 0.939) for the ANS in image-only mode. In the external test dataset, the KIDS achieved AUCs of 0.867–0.936 for all 5 pathological types in the hybrid model, and it also achieved similar performances of 0.879 (95% CI: 0.824, 0.933) for

DN and 0.711 (95% CI: 0.593, 0.829) for MCD/FSGS in the image-only model. The models' performance in the external test for each pathological type of CKD was similar to that of the internal test dataset. In addition, adding retinal images significantly improved the performance of the hybrid model. The continuous NRIs and IDIs on the prospective test and external test were similar to those on the internal test (Table 2).

To evaluate the model's diagnostic capability across diverse populations, we conducted further validation using multi-ethnic CKD patient datasets from Kashi, Taiyuan, and Xi'an. The model achieved AUCs ranging from 0.790 to 0.932 for IgAN, MN, DN, ANS, and MCD/FSGS (Supplementary Fig. 4). Given the underdeveloped economic conditions and healthcare infrastructure in Somalia (one of the least developed regions globally), renal biopsies are not feasible. However, accurate pathological diagnosis is essential for treatment decisions in CKD patients. We conducted an external validation of the simplified model's diagnostic performance using patient datasets from Somalia. The diagnoses were determined by a panel of senior nephrologists and nephropathologists from China, who served as the reference standard.

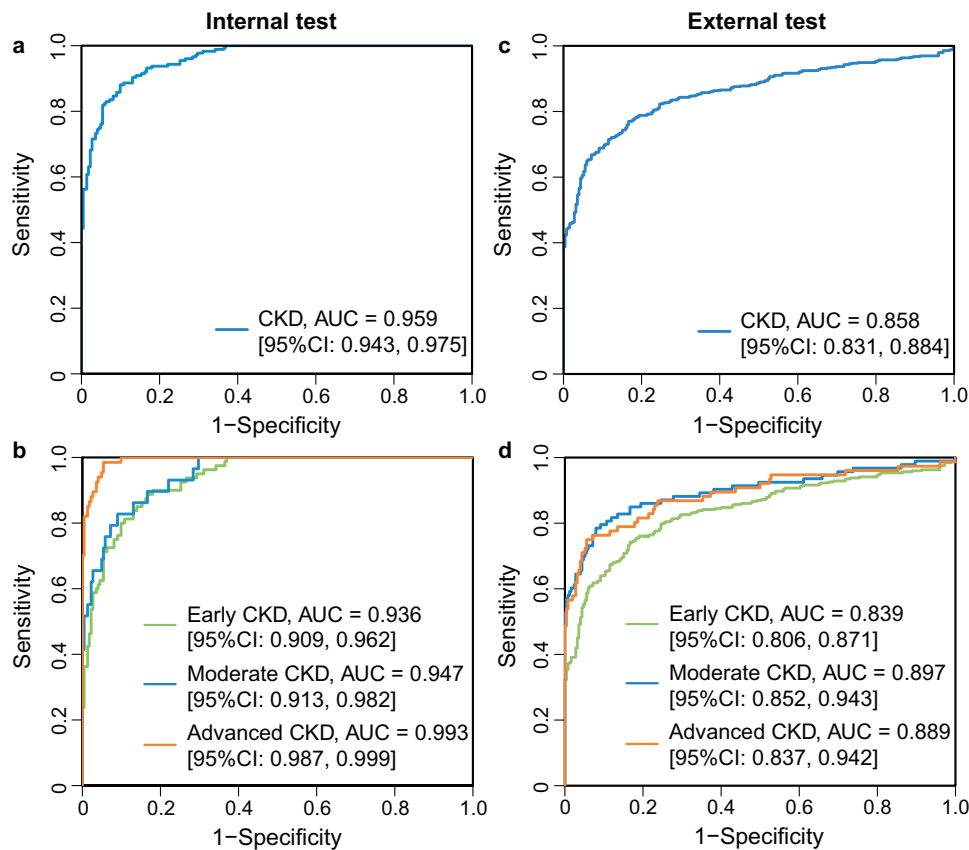


Fig. 2 | Performance of the CKD screening AI model using retinal images. ROC curves showing the model's performance in detecting CKD (a, c) and in classifying CKD stages (early, moderate, and advanced) (b, d) using retinal images alone. Results are shown for the internal test set (a, b) and the external test set (c, d). CKD,

chronic kidney disease; eGFR, estimated glomerular filtration rate. Early CKD, eGFR ≥ 60 mL/min/1.73 m²; moderate CKD, eGFR 30–59 mL/min/1.73 m²; advanced CKD, eGFR <30 mL/min/1.73 m². AUC area under the curve, CI confidence interval. Source data are provided as a Source Data file.

The simplified model achieved AUCs between 0.630 and 0.870 in the Somalia dataset (Supplementary Fig. 5). The models exhibited satisfactory accuracy across diverse populations in China, as well as Somali patients.

Pathological staging AI model. We developed an AI model for the pathological staging of CKD patients for the KIDS, and its AUC reached 0.901 (95% CI: 0.847, 0.955) in the internal test dataset for identifying patients with high PSG without renal biopsy. When evaluated via prospective and external tests, the KIDS achieved similar AUCs of 0.883 (95% CI: 0.813, 0.954) and 0.919 (95% CI: 0.881, 0.958), respectively (Supplementary Fig. 6). In the real-world test datasets, the KIDS system achieved AUCs ranging from 0.901 to 0.910 (Supplementary Fig. 4c, d), reflecting its robustness across diverse populations. The levels of creatinine, blood urea nitrogen, and hemoglobin were found to be crucial for the construction of this model (Supplementary Fig. 7).

Progression prediction AI model

In our retrospective cohort, 999 patients were followed for kidney outcomes over the period (median, 44 months; IQR, 22–74 months; range, 0.6–156 months) after biopsy, and 272 participants developed renal endpoints (Fig. 4a and Supplementary Figs. 8 and 9). A Cox proportional hazards (CPH) regression model was built for the KIDS to predict the progression of CKD with confirmed pathological types and PSG classification. The KIDS achieved an average C-index of 0.811 (95% CI: 0.800, 0.822), and the AUCs of the time-dependent receiver operating characteristic (ROC) curves were 0.854 (95% CI: 0.810, 0.897) at 1 year, 0.845 (95% CI: 0.811, 0.879) at 3 years and 0.821 (95% CI: 0.786, 0.857) at 5 years (Fig. 4c). The Kaplan–Meier curves for the

risk stratification from the model are illustrated in Supplementary Fig. 10. As shown, the KIDS had good discrimination ability in stratifying patients into low-, medium- and high-risk (risk score $<Q1$, $Q1$ – $Q3$, $>Q3$) subgroups ($p < 0.0001$).

Moreover, to assess the clinical importance of retinal image data in predicting CKD prognosis, we employed the probability of pathological classification and the probability of a high rate ($\geq 75\%$) of glomerulosclerosis based on retinal images as a substitute for the renal biopsy results to construct an alternative risk prediction model for the KIDS. Risk stratification extracted from the KIDS resulted in significant separations of the low-, medium- and high-risk groups ($p < 0.0001$), and the Kaplan–Meier curves are illustrated in Fig. 4b. The CPH models from the KIDS achieved an average C-index of 0.804 (95% CI: 0.783, 0.824) and AUCs of 0.852 (95% CI: 0.811, 0.892) at 1 year, 0.833 (95% CI: 0.799, 0.866) at 3 years and 0.809 (95% CI: 0.772, 0.846) at 5 years for the time-dependent ROC curves (Fig. 4d).

The performance of the KIDS and nephrologists in the diagnosis of different types of CKD

We used a multicenter prospective dataset labeled with renal biopsy results for a human–machine comparison to preliminarily assess the differences in pathological diagnostic ability between the KIDS and nephrologists. The demographics and clinical information of the patients with different pathological types are listed in Supplementary Table 5. The test dataset was labeled by nine nephrologists with varying levels of clinical experience in China and three nephrologists in Somalia. The performance of the KIDS was subsequently compared with that of nephrologists via ROC curve plots. As shown in Fig. 5C, the plots revealed that the KIDS achieved greater sensitivity than did all the

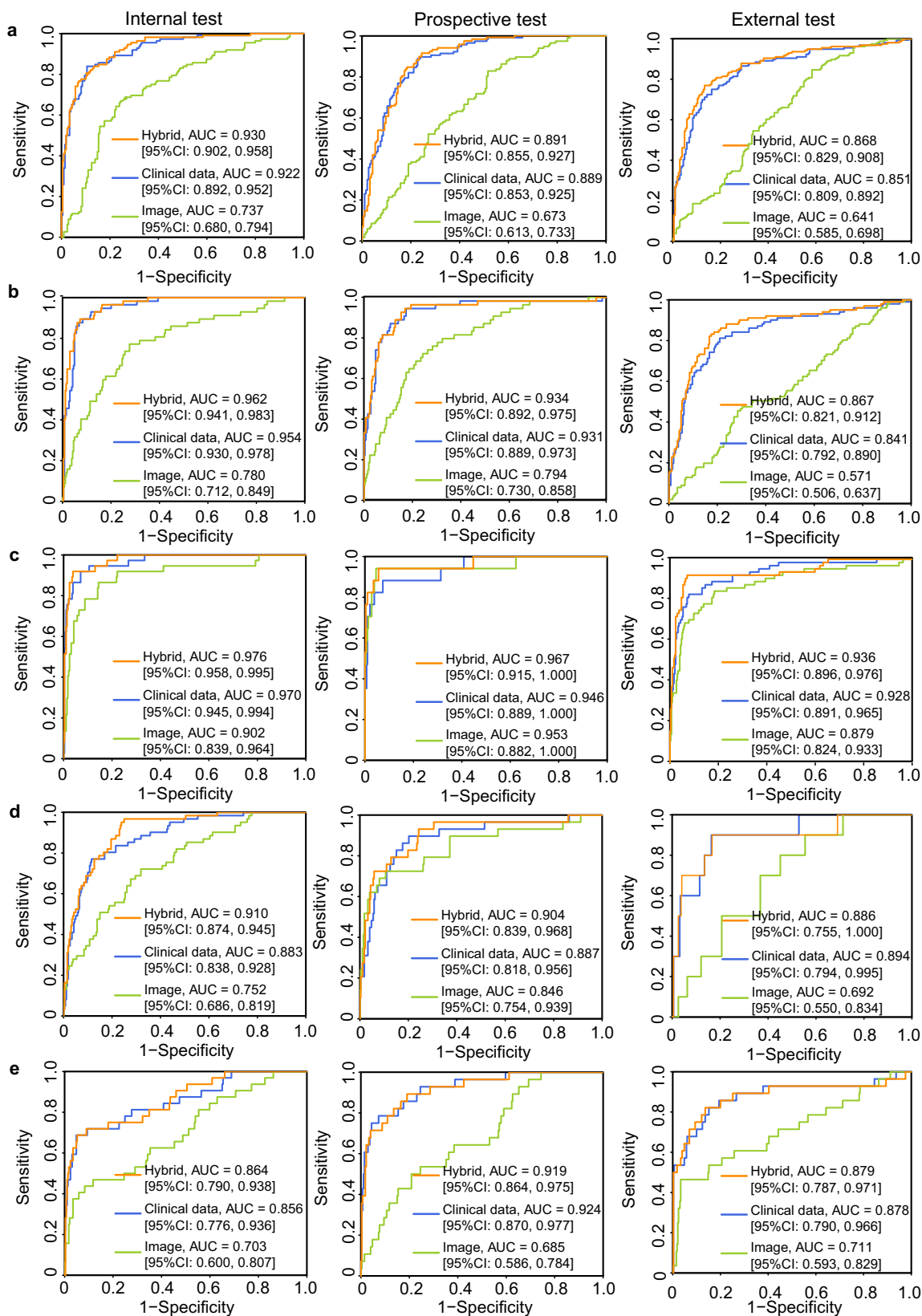


Fig. 3 | Performance of the pathological diagnosis AI models in image-only, clinical data-only, and hybrid models. ROC curves for the performance of pathological diagnosis of IgAN, MN, DN, ANS, and MCD/FSGS (a–e) in the internal test set, prospective test set, and external test set. DN diabetic nephropathy, IgAN

IgA nephropathy, MN idiopathic membranous nephropathy, ANS arterionephrosclerosis, MCD/FSGS idiopathic minimal change disease and focal segmental glomerulosclerosis, AUC area under the curve, CI confidence interval. Source data are provided as a Source Data file.

Table 2 | Net reclassification improvement and integrated discrimination improvement by incorporating retinal images into hybrid models of pathological diagnosis AI models via internal, prospective, and external tests

	NRI		IDI	
	Internal test	Prospective test	Internal test	External test
DN	1.134 (0.808, 1.424)	1.358 (0.955, 1.704)	0.056 (0.002, 0.111)	0.085 (0.046, 0.125)
IgAN	0.491 (0.258, 0.706)	0.204 (-0.032, 0.423)	0.029 (0.011, 0.046)	0.061 (0.037, 0.086)
MN	1.013 (0.754, 1.245)	0.969 (0.708, 1.204)	0.063 (0.040, 0.086)	0.064 (0.042, 0.086)
ANS	0.594 (0.319, 0.888)	0.609 (0.253, 0.971)	0.050 (0.026, 0.074)	0.023 (-0.051, 0.097)
MCD/FSGS	0.599 (0.227, 0.952)	0.740 (0.360, 1.140)	0.048 (0.011, 0.085)	0.059 (0.010, 0.109)

The NRI assessed in this study is referred to as the continuous NRI without predefined thresholds. The IDI and NRI estimates are given with 95% confidence intervals. Source data are provided as a Source Data file. NRI net reclassification improvement, IDI integrated discrimination improvement, DN diabetic nephropathy, IgAN IgA nephropathy, MN idiopathic membranous nephropathy, ANS arterionephrosclerosis, MCD/FSGS idiopathic minimal change disease and focal segmental glomerulosclerosis.

other nephrologists, except for the diagnosis of ANS, for which the sensitivities of the two experts were slightly greater than those of the KIDS. In addition to accuracy, the KIDS also exhibited a greater advantage in terms of stability.

Model visualization and explanation

Saliency maps were used to enhance the interpretability of the AI models by highlighting key areas in retinal images. These changes were related to abnormal exudation, vascular rigidity, and a greater cup-to-disc ratio in CKD patients (Supplementary Fig. 11a–c). Distinctive features of various pathological types have been identified, including the presence of microaneurysms and exudation in the DN, vascular narrowing and papilledema in the ANS, and increased retinal drusen in IgAN. Conversely, the retinopathy observed in MN and MCD/FSGS was relatively mild (Supplementary Fig. 11d–f), which suggested that the signals utilized for prediction may be correlated with the severity of vascular damage associated with CKD of different etiologies. Clinicians and researchers are expected to gain a better understanding of the underlying mechanisms and potential treatments for CKD-related retinal pathology.

A detailed explanation of feature selection and interpretation of the CPH model used in predicting CKD progression is displayed in Supplementary Fig. 12. On the basis of the CPH model, a nomogram was developed as an intuitive and visual method for the individualized prediction of CKD progression in patients without renal biopsy data for clinical use. An example of the nomogram using the predicted probability of pathological type and high PSG is shown in Supplementary Figs. 13 and 14.

Discussion

In this study, the KIDS was developed to implement comprehensive process management in CKD diagnosis and treatment decision-making, including (1) detecting CKD in populations, (2) conducting pathological diagnosis and staging in identified CKD patients, and (3) predicting disease prognosis on the basis of the pathological diagnosis. When a participant arrives at a primary care setting, the KIDS can identify potential CKD cases from retinal images and recommend referrals for suspicious cases to specialized clinics, enabling early detection and timely intervention for CKD patients. In specialist nephrology clinics, the KIDS is capable of providing objective pathological diagnosis and predicting adverse outcomes without the demand of renal biopsy, assisting nephrologists in precise treatment decision-making and personalized management (Fig. 5A). The KIDS is expected to play a significant role in improving kidney care and preventing kidney failure, especially in less developed regions with limited medical resources.

Current methods used to evaluate renal function and detect CKD, such as clinical information, imaging techniques (ultrasound, computed tomography, and magnetic resonance imaging)^{30,31}, and routine laboratory tests (blood and urine tests)^{11,32–34}, may not be specific enough to differentiate pathological types of CKD. Renal biopsy remains the gold standard for pathological diagnosis, and studies have suggested that prebiopsy and postbiopsy diagnoses differ in up to one-third of patients, highlighting the utility and significance of renal biopsy for patients with unclear clinical conclusions³⁵. Although early biopsy-guided initiation of therapy can prevent complications of decline in kidney function and mitigate disease progression⁸, the limited availability of specialists, strict indications, and accompanying complications hinder accessibility to renal biopsy^{8–11,36–38}. Percutaneous kidney biopsy procedures are not routinely performed, particularly in low- and middle-income countries (LMICs) and regions, and they may be associated with high complication rates (reported to be as high as 52.6% in these regions) and a lack of nephrologists³⁹. In particular, renal biopsy is rarely an option in the least developed regions, such as Somalia. Therefore, a noninvasive

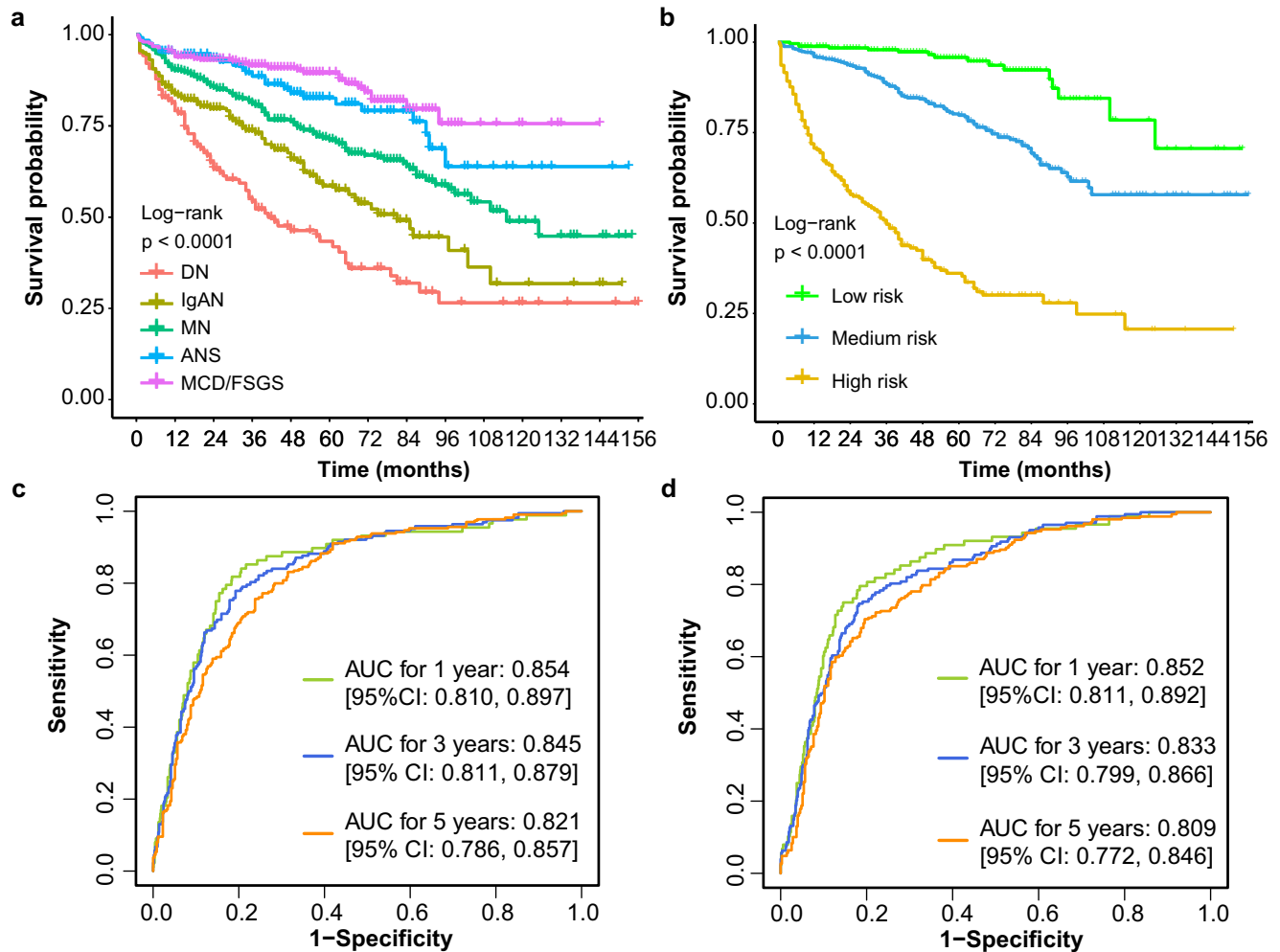


Fig. 4 | Progression of CKD in different pathological types and performance of Cox proportional hazards regression models in progression prediction.

a Kaplan–Meier curves for the progression of CKD according to different pathological types. IgAN IgA nephropathy, MN idiopathic membranous nephropathy, DN diabetic nephropathy, ANS arterionephrosclerosis, MCD/FSGS idiopathic minimal change disease and focal segmental glomerulosclerosis. **b** Kaplan–Meier curves for risk stratification from the Cox proportional hazards regression model with predicted pathological types from the image-only AI model. Survival curves represent

the high-risk, medium-risk, and low-risk subgroups (risk score $<Q1$, $Q1$ – $Q3$, $>Q3$), and 95% CI regions are represented as shaded areas around the curve. ROC curves showing the performance of Cox proportional hazards regression models at 1, 3, and 5 years with 10-fold cross-validation using **c** pathological types from renal biopsy and **d** predicted pathological types from the image-only AI model. The P value was calculated via the log-rank test among all groups. AUC area under the curve, CI confidence interval. Source data are provided as a Source Data file.

approach with good compliance that can provide a pathological diagnosis of CKD and guide clinical treatment will be the state-of-the-art in CKD management, and will potentially improve the basic quality of kidney care in LMICs. Retinal photography is the most common examination in ophthalmology, and has the advantages of convenience, affordability, and noninvasiveness. In the U.S., the cost of conventional CKD screening is approximately \$80, approximately five times greater than that of retinal photography. This indicates that CKD screening using retinal images is more cost-effective and pragmatic^{40,41}, making it suitable for CKD screening and diagnosis in LMICs.

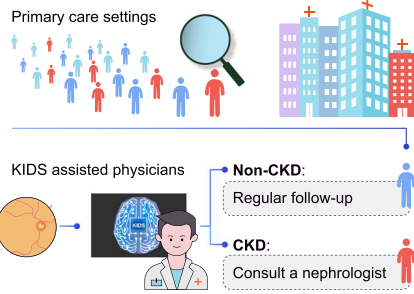
Advances in CKD screening using DL algorithms based on fundus images have been achieved previously^{27,28}. Similarly, we have also demonstrated that AI models employing retinal images serve as attractive tools for CKD screening and provide high accuracy. More importantly, the major strength of this study is the ability of the AI model to provide both pathological diagnosis and prediction of CKD based on retinal images. The KIDS can identify the five most common pathological types of CKD from retinal images. Among these, the identification of DN and ANS achieves higher AUCs, which is likely

attributable to the obvious microangiopathy in the retina caused by hyperglycemia or hypertension^{42–44}. When clinical data such as routine blood, urine, and kidney ultrasound data were combined, the AUCs of the hybrid models reached 0.864–0.976. To assess the effect of adding retinal images to hybrid models, we quantified the improvement of model performance via the NRI^{45,46}. We found that the use of retinal images in addition to clinical data markedly enhanced the pathological prediction (NRI: 0.491–1.134). Although NRI has faced concerns over poorly fitted models, NRI can still be used when the models are well-calibrated. In this study, our analysis, including calibration plots and Hosmer–Lemeshow goodness-of-fit tests, demonstrates that the models exhibit strong calibration. Herein, the use of NRI is appropriate in this context, as it provides valuable insight into model reclassification after calibration.

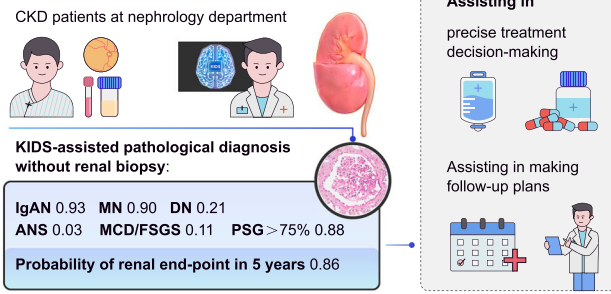
Furthermore, the model’s robustness and generalization were validated across multicenter and multi-ethnic external datasets, showing consistent and satisfactory performance across diverse populations. Notably, although the Shanxi dataset lacked retinal images, the clinical data-only model was used for pathological diagnosis and demonstrated strong performance with excellent AUC values

A) The KIDS can be applied in the following clinical scenarios

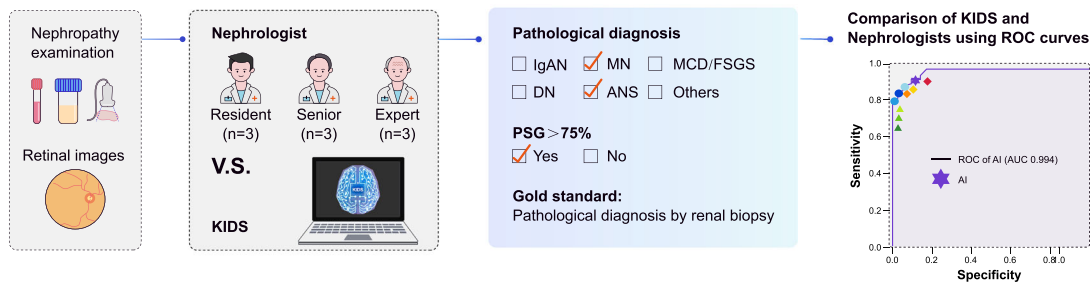
1 Scenarios 1: CKD Screening



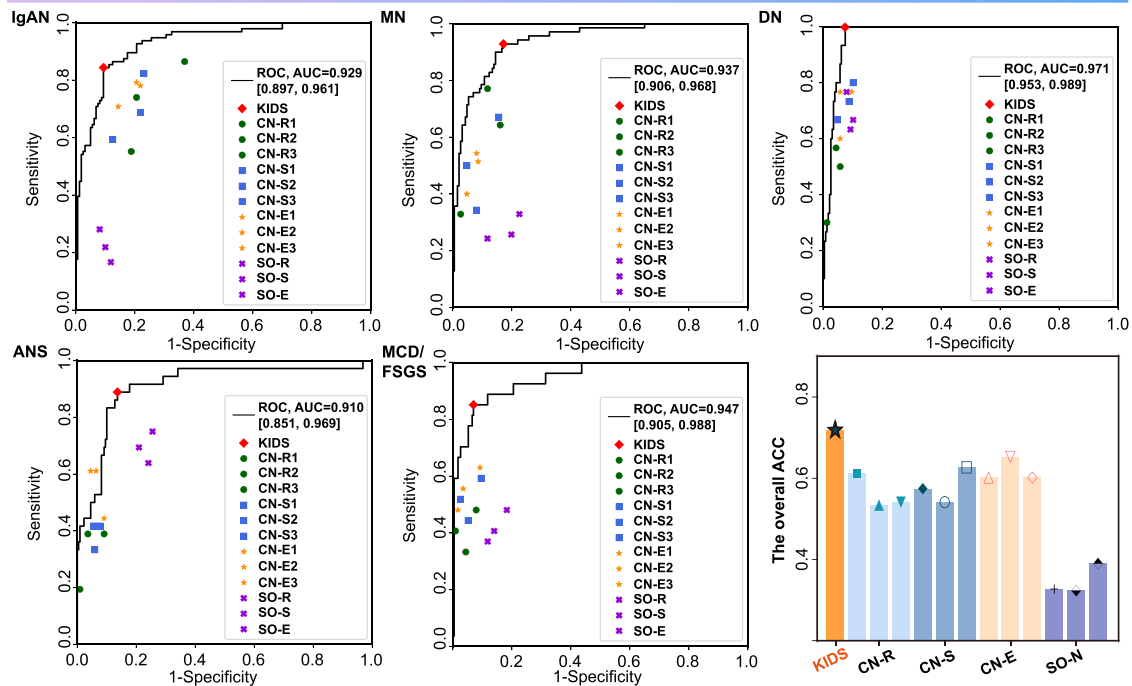
2 Scenarios 2: Pathological Diagnosis



B) Comparison between the KIDS and clinical nephrologists in a prospective validation



C) Performances of the KIDS and nephrologists in diagnosis of different CKD



(0.790–0.932). This indicates the advantages of the KIDS system in accommodating different data types, ensuring that even in the absence of retinal images, accurate diagnostic predictions can still be achieved using clinical data alone. In Somalia, one of the least undeveloped countries, kidney care is extremely limited, making it challenging to perform a renal biopsy and to locate standardized and

comprehensive medical records. The results of human-machine comparisons also highlighted the poor diagnostic capabilities of nephrologists in Somalia, underscoring the urgent need to provide assistance in pathological diagnosis and improve the level of kidney care. Herein, we simplified the diagnostic model to fit the clinical data available in Somalia. This simplification resulted in AUCs ranging from 0.630 to

Fig. 5 | Description of the clinical application scenarios for the KIDS and human and artificial intelligence comparison studies. **A** The KIDS system can be applied in the following two clinical scenarios: scenario 1 involves CKD screening based on retinal images. When a participant arrives at a primary care setting, the KIDS can identify potential CKD cases from retinal images and recommend referrals for suspicious cases to specialized clinics, enabling early detection and timely intervention for CKD patients. Scenario 2: Noninvasive pathological diagnosis. In specialist nephrology clinics, the KIDS is expected to provide objective pathological diagnosis and predict adverse outcomes without the need for invasive renal biopsy. This information can assist nephrologists in making precise treatment decisions and managing patients with personalized care. **B** A human and artificial intelligence comparison study was performed to evaluate the performance of the KIDS

compared with that of nephrologists in diagnosing pathology via ROC curves. **C** ROC curve plots revealed that the KIDS achieved greater sensitivity than did all the other nephrologists, except for the diagnosis of ANS, for which the sensitivities of the two experts were slightly greater than those of the KIDS. In terms of accuracy, KIDS also exhibited a greater and more stable advantage. KIDS kidney intelligent diagnosis system, ROC receiver operating characteristic, IgAN IgA nephropathy, MN idiopathic membranous nephropathy, DN diabetic nephropathy, ANS arterionephrosclerosis, MCD/FSGS idiopathic minimal change disease and focal segmental glomerulosclerosis. CN-R Chinese resident nephrologist, CN-S Chinese senior nephrologist, CN-E Chinese expert nephrologist, SO-R Somali resident nephrologist, SO-S Somali senior nephrologist, SO-E Somali expert nephrologist, SO-N Somali nephrologists. Source data are provided as a Source Data file.

0.870, demonstrating the model's effectiveness in providing accurate diagnosis despite limited existing data. These results highlight the model's generalizability and accessibility, showcasing its great potential advantages for basic-level hospitals that lack access to renal biopsies, especially in undeveloped regions. In addition, PSG, which is based on the creatinine value and varies with metabolism, medications, and calculation methods, was used as an estimate for chronic kidney injury, revealing features that are more objective and stable than the eGFR^{47,48}. Therefore, we also developed a clinical data-based pathological staging model in the KIDS to identify patients with high PSG prior to renal biopsy. It can help nephrologists decide whether a renal biopsy is needed for a CKD patient, thus preventing patients with severe sclerosis from sustaining undesirable damage and financial burden.

Accurate progression prediction plays an important role in facilitating specific management and improving the outcomes of CKD patients^{49,50}. Most previous AI models have provided prognostic predictions for only one pathological type of CKD, including the time-to-ESRD or high creatinine value within 5 years^{51–53}. Additionally, these models are always targeted at patients with proven biopsies. In this study, we confirmed that the prognoses of CKD patients are related to their pathological types and high PSG values ($\geq 75\%$). Moreover, we developed the KIDS to predict the time point of deterioration of renal function and the outcome events within 1, 3, and 5 years, including a 50% decrease in the eGFR, renal transplantation, and the initiation of dialysis.

The KIDS exhibited robust performance and generalizability in the prospective dataset, accurately identifying pathological types and even outperforming experienced nephrologists. Considering the lack of renal biopsy in the least developed regions, the KIDS can serve as an alternative diagnostic method and is expected to significantly improve the level of kidney care. Our system has three unique advantages. First, on the basis of existing renal biopsy data, Kaplan–Meier curves and a CPH regression model were established in the KIDS, which can be used for risk stratification and prognosis prediction of the five pathological types of CKD, and show good performance. Second, in the absence of a renal biopsy, the KIDS can take advantage of the probability of pathological diagnosis for prognosis prediction. Third, the KIDS is simple and easy to use, exhibiting good accessibility. All the input variables used in the prediction models are from routine tests in clinical practice. Additionally, a nomogram was drawn for the individualized prediction of CKD progression in patients, which is simple and effective for physicians.

Our study has several limitations that we aim to address in future research. First, since the pathological diagnosis models were trained using the data from CKD patients who underwent renal biopsy to ensure the accurate and effective prediction of pathological types of CKD, the sample size was limited because of the invasiveness and accessibility constraints of renal biopsy. Nevertheless, our study included a greater cohort of participants than did an analogous study involving renal biopsy, where the number of participants usually remained below 2000⁵⁴. Regarding the quality of data collection, the

retinal images collected in external tests have insufficient brightness, clarity, and integrity, and the differences observed in retinal images acquired from various instruments can markedly impede the model's performance. To solve these issues, we have enhanced the ability of KIDS to process retinal images from different instruments. Through validation with retinal images from different instruments at external test centers, the models showed good generalizability. Finally, because our study cohort included Chinese and Somali patients, the generalizability in a wider range of countries and ethnicities will need to be validated. To enhance the clinical utility of the KIDS system across more regions and populations, we have implemented its deployment on a cloud platform (<https://zockids.gzzoc.com/modelstore/#/login>), enabling broader accessibility and application. Model training on more diverse clinical and demographic cohorts may further improve the diagnostic accuracy and generalizability for its clinical use. Additionally, we also aim to expand the model's capability to predict a wider range of pathological types. Future studies will also focus on the histological characteristics of CKD to examine their prognostic relevance.

In conclusion, the KIDS is not only suitable for CKD screening of the population, but it also provides accurate diagnostic and prognostic information for individual patients without the requirement for invasive operations. The accuracy and stability of our model in pathological diagnoses are better than those of human doctors, which is extremely beneficial for the CKD patients who cannot receive a renal biopsy. In addition, due to the ease of access to retinal images or clinical data in clinical practice, this noninvasive model shows great potential for adoption for the comprehensive management of CKD and improving kidney care.

Methods

This was a multicenter, retrospective cohort study, and the KIDS was developed and validated utilizing retinal images and clinical data from participants across 8 hospitals in China and 1 hospital in Somalia. The dataset collected from FAH and ZOC was used for model development and internal, prospective testing. External testing was performed using datasets from ZPH, FPH, and AHY. Additionally, small-scale real-world validation was conducted using multicenter and multi-ethnic datasets from FPHK, SPTCMI, SAH, and BH. The study was registered at ClinicalTrials.gov (identifier: NCT05223712) and approved by the Institutional Review Board/Ethics Committee of ZOC (2021KYPJ167), the ICE for Clinical Research and Animal Trials of FAH (2022097), the Ethics Committee for Clinical Research and Experimental Animals of ZPH (K2022-162-2), the Medical Ethics Committee of FPH (2022-90), the Medical Ethics Committee of AHY (YYFY-LL-2022-106), the Medical Ethics Committee of SPTCMI (2025Y-08001), the Medical Ethics Committee of FPHK (2024-69), the Medical Ethics Committee of SAH (2024-256), and the Research Committee of BH (BH/IRVD2024007/001). The study was conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from participants prospectively enrolled at the FAH for model prospective testing. For all other cohorts, informed consent was waived by the respective

institutional review boards due to the retrospective design and use of de-identified data. All ethics committees listed above reviewed and approved the study protocol and consent procedures. (Supplementary Fig. 1).

Diagnostic criteria

According to the international guidelines of the Kidney Disease: Improving Global Outcomes (KDIGO), CKD was defined as either an eGFR <60 mL/min/1.73 m² or kidney damage (albuminuria) for 3 months or longer⁵. Non-CKD controls were defined as individuals with an eGFR ≥ 60 mL/min/1.73 m², absence of albuminuria, and no documented history of CKD. CKD was categorized into early CKD (eGFR ≥ 60 mL/min/1.73 m² with albuminuria), moderate CKD (eGFR 30–59 mL/min/1.73 m²), and advanced CKD (eGFR < 30 mL/min/1.73 m²)⁵⁵.

The pathological diagnosis was established through collaborative assessments by nephrology and pathology specialists, based on renal biopsy results (Supplementary Methods). We developed prediction models for identifying the five most common types of CKD, whose distribution is consistent with epidemiological surveys of the biopsy-proven spectrum of kidney diseases in China^{56,57}: (1) IgAN: IgA nephropathy; (2) MN: idiopathic membranous nephropathy; (3) ANS: arterionephrosclerosis; (4) DN: diabetic nephropathy; and (5) MCD/FSGS: idiopathic minimal change disease (MCD) and focal segmental glomerulosclerosis (FSGS). MCD and FSGS are grouped as podocytopathies owing to their shared pathological characteristics and mechanisms involving podocytes⁵⁸. The definitions and management principles of these 5 pathological classifications are described in the Supplementary Material. The PSG is a crucial prognostic marker for long-term implications that is used for assessing the severity of kidney diseases. In the pathological staging model, glomerular lesions with $\geq 75\%$ sclerotic glomeruli are classified as having high PSG, indicating substantial glomerular damage and fibrosis⁵⁹.

The composite renal endpoint event was defined as a renal outcome if it met any of the following criteria: I. deterioration in kidney function, defined as a $\geq 50\%$ decline in eGFR from baseline (as measured by serum creatinine at admission and follow-up); II. development of end-stage renal disease (eGFR < 15 mL/min/1.73 m²); and III. initiation of renal replacement therapy (dialysis or kidney transplantation)⁶⁰. Kidney survival was calculated after diagnostic renal biopsy to renal endpoints.

Clinical and image datasets

Dataset for the CKD screening AI model. For CKD detection, we used retinal images of CKD patients and non-CKD participants from retrospective datasets. The development and internal test dataset consisted of non-CKD participants who had annual health examinations, including routine systemic (albuminuria and creatinine tests) and ophthalmic tests at the Health Examination Center, and CKD patients from the Department of Nephrology of FAH between January 3, 2009, and February 28, 2023. The external test set included participants who underwent an annual health check at ZOC and outpatients at ZPH, FPH, and AHY from December 23, 2016, to March 13, 2024.

Dataset for the pathological diagnosis and staging AI model. Patients who met the diagnostic criteria for CKD and underwent both ophthalmological examinations with retinal imaging and kidney specialist examinations with renal biopsy during the same hospitalization were included in the study. The development and internal test datasets were collected from patients of the Department of Nephrology of FAH between April 28, 2009, and August 8, 2022. Validation datasets were provided by a prospective cohort collected in the Department of Nephrology of FAH from August 8, 2022, to April 27, 2023.

For external testing, we obtained 3 independent retrospectively collected datasets: (1) Zhongshan City People's Hospital [ZPH,

Zhongshan City, China]: patients from December 23, 2016, to November 7, 2022; (2) First People's Hospital of Foshan [FPH, Foshan City, China]: patients from January 13, 2021, to February 28, 2023; and (3) Affiliated Hospital of Youjiang Medical University for Nationalities [AHY, Youjiang City, China]: patients from May 7, 2022, to February 3, 2023. These hospitals are located in three cities across two provinces in southern China, each representing different economic and medical profiles. All these datasets consisted of retinal images and clinical data, including blood and urine test results, kidney ultrasound scans (as shown in Supplementary Table 2), pathological diagnosis, and PSGs derived from renal biopsies.

For real-world validation across multinational and multi-ethnic settings, we collected four datasets: (1) First People's Hospital of Kashi [FPHK, Kashi, China], spanning December 14, 2020, to November 10, 2024; (2) Shanxi Provincial Traditional Chinese Medicine Institute [SPTCMI, Taiyuan, China], from December 26, 2018, to September 29, 2024; (3) The Second Affiliated Hospital of Xi'an Jiaotong University [SAH, Xi'an, China], covering November 11, 2020, to December 21, 2023; and (4) Banadir Hospital [BH, Mogadishu, Somalia], from July 20, 2023, to November 18, 2024. Due to constraints in local medical resources, the datasets from SPTCMI and BH comprised only clinical data.

Dataset for progression prediction models. For the CKD progression prediction AI model, CKD patients (eGFR ≥ 15 mL/min/1.73 m² and without dialysis) before the kidney failure stage from the retrospective cohort were included. Administrative and follow-up data from the retrospective cohort were collected for development and validation. Participants were monitored for disease outcomes, including renal dialysis, kidney transplantation, hospitalization, mortality, and cause of death, from May 12, 2009, to June 22, 2022. Follow-up information regarding the participants' examinations after the initial assessment was obtained through electronic health records, phone interviews, or clinical visits. The drop-out time and reason for the missed visits were recorded. Participants with preexisting renal dialysis or kidney transplantation at the time of administration were excluded from the cohort analysis (Supplementary Methods).

Image acquisition and data quality control

The retinal images of all the datasets consisted of one macula-centered fundus image per eye with a 45° primary field of view. For the datasets collected from FAH, the retinal images were captured with CR-2 AF (Canon) and RetiCam 3100 (SYSEYE). For the external testing dataset, retinal images from ZPH, FPH, and AHY were obtained via KOWA Nonmyd 7 (Kowa), AFC-330 (NIDEK), and CR-2 AF (Canon), respectively. The retinal images included in the analysis met the following criteria: integrity of the macula and optic disc, absence of artifacts, and sufficient resolution. For each subject, one retinal image per eye that met the specified criteria was selected. For subjects with only one eligible eye, the best image from that eye was retained. Out of a total of 15,127 images collected, we excluded 1983 (13.1%) ineligible fundus images.

Model development

CKD screening AI model. The model was developed based on retinal images. The images were assigned randomly to training, validation, and test sets at a ratio of 8:1:1 at the participant level, and there were no samples overlapping in the three datasets. In this study, the model's input comprised retinal images from both the right and left eyes, and the images were processed separately. Predictions were made with the AI model at the image level, and then the image-level outputs were averaged at the participant level as a final prediction for each participant. For participants whose only one eye image was eligible, the retinal image was processed, and its model output was taken as participant-level prediction. To further evaluate the performance of the CKD screening DL model when applied to populations with

different prevalences of CKD, we conducted tests based on the internal test and external test datasets, with simulated prevalences of 5%, 10%, and 20%. We created 1000 simulated datasets by randomly sampling $1000 \times (1 - \text{prevalence})$ non-CKD and $1000 \times \text{prevalence}$ CKD samples from each test dataset with replacement for each prevalence.

A noninvasive model for pathological diagnosis. To identify the presence of the five categories of pathological diagnosis, we trained a noninvasive model consisting of three submodels to perform five separate binary classification tasks: (1) a retinal image model; (2) a clinical data model; and (3) a hybrid model. In the retinal image-only model, the input comprised retinal images, specifically one macular-centered image per eye, and the output was the probabilities of the presence of the five pathological categories. The clinical data model was first built via a multilayer perceptron (MLP; 56, 256, 256, 5) with full features. Then, the importance of the features was measured via the permutation feature importance method on the validation set, and features with 95% lower confidence limits >0 were selected. Finally, the MLP (35, 256, 256, 5) model with the selected features and the highest accuracy on the validation set was constructed as the final meta-model. In addition, we developed a simplified version of the meta-model for application based on tests and ultrasounds that may be lacking in the local medical conditions. For the hybrid model, we developed an EfficientNet-B5-based multimodal model by integrating retinal images and clinical data as inputs. The average of the probabilities of each category for both eyes from a patient was calculated as the probability of the presence of the category at the participant level. (Supplementary Fig. 2).

Pathological staging model. The pathological staging model was built via XGBoost with features selected via least absolute shrinkage and selection operator with 10-fold cross-validation. The datasets used were the same as those used in the aforementioned clinical data model, which requires 6 clinical features as inputs: hemoglobin, creatinine, uric acid, and renal ultrasound measurements (left renal length; whether the demarcation of the cortex is clear; and whether the medulla and hyperechoic medulla are greater). The output was the probability of a PSG score $\geq 75\%$.

Progression prediction models. The pathological types confirmed by renal biopsy, combined with clinical features, were used to develop models for predicting the renal endpoints. We also used the predicted probabilities of pathological classification from retinal images and the predicted probability of high glomerulosclerosis ($\geq 75\%$) results instead of renal biopsy results to develop models with the same methods. CPH regression and 10-fold cross-validation were adopted to train and validate the models. For risk stratification, the individual risk scores were calculated with the models. Patients were subsequently categorized into three groups according to upper and lower quartiles: low-risk (risk score $< Q1$), medium-risk ($Q1 - Q3$), and high-risk (risk score $> Q3$) according to a previous study²⁸.

Model visualization and explainability. As a graphical representation of the prognostic prediction model based on Cox regression, the nomogram consists of a set of axes representing the distribution of selected variables, with corresponding points assigned to different levels or values of each variable. The points for each variable are added to obtain a total score, which is then translated into a predicted probability of CKD progression at 1, 3, and 5 years on separate axes. Furthermore, to improve the interpretability of the models, a forest plot was adopted for the Cox regression models, and the importance of the features in the final meta-model was measured via the permutation feature importance method on the validation set. A SmoothGrad saliency map⁶¹ was used to draw heatmaps by highlighting pixels that strongly influenced the prediction of the models for the images in the test datasets.

The details of model development are provided in the Supplementary Methods.

Comparison between the KIDS and clinical nephrologists in a prospective multicenter validation

We designed an AI and clinical nephrologists' comparison study to evaluate the performance of the KIDS compared with that of nephrologists in diagnosing pathological types of CKD and to assess the real-world applicability of the KIDS in a clinical setting. The test set prospectively enrolled 256 CKD patients, none of whom had been used in training and validation before, from multiple centers (FAH, FPH, ZPH, and AHY) with single or multiple pathological diagnoses confirmed by renal biopsy between May 1, 2023, and November 22, 2023. All the cases were subsequently assessed via the KIDS and individually by nine nephrologists at different levels of expertise in China: three at the resident level (X.L. [FPH], Q.H. [ZPH], Y.X. [AHY]), three at the senior level (L.J. [FAH], P.Y. [FPH], S.Y. [Dongguan TCM hospital]), and three at the expert level (N.H. [FAH], H.Y. [FAH], J.Y. [FAH]). Three nephrologists, I.M.A. (resident level), N.I.A. (senior level), and M.H.R.H. (expert level) from Banadir Hospital in Somalia, also conducted pathological diagnoses of CKD patients. Nephrologists were asked to determine the pathological type on the basis of the clinical data and retinal images provided for each case. We used ROC curves and accuracy (ACC) values to compare the pathological diagnostic performance of the KIDS and nephrologists.

$$\text{Accuracy, ACC} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}$$

The ACC for each patient is defined as the proportion of correctly predicted pathological diagnosis labels to the total number of labels (both predicted and actual) for that patient. The ACC for the test dataset is calculated as the average ACC across all 256 patients (n : total number of patients; Y_i : actual label set, $Y_i = (y_1, y_2, \dots, y_k)$, $y_j \in \{0, 1\}$, $1 \leq j \leq k$, k is the number of labels; Z_i : predicted label set, $Z_i = (z_1, z_2, \dots, z_k)$, $z_j \in \{0, 1\}$, $1 \leq j \leq k$)^{62,63}.

Statistical analysis

The performance of the models in CKD screening and pathological diagnosis and staging for binary classification prediction was assessed via ROC curves, in which sensitivity was plotted against 1-specificity at different probability thresholds. The AUCs are reported along with 95% DeLong confidence intervals (CIs). The paired DeLong test, continuous NRI, and IDI were employed to assess the improvement in pathological classification performance by adding retinal imaging information. Bonferroni correction was adopted for multiple AUC comparisons. Calibration plot and Hosmer-Lemeshow goodness-of-fit test were used to assess the consistency between the observed probabilities and the probabilities predicted by the metadata model and hybrid model. Using the optimal operating thresholds determined with the Youden index, the sensitivities and specificities for each type of pathological diagnosis made by the KIDS were measured and visualized via ROC curves to compare the KIDS with those of clinical nephrologists. Kaplan-Meier curves and log-rank tests were used to compare the overall survival probability of CKD patients in different pathological types and groups. The concordance index (C-index) was used to assess the overall predictive accuracy of the CPH model. Time-dependent ROC curves with AUCs and 95% CIs at 1 year, 3 years, and 5 years were also used to measure model performance. All the statistical analyses were performed via Python (version 3.8.0) and R (version 4.1.1).

Reporting summary

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

Data availability

All data supporting the findings described in this manuscript are available in the article and in the Supplementary Information or/and from the corresponding author upon request. Data used to generate the main and supplementary figures are provided in the Source Data file. Due to patient privacy and ethical restrictions, the raw clinical data cannot be shared publicly. In the case of noncommercial use, requests can be directed to the lead corresponding author, H.L. [linht5@mail-sysu.edu.cn], using the full data access form available on Github [<https://github.com/ZOC-iKEA/KIDS>]. Access requests will be reviewed by the institutional data governance committee, and a response will be provided within 30 days. If approved, data will be shared via secure transfer and be available for a period of 12 months, and only for the purposes outlined in the approved application. Source data are provided with this paper.

Code availability

Deep learning models are developed and deployed using the publicly available EfficientNet-b5 architecture, starting from pre-trained models available at <https://github.com/pytorch/vision>. The source code for our model is available on GitHub (<https://github.com/ZOC-iKEA/KIDS>). KIDS models are also deployed online (<https://zockids.gzzoc.com/modelstore/#/login>), and interested participants can contact the lead corresponding author to request an account for a trial.

References

- GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. **395**, 709–733 (2020).
- Bello, A. K. et al. Assessment of global kidney health care status. *JAMA*. **317**, 1864–1881 (2017).
- Baumann, M., Burkhardt, K. & Heemann, U. Microcirculatory marker for the prediction of renal end points: a prospective cohort study in patients with chronic kidney disease stage 2 to 4. *Hypertension* **64**, 338–346 (2014).
- Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 Clinical Practice Guideline for the evaluation and management of chronic kidney disease. *Kidney Int.* **105**, S117–S314 (2024).
- Levey, A. S. & Coresh, J. Chronic kidney disease. *Lancet*. **379**, 165–180 (2012).
- Whittier, W. L. & Korbet, S. M. Timing of complications in percutaneous renal biopsy. *J. Am. Soc. Nephrol.* **15**, 142–147 (2004).
- Corapi, K. M. et al. Bleeding complications of native renal biopsy: a systematic review and meta-analysis. *Am. J. Kidney Dis.* **60**, 62–73 (2012).
- Zhu, M. S., Chen, J. Z. & Xu, A. P. Factors that can minimize bleeding complications after renal biopsy. *Int. Urol. Nephrol.* **46**, 1969–1975 (2014).
- Turk, A. A. A., Estiverne, C., Agrawal, P. R. & Michaud, J. M. Trends and outcomes of the use of percutaneous native renal biopsy in the United States: 5-year data analysis of the Nationwide Inpatient Sample. *Clin. Kidney J.* **11**, 330–336 (2018).
- Bello, A.K. et al. *ISN–Global Kidney Health Atlas: A Report by the International Society of Nephrology: An Assessment of Global Kidney Health Care Status Focusing on Capacity, Availability, Accessibility, Affordability and Outcomes of Kidney Disease*. <http://www.theisn.org/global-atlas> (International Society of Nephrology, 2023).
- Siwy, J. et al. Noninvasive diagnosis of chronic kidney diseases using urinary proteome analysis. *Nephrol. Dial. Transpl.* **32**, 2079–2089 (2017).
- Connor, K. L. & Denby, L. MicroRNAs as non-invasive biomarkers of renal disease. *Nephrol. Dial. Transpl.* **36**, 428–429 (2021).
- Klinkhammer, B. M. et al. Non-invasive molecular imaging of kidney diseases. *Nat. Rev. Nephrol.* **17**, 688–703 (2021).
- Kim, T. et al. Plasma proteins associated with chronic histopathologic lesions on kidney biopsy. *J. Am. Soc. Nephrol.* **35**, 910–922 (2024).
- Kuo, C. C. et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *npj Digit. Med.* **26**, 29 (2019).
- Zhao, D. et al. Current progress in artificial intelligence-assisted medical image analysis for chronic kidney disease: a literature review. *Comput. Struct. Biotechnol. J.* **21**, 3315–3326 (2023).
- Farrar, T. E. et al. The eye, the kidney, and cardiovascular disease: old concepts, better tools, and new horizons. *Kidney Int.* **98**, 323–342 (2020).
- Poll-The, B. T., Maillette de Buy Wenniger-Prick, L. J., Barth, P. G. & Duran, M. The eye as a window to inborn errors of metabolism. *J. Inherit. Metab. Dis.* **26**, 229–244 (2003).
- Arnould, L. et al. Association between the retinal vascular network with Singapore “I” Vessel Assessment (SIVA) software, cardiovascular history and risk factors in the elderly: the Montrachet study, population-based study. *PLoS ONE*. **13**, e0194694 (2018).
- Rim, T. H. et al. Prediction of systemic biomarkers from retinal photographs: development and validation of deep-learning algorithms. *Lancet Digit. Health* **2**, e526–e536 (2020).
- Xiao, W. et al. Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multi-centre study. *Lancet Digit. Health* **3**, e88–e97 (2021).
- Ting, D. S. W. et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* **318**, 2211–2223 (2017).
- Cheung, C. Y. et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat. Biomed. Eng.* **5**, 498–508 (2021).
- Kermany, D. S. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**, 1122–1131 (2018).
- Xu, X. et al. Retinal image measurements and their association with chronic kidney disease in Chinese patients with type 2 diabetes: the NCD study. *Acta Diabetol.* **58**, 363–370 (2021).
- Lim, L. S. et al. Structural changes in the retinal microvasculature and renal function. *Investig. Ophthalmol. Vis. Sci.* **54**, 2970–2976 (2013).
- Sabanayagam, C. et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit. Health* **2**, e295–e302 (2020).
- Zhang, K. et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal images. *Nat. Biomed. Eng.* **5**, 533–545 (2021).
- Saunders, M. R., Cifu, A. & Vela, M. Screening for chronic kidney disease. *JAMA*. **6**, 615–6166 (2015).
- Kuo, C. C. et al. Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *npj Digit. Med.* **2**, 29 (2019).
- Calvert, N. D. et al. Direct mapping of kidney function by DCE-MRI urography using a tetrazinane organic radical contrast agent. *Nat. Commun.* **1**, 3965 (2023).
- Chen, Z. et al. Using elastography-based multilayer perceptron model to evaluate renal fibrosis in chronic kidney disease. *Ren. Fail.* **1**, 2202755 (2023).
- Chebotaeva, N. et al. Urinary protein and peptide markers in chronic kidney disease. *Int. J. Mol. Sci.* **22**, 12123 (2021).
- Barinotti, A. et al. Serum biomarkers of renal fibrosis: a systematic review. *Int. J. Mol. Sci.* **22**, 14139 (2022).

35. Al Turk, A. A., Estiverne, C., Agrawal, P. R. & Michaud, J. M. Trends and outcomes of the use of percutaneous native kidney biopsy in the United States: 5-year data analysis of the Nationwide Inpatient Sample. *Clin. Kidney J.* **11**, 330–336 (2018).
36. Hogan, J. J., Mocanu, M. & Berns, J. S. The native renal biopsy: update and evidence for best practice. *Clin. J. Am. Soc. Nephrol.* **2**, 354–362 (2016).
37. Luciano, R. L. & Moeckel, G. W. Update on the native renal biopsy: core curriculum. *Am. J. Kidney Dis.* **3**, 404–415 (2019).
38. Korbet, S. M., Volpini, K. C. & Whittier, W. L. Percutaneous renal biopsy of native kidneys: a single-center experience of 1,055 biopsies. *Am. J. Nephrol.* **2**, 153–162 (2014).
39. Kajawo, S. et al. A systematic review of complications associated with percutaneous native kidney biopsies in adults in low- and middle-income countries. *Kidney Int. Rep.* **6**, 78–90 (2020).
40. Centers for Medicare & Medicaid Services. Physician fee schedule look-up tool. <http://www.cms.gov/Medicare/MedicareFee-for-Service-Payment/PFSlookup/index.html> (2023).
41. Cusick, M. M. et al. Population-wide screening for chronic kidney disease: a cost-effectiveness analysis. *Ann. Intern. Med.* **176**, 788–797 (2023).
42. Yan, Y. et al. Retinal microvascular changes in diabetic patients with diabetic nephropathy. *BMC Endocr. Disord.* **1**, 101 (2023).
43. Min, J. W. et al. Relationship between retinal capillary nonperfusion area and renal function in patients with type 2 diabetes. *Investig. Ophthalmol. Vis. Sci.* **14**, 14 (2020).
44. Aronov, M. et al. Retinal vascular signs as screening and prognostic factors for chronic kidney disease: a systematic review and meta-analysis of current evidence. *J. Pers. Med.* **7**, 665 (2021).
45. Kerr, K. F. et al. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology* **1**, 114–121 (2014).
46. Pencina, M. J. et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* **2**, 157–172 (2008).
47. Escofet, X. et al. The presence of glomerular sclerosis at time zero has a significant impact on function after cadaveric renal transplantation. *Transplantation* **3**, 344–346 (2003).
48. Denic, A. et al. Changes in glomerular volume, sclerosis, and ischemia at 5 years after kidney transplantation: incidence and correlation with late graft failure. *J. Am. Soc. Nephrol.* **2**, 346–358 (2023).
49. Akihisa, T. et al. Initial decline in eGFR to predict tolvaptan response in autosomal-dominant polycystic kidney disease. *Clin. Exp. Nephrol.* **6**, 540–551 (2022).
50. Pencina, M. J. et al. Interpreting incremental value of markers added to risk prediction models. *Am. J. Epidemiol.* **6**, 473–481 (2012).
51. Diciolla, M. et al. Patient classification and outcome prediction in IgA nephropathy. *Comput. Biol. Med.* **66**, 278–286 (2015).
52. Naumovic, R. et al. Application of artificial neural networks in estimating predictive factors and therapeutic efficacy in idiopathic membranous nephropathy. *Biomed. Pharmacother.* **64**, 633–638 (2010).
53. Mulugeta, G. et al. Classification of imbalanced data using machine learning algorithms to predict the risk of renal graft failures in Ethiopia. *BMC Med. Inform. Decis. Mak.* **23**, 98 (2023).
54. Kers, J. et al. Deep learning-based classification of kidney transplant pathology: a retrospective, multicentre, proof-of-concept study. *Lancet Digit. Health* **1**, e18–e26 (2022).
55. Ladin, K. et al. Effectiveness of an intervention to improve decision making for older patients with advanced chronic kidney disease: a randomized controlled trial. *Ann. Intern. Med.* **176**, 29–38 (2023).
56. Hou, J. H. et al. Changes in the spectrum of kidney diseases: an analysis of 40,759 biopsy-proven cases from 2003 to 2014 in China. *Kidney Dis.* **4**, 10–19 (2018).
57. Chen, X. J. et al. Changes in spectrum of biopsy-proven kidney diseases within decade: an analysis based on 10 199 cases from South China. *Postgrad. Med J.* **100**, 20–27 (2023).
58. Kopp, J. B. et al. Podocytopathies. *Nat. Rev. Dis. Prim.* **1**, 68 (2020).
59. Jennette, J. C., D’Agati, V. D., Fogo, A. B., Nickleleit, V. & Stokes, M. B. *Heptinstall’s Pathology of the Kidney* (Wolters Kluwer, 2023).
60. Levey, A. S. et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am. J. Kidney Dis.* **64**, 821–935 (2014).
61. Smilkov, D., Thorat, N., Kim, B., Viégas, F.B. & Wattenberg, M. SmoothGrad: removing noise by adding noise. <http://arxiv.org/abs/1706.03825> (2017).
62. Mohammad, S. & Sorower, A. Literature survey on algorithms for multi-label learning. <https://api.semanticscholar.org/CorpusID:13222909> (2010).
63. Godbole, S. & Sarawagi, S., Discriminative methods for multi-labeled classification. In *Proc. Advances in Knowledge Discovery and Data Mining. PAKDD. Lecture Notes in Computer Science* Vol. 3056 (eds Dai, H., Srikant, R. & Zhang, C.) https://doi.org/10.1007/978-3-540-24775-3_5 (Springer, 2004).

Acknowledgements

We appreciate Prof. Richard J. Johnson (Division of Renal Diseases and Hypertension, University of Colorado Anschutz Medical Campus, U.S.), Prof. Tien Yin Wong (Vice Provost of Tsinghua University, Senior Vice Chancellor of Tsinghua Medicine, China), Prof. Thomas Coffman (Dean, Duke-NUS Medical School, Singapore) and Prof. Yemisi Takwoingi (Institute of Applied Health Research, University of Birmingham, UK) for their insightful suggestions regarding this study. We thank Nor Idiris Ahmed and Ibrahim Mohamad Abdullahi (Banadir Hospital, Mogadishu, Somalia), Shuhuan Yao (Dongguan Traditional Chinese Medicine Hospital, Dongguan, China) and Qiongyi Huang (Department of Nephrology, Zhongshan City People’s Hospital, Zhongshan, China) for participating in the human and artificial intelligence comparison tests. We would like to thank Dongmei Xiang (Affiliated Qingyuan Hospital, Guangzhou Medical University, Qingyuan People’s Hospital), Hongwen Jiang and Buajier Ainiwan (Department of Ophthalmology, the First People’s Hospital of Kashi, Kashi Prefecture, China) for their assistance with the patient recruitment and data collection. The authors would like to acknowledge financial support from the National Natural Science Foundation of China (Grant Nos. 92368205 to H.L., 82370707 to W.C. and 82371111 to D.R.L.). Science and Technology Program of Guangzhou (Grant No. 202201020337 to H.L.). The High-level Science and Technology Journals Projects of Guangdong Province (Grant No. 2021B1212010003 to H.L.). Basic scientific research projects of Sun Yat-sen University (Grant No. 23ykcxqt002 to H.L.). Young Talent Support Project of Guangzhou Association for Science and Technology (Grant No. QT-2023-028 to D.L.). The Science and Technology Planning Projects of Guangdong Province (Grant No. 2018B010109008 to H.L.). Guangdong Natural Science Funds for Distinguished Young Scholars (Grant No. 2023B1515020100 to D.R.L.). The funders had no involvement in the study’s design, data collection and analysis, decision to publish, or manuscript preparation. To provide insights into the associations between eye and kidney diseases, Zhongshan Ophthalmic Center and the Department of Nephrology, the First Affiliated Hospital of Sun Yat-sen University, have collaboratively launched the “intelligent eye–kidney alliance” (iEKA) project. This initiative encompasses multiple clinical centers strategically located across various regions, including Zhongshan City People’s Hospital [Zhongshan, China], the First People’s Hospital of Foshan [Foshan, China], the Affiliated Hospital of Youjiang Medical University for Nationalities [Baise, China], the First People’s Hospital of Kashi [Kashi Prefecture, China] and Banadir Hospital [Mogadishu, Somalia]. Parts of the findings were presented (abstract 30074439) at the annual meeting of the American Academy of Ophthalmology from November 3 to November 6, 2023, San Francisco, CA.

Author contributions

Conceptualization: H.L., W.C., Q.W., J.L. and D.R.L.; methodology: L.Z., Q.W. and J.Y.W.; software and validation: L.Z. and J.Y.W.; resources: H.L. and W.C.; data collection and curation: J.L., Q.W., J.Y.W., Y.W., Y.Q.W., N.H., L.J., Z.D., S.Y., W.F.C., Q.L., H.M.L., P.W., W.C.I., J.F., Y.W., H.H., P.Y., Q.Y., C.X., M.Z., X.L., Y.K., J.W., R.M., Y.X., J.G., M.H.R.H., A.H.N., G.C., R.F. S.Z., Y.J. and W.Z.; writing—original draft: Q.W., D.L. and J.Y.W.; participating in the human and artificial intelligence comparison test: X.L., Y.X., L.J., P.Y., N.H., H.J.Y., J.Y. and M.H.R.H.; writing—review and editing: H.L., W.C., D.L., L.Z., D.R.L., K.K.C., W.B.C., R.L., Y.K., J.O.C.L., D.S.W.T. and C.S.; funding acquisition: H.L., W.C., D.R.L. and D.L.; supervision: H.L., W.C. and D.R.L. Q.W., J.L., L.Z., D.L., J.Y.W., Y.W., and Y.Q.W. contributed equally to this work.

Competing interests

Zhongshan Ophthalmic Center and The First Affiliated Hospital, Sun Yat-sen University have filed for patent protection for H.L., W.C., Q.W., L.Z., J.L., D.L., and J.W. for work related to noninvasive screening, diagnosis, and prognosis prediction of CKD. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62273-0>.

Correspondence and requests for materials should be addressed to Duoru Lin, Wei Chen or Haotian Lin.

Peer review information *Nature Communications* thanks Chirag Parikh, Wisit Cheungpasitporn, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

© The Author(s) 2025, corrected publication 2025

¹State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Vision Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou, China. ²Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ³NHC Key Laboratory of Clinical Nephrology (Sun Yat-sen University) and Guangdong Provincial Key Laboratory of Nephrology, Guangzhou, China. ⁴Department of Ophthalmology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ⁵Department of Pathology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ⁶Banadir Hospital, Mogadishu, Somalia. ⁷Department of Nephrology, The First People's Hospital of Kashi, Kashi Prefecture, China. ⁸Health Management Center, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ⁹Department of Nephrology, Zhongshan City People's Hospital, Zhongshan, China. ¹⁰The First People's Hospital of Foshan, Foshan, China. ¹¹Department of Nephrology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China. ¹²Department of Nephrology, Shanxi Provincial Traditional Chinese Medicine Institute, Taiyuan, China. ¹³Department of Nephrology, Second Affiliated hospital of Xi'an Jiaotong University, Xi'an, China. ¹⁴Department of Anesthesiology, Singapore General Hospital, Singapore, Singapore. ¹⁵Duke-NUS Medical School, Singapore, Singapore. ¹⁶Division of Pharmacy, Singapore General Hospital, Singapore, Singapore. ¹⁷Ophthalmology and Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore. ¹⁸Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. ¹⁹Institute of Applied Health Research, University of Birmingham, Birmingham, UK. ²⁰Center for Precision Medicine and Department of Genetics and Biomedical Informatics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. ²¹Hainan Eye Hospital and Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Haikou, China. ²²These authors contributed equally: Qianni Wu, Jianbo Li, Lanqin Zhao, Dong Liu, Jingyi Wen, Yunuo Wang, Yiqin Wang. ✉e-mail: lindr5@mail.sysu.edu.cn; chenwei99@mail.sysu.edu.cn; linht5@mail.sysu.edu.cn