nature biomedical engineering

Article

https://doi.org/10.1038/s41551-025-01413-9

A deep learning system for detecting silent brain infarction and predicting stroke risk

Received: 12 February 2024

Accepted: 24 April 2025

Published online: 06 June 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Current brain imaging to detect silent brain infarctions (SBIs) is not feasible for the general population. Here, to overcome this challenge, we developed a retinal image-based deep learning system, DeepRETStroke, to detect SBI and refine stroke risk. We use 895,640 retinal photographs to pretrain the DeepRETStroke system, which encodes a domain-specific foundation model for representing eye–brain connections. Then, we validated the downstream clinical tasks of DeepRETStroke using 213,762 retinal photographs from diverse datasets across China, Singapore, Malaysia, the USA, the UK and Denmark to detect SBI and predict stroke events. DeepRETStroke performed well in internal validation datasets, with areas under the curve of 0.901 for predicting incident stroke and 0.769 for predicting recurrent stroke. External validations demonstrated consistent performances across diverse datasets. Finally, in a prospective study comprising 218 participants with stroke, we assessed the performance of DeepRETStroke compared with clinical traits in guiding strategies for stroke recurrence prevention. Altogether, the retinal image-based deep learning system, DeepRETStroke, is superior to clinical traits in predicting stroke events, especially by incorporating the detection of SBI, without the need for brain imaging.

Stroke is one of the leading causes of death and long-term disability worldwide[1]. Conventional stroke risk assessments rely on using information about clinical risk factors, mostly from self-reported data such as smoking and history of ischaemic stroke[2–5], and fall short in accurately identifying those at risk. Prior studies demonstrated that the accuracy of conventional stroke risk prediction models was just modest (concordance index (C-index) 0.58–0.73), especially in multiethnic populations[2,6].

Brain imaging such as magnetic resonance imaging (MRI) scans may detect the presence or absence of subclinical cerebrovascular disease[7], and incorporating these brain imaging features into risk assessment may be helpful to more precisely identify individuals at high risk of stroke[7,8]. For example, silent brain infarctions (SBIs), which affect nearly 20% of the general population[9–12], indicate underlying ischaemic cerebrovascular disease and are associated with an increased risk of future stroke[12,13]. Thus, detection of SBI even in asymptomatic patients could potentially allow physicians to refine stroke risk classification and allow patients to be better managed. Scientific statements from the American Heart Association and American Stroke Association suggest that patients with SBI should follow primary prevention guidelines to prevent symptomatic stroke[5,14,15] (Fig. 1a). However, identification of SBIs relies primarily on brain imaging such as MRI and computed tomography (CT), which is impractical and not cost-effective for general stroke screening. Thus, the American Heart Association and American Stroke Association do not recommend screening asymptomatic general population with MRI to detect SBI[14]. This underscores a key clinical gap—how to detect SBI in a simple and cost-effective manner in the general population without the need for brain imaging scans[5,14,16].

Recent advances in medical imaging and deep learning (DL) have highlighted the retina as a unique window to the brain[17]. The retinal vasculature shares embryological, anatomical and physiological similarities with the cerebral vasculature, offering a non-invasive surrogate to detect and predict early cerebrovascular changes[18]. Retinal photography as a non-invasive retinal imaging approach is now widely used across various clinical and community settings for screening eye diseases such as diabetic retinopathy[19], but has also been used with DL techniques to detect various systemic and neurological conditions[20–22].

e-mail: yongjunwang@ncrcnd.org.cn; daiqh@tsinghua.edu.cn; wpjia@sjtu.edu.cn; huarting99@sjtu.edu.cn; shengbin@sjtu.edu.cn; wongtienyin@tsinghua.edu.cn

https://doi.org/10.1038/s41551-025-01413-9

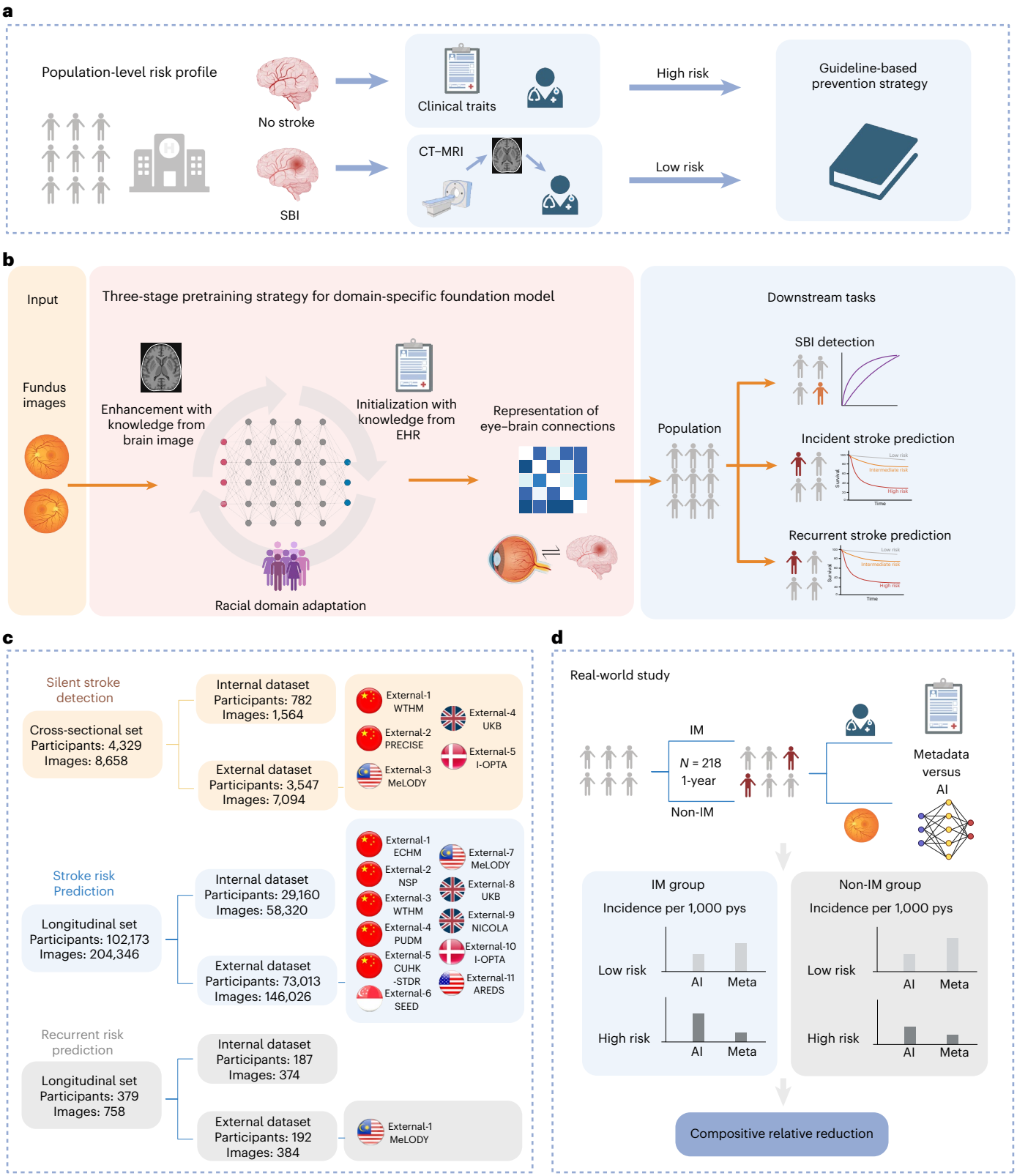


Fig. 1 | Study design of the DeepRETStroke system. a, The established primary care workflow: either high-risk patients or those with SBI were recommended to follow guideline-based primary prevention strategies. **b**, A schematic overview of the DeepRETStroke system. The DeepRETStroke system encodes a domain-specific foundation model representing eye–brain connections, which can be applied to several downstream tasks such as SBI detection and future stroke prediction. EHR, electronic health record. **c**, Multicentre datasets used to develop and validate the DeepRETStroke system. **d**, The design of the real-world prospective observational study to evaluate patient outcomes. pys, person-years. Created with BioRender.com.

https://doi.org/10.1038/s41551-025-01413-9

Table 1 | Characteristics of the developmental, internal and external validation datasets for the detection of SBI

	Developmental	Internal	External-1	External-2	External-3	External-4	External-5
		SDPP	WTHM	PRECISE	MeLODY	UKB	I-OPTA
Fundus images	1,080	484	876	5,576	76	312	254
Participants	540	242	438	2788	38	156	127
Race, *n* (%)							
Chinese	540 (100.0)	242 (100.0)	438 (100.0)	2,788 (100.0)	15 (39.5)	–	–
Indian	–	–	–	–	8 (21.0)	–	–
Malay	–	–	–	–	15 (39.5)	–	–
Asian	–	–	–	–	–	4 (2.6)	–
White	–	–	–	–	–	137 (87.8)	127 (100.0)
Black	–	–	–	–	–	1 (0.6)	–
Other	–	–	–	–	–	14 (9.0)	–
Age (years)	50.1±8.7	51.1±7.9	61.4±12.6	61.0±6.6	65.6±8.9	58.2±9.2	71.1±8.1
Male, *n* (%)	377 (69.8)	200 (82.6)	323 (73.7)	1,461 (52.4)	21 (55.3)	76 (48.7)	79 (62.2)
Smoking, *n* (%)	137 (25.4)	81 (33.5)	109 (24.9)	950 (34.1)	–	71 (45.5)	53 (41.7)
BMI ($kg\,m^{-2}$)	24.5±3.0	25.2±2.9	25.3±2.5	23.8±3.0	26.4±1.9	26.6±4.1	27.0±4.6
Systolic BP (mmHg)	120.9±15.2	123.1±14.8	130.2±17.3	129.2±3.0	145.8±19.2	137.2±19.9	145.1±6.2
Diastolic BP (mmHg)	74.2±10.9	76.2±10.2	78.9±11.0	75.2±9.0	76.4±8.1	81.2±10.6	84.8±13.8
FPG ($mmol\,l^{-1}$)	5.3±0.9	5.4±1.0	6.8±2.7	5.9±1.6	10.5±7.9	5.1±1.4	–
HbA1c (%)	5.7±0.6	5.8±0.7	6.9±1.6	5.9±0.9	8.0±2.2	5.4±0.7	6.5±1.5
TC ($mmol\,l^{-1}$)	4.9±1.1	5.0±0.9	4.2±1.0	5.3±1.0	4.5±0.7	5.8±1.2	4.8±0.8
TG ($mmol\,l^{-1}$)	1.6±1.2	1.8±1.6	1.9±1.8	1.8±1.2	1.6±0.9	1.7±1.0	1.4±0.7
HDL-C ($mmol\,l^{-1}$)	1.3±0.3	1.3±0.3	1.1±0.3	2.8±0.8	1.2±0.3	1.5±0.4	1.6±0.3
LDL-C ($mmol\,l^{-1}$)	3.0±0.7	3.2±0.8	2.5±0.9	1.4±0.3	2.5±0.7	3.6±0.9	2.5±0.9
Hypertension, *n* (%)	182 (33.7)	104 (43.0)	127 (29.0)	1,196 (42.9)	11 (28.9)	77 (49.4)	39 (30.7)
Antihypertensives, *n* (%)	109 (20.2)	68 (28.1)	87 (19.9)	2,043 (73.3)	12 (31.6)	27 (17.3)	49 (38.6)
Diabetes, *n* (%)	50 (9.3)	30 (12.4)	39 (8.9)	269 (9.6)	38 (100.0)	6 (3.8)	9 (7.1)
SBI, *n* (%)	11 (2.0)	5 (2.1)	40 (9.1)	295 (10.6)	4 (10.5)	18 (11.5)	1 (0.8)

Data are presented as mean±s.d. or *n* (%) as appropriate. The en dash means not available or not applicable. PRECISE, Polyvascular Evaluation for Cognitive Impairment and Vascular Events; MeLODY, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia Diabetes Registry in Malaysia cohort; UKB, UK Biobank; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD; FPG, fasting plasma glucose; BP, blood pressure; TC, total cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol.

We introduce DeepRETStroke, a DL system that encodes a domain-specific foundation model for representing eye–brain connections. It utilizes retinal photographs for downstream clinical tasks of detecting SBI and predicting future stroke events, demonstrating the capability of this retinal image-based system to enhance stroke risk assessment in the community. In multiethnic and multicountry datasets, we conducted our study by three stages. First, we used DeepRETStroke to detect SBI from retinal images. We then used SBI features detected from retina to improve incident stroke risk prediction and fine-tune models to predict recurrent stroke. Finally, we conducted a real-world proof-of-concept study to demonstrate the effectiveness of DeepRETStroke in guiding stroke prevention strategies compared with clinical risk prediction models.

Results

Study sample and modelling strategy

The DeepRETStroke system was pretrained using 895,640 retinal photographs from the Shanghai Integrated Diabetes Prevention and Care System (Shanghai Integration Model) and the China National Diabetic Complications Study. The main aim of this study was twofold: (1) to output participant-level SBI detection results (that is, SBI or no SBI) and (2) to output participant-level incident stroke risk results (that is, 5-year stroke risk), by recognizing SBI features from fundus images through knowledge transfer during the joint training. We also fine-tuned the model to enable the prediction of recurrent stroke. Therefore, we trained, validated and tested the DeepRETStroke system for detecting SBI and predicting incident or recurrent stroke from retrospectively collected retinal photographs and clinical traits from diverse datasets from China, Singapore, Malaysia, the USA, the UK and Denmark. Figure 1b gives an overview of the construction and validation of DeepRETStroke (that is, fundus model). To evaluate the performance, we then developed the metadata model and the combined model for comparing with the DeepRETStroke system. For the SBI detection, the metadata model was a logistic-regression classifier with a series of cardiovascular risk factors that were available in these datasets at baseline (more details provided in Methods). For the incident/recurrent stroke prediction, the metadata model was a Cox-proportional hazards model with the same risk factors as those used in the SBI detection task. The combined model was based on both the fundus model and the metadata model. Figure 1c shows the entire validation design of DeepRETStroke in multicountry datasets. Finally, we conducted a prospective real-world study to test the effectiveness

https://doi.org/10.1038/s41551-025-01413-9

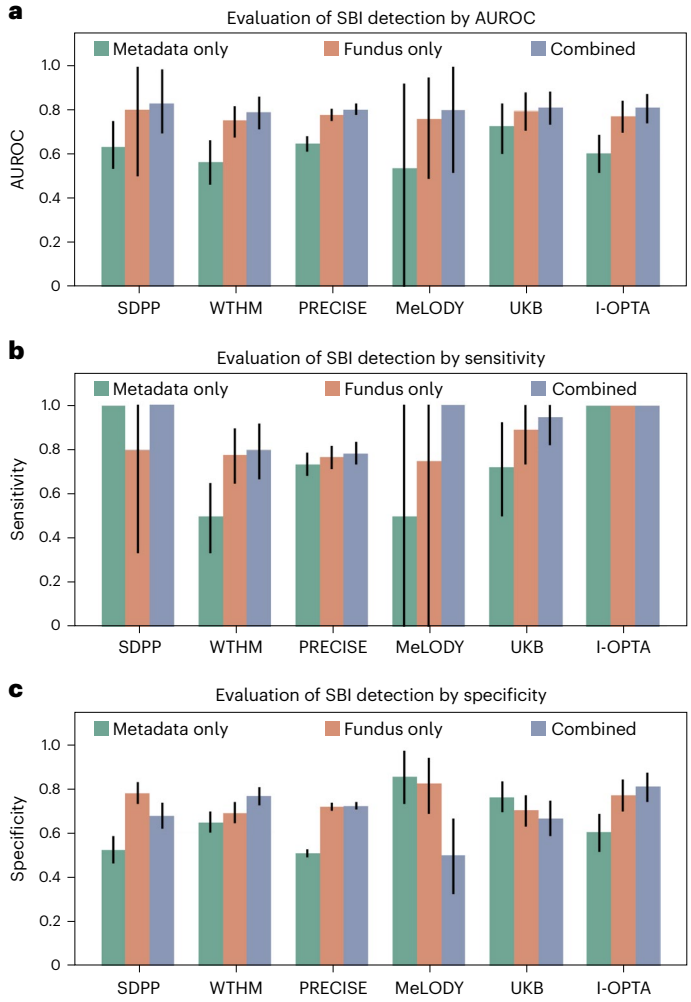


Fig. 2 | Performance of the DeepRETStroke system for the detection of SBI. **a**, The AUROC of the DeepRETStroke system for the detection of SBI. **b**, The sensitivity of the DeepRETStroke system for the detection of SBI. **c**, The specificity of the DeepRETStroke system for the detection of SBI. The error bars represent bootstrapped (n = 1,000) 95% CIs. PRECISE, Polyvascular Evaluation for Cognitive Impairment and Vascular Events; MeLODY, Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia; UKB, UK Biobank; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD.

of the DeepRETStroke system in predicting recurrent stroke (Fig. 1d). For all participants, we used two macular-centred retinal photographs (that is, one for each eye). More details on image quality control and retinal photograph enhancement can be found in Methods.

SBI detection

For the detection of SBI, we developed and internally validated the DeepRETStroke system based on the participants with MRI scans from the Shanghai Diabetes Prevention Program (SDPP) cross-sectional study and externally validated it on five multiethnic datasets. The clinical characteristics of participants are summarized in Table 1. As shown in Fig. 2 and Extended Data Table 1, the fundus model performed well in the internal datasets with an area under the curve (AUC) of 0.797 (95% confidence interval (CI) 0.500 to 0.995), higher than that of the metadata model (0.633, 95% CI 0.533 to 0.748). The sensitivity of the fundus model was 0.800 (0.333 to 1.000) and the specificity was 0.781 (0.730 to 0.830). In independent external datasets, the fundus model achieved AUCs ranging from 0.751 to 0.792, demonstrating the accurate detection performance of DeepRETStroke.

Incident stroke prediction

For the prediction of incident stroke, the DeepRETStroke system was developed and internally validated based on the participants without MRI scans from the SDPP cohort (that is, internal dataset), using the soft labels obtained in the recognition of SBI features, allowing the incorporation of SBI features to augment stroke risk prediction. Then, DeepRETStroke was validated on 11 external multiethnic datasets. The clinical characteristics of the cohort participants are summarized in Table 2. Figure 3 and Extended Data Table 2 show the model performance of DeepRETStroke in predicting incident stroke over a 5-year period. On the internal validation dataset, the fundus model achieved an AUC of 0.901 (95% CI 0.846 to 0.940) and C-index of 0.910 (95% CI 0.853 to 0.950). On the external validation datasets, the AUCs ranged from 0.728 to 0.895. Of note, the AUCs of the fundus model outperformed the metadata model, indicating the possibility of inputting only fundus images to make an accurate prediction of incident stroke via the DeepRETStroke system.

In addition, we used time-dependent analysis at 1–5 years to assess the prognostic accuracy of the three models for stroke prediction. The results of all validation cohorts are shown in Extended Data Fig. 1. Most AUCs of the fundus model were higher than those of the metadata model (that is, clinical traits). In the results of external validation, the performance of the metadata model in the first two years was inconsistent with the fundus model across multiple external datasets, but in the last three years, most AUCs of the fundus model were better than those of the metadata model, which reflects the high concordance and strong calibration of DeepRETStroke system in the long run.

Furthermore, we conducted subgroup analyses for incident stroke prediction according to the baseline health status of the cohort participants (diabetes, hypertension and carotid atherosclerosis). As shown in Extended Data Table 3, the ability of the model to predict incident stroke was consistent among participants with and without diabetes, hypertension and carotid atherosclerosis in the internal cohort and most external validation cohorts. These results demonstrated robust performances of our DeepRETStroke system in predicting incident stroke.

Recurrent stroke prediction

To broaden the application scope and enhance the versatility of the system, the model is further fine-tuned to enable the prediction of recurrent stroke. Extended Data Table 4 presents the clinical characteristics of the cohort participants in developing and validating the system. Extended Data Table 5 shows the model performance of predicting recurrent stroke in 5 years. The fundus model achieved an AUC of 0.769 (95% CI 0.375 to 1.000) and the combined model achieved 0.833 (95% CI 0.500 to 1.000), on the internal dataset. Likewise, the AUCs of the fundus model on the external dataset are higher than those of the metadata model (0.727 versus 0.705).

Real-world prospective exploratory study

To further evaluate the outcome of the integration with clinical workflows, we additionally conducted a real-world study within a prospective cohort, where 218 patients with prior stroke or SBI received either integrated management (IM; integrated hospital-community management programme) or not, according to participant preference and clinical considerations (Fig. 1d). There were 56 participants in the IM group and 162 participants in the non-IM group (Supplementary Table 1). Participants in the IM group were provided with regular clinical and metabolic measurements, advised by specialists in comprehensive hospitals and received lifestyle guidance and peer support at community health service centres. Participants in this programme were followed up for stroke event over a 1-year period. Both the fundus model and the metadata model for recurrent stroke divided the IM group and the non-IM group into low-risk and high-risk groups according to cohort median risk[23]. We calculated the adjusted relative reduction (aRR) of

Table 2 | Characteristics of the developmental, internal and external validation datasets for the prediction of incident stroke

	Developmental set		Internal	External-1	External-2	External-3	External-4	External-5	External-6	External-7	External-8	External-9	External-10	External-11
	Training set	Tuning set	SDPP	ECHM	NDSP	WTHM	PUDM	CUHK-STDR	SEED	MeLODY	UKB	NICOLA	I-OPTA	AREDS
Fundus images	40,910	5,800	11,610	11,940	10,572	4,548	582	1,002	13,198	11,444	77,682	5,928	260	8,870
Participants	20,455	2,900	5,805	5,970	5,286	2,274	291	501	6,599	5,722	38,841	2,964	130	4,435
Race, *n* (%)														
Chinese	20,455 (100.0)	2,900 (100.0)	5,805 (100.0)	5,970 (100.0)	5,286 (100.0)	2,274 (100.0)	291 (100.0)	501 (100.0)	2,605 (39.5)	1,391 (24.3)	–	–	–	–
Indian	–	–	–	–	–	–	–	–	2,141 (32.4)	1,761 (30.8)	–	–	–	–
Malay	–	–	–	–	–	–	–	–	1,853 (28.1)	2,436 (42.6)	–	–	–	–
Asian	–	–	–	–	–	–	–	–	–	–	2,119 (5.5)	–	–	8 (0.2)
White	–	–	–	–	–	–	–	–	–	–	33,428 (86.1)	2,964 (100.0)	130 (100.0)	4,243 (95.7)
Black	–	–	–	–	–	–	–	–	–	–	374 (0.9)	–	–	161 (3.6)
Other	–	–	–	–	–	–	–	–	–	134 (2.3)	2,920 (7.5)	–	–	23 (0.5)
Age (years)	45.9±11.4	45.7±11.8	45.8±11.3	45.8±11.4	61.9±3.9	48.6±11.2	52.1±10.9	60.0±13.1	57.3±9.4	59.3±11.3	56.7±8.2	63.5±8.7	71.1±8.1	69.3±5.1
Male, *n* (%)	13,676 (66.9)	1,981 (68.3)	3,836 (66.1)	3,953 (66.2)	2,242 (42.4)	1,521 (66.9)	164 (56.4)	273 (54.5)	3,437 (52.1)	2,761 (48.3)	17,557 (45.2)	1,560 (52.6)	82 (63.1)	1,948 (43.9)
Smoking, *n* (%)	6,990 (34.2)	998 (34.4)	1,952 (33.6)	2,043 (34.2)	1,263 (23.9)	760 (33.4)	118 (40.5)	52 (10.4)	1,815 (27.5)	–	17,270 (44.5)	1,384 (46.7)	53 (40.8)	2,446 (55.2)
BMI (kg m^{-2})	24.5±3.3	24.5±3.3	24.4±3.2	24.4±3.3	25.0±3.3	24.4±3.4	25.3±3.2	25.9±4.5	25.3±4.5	29.0±3.2	27.4±4.8	28.9±5.1	26.9±4.6	27.5±4.9
Systolic BP (mmHg)	121.8±15.5	121.9±15.9	121.7±15.6	121.6±15.4	134.0±15.1	123.8±15.9	123.4±14.5	137.8±20.9	137.3±20.6	138.7±13.8	137.1±18.4	132.0±18.3	144.8±6.5	135.8±14.6
Diastolic BP (mmHg)	73.6±10.3	73.5±10.7	73.4±10.4	73.4±10.4	83.3±7.8	80.9±11.6	79.4±10.4	77.4±10.6	78.2±10.3	77.2±9.2	81.8±10.0	82.4±10.3	84.8±13.8	77.9±9.3
FPG (mmol l^{-1})	5.6±1.2	5.7±1.3	5.6±1.1	5.6±1.2	6.3±1.4	5.2±1.2	8.6±2.7	7.8±3.1	–	8.5±4.1	5.2±1.1	5.9±1.8	–	–
HbA1c (%)	5.6±0.7	5.6±0.7	5.6±0.7	5.6±0.7	5.9±0.9	5.7±0.7	7.5±1.4	7.5±1.4	6.2±1.2	8.0±2.0	18.1±1.4	5.8±0.8	6.4±1.5	–
TC (mmol l^{-1})	5.0±0.9	5.0±0.9	5.0±0.9	5.0±0.9	5.2±1.0	4.7±0.9	4.6±1.3	4.3±0.9	5.4±1.1	4.7±1.0	5.7±1.1	5.5±1.3	4.8±0.8	4.4±2.8
TG (mmol l^{-1})	1.7±1.4	1.7±1.5	1.7±1.4	1.7±1.3	1.7±1.5	1.7±1.5	1.9±2.2	1.4±0.9	1.8±1.2	1.8±1.3	1.7±1.0	1.6±0.9	1.3±0.7	–
HDL-C (mmol l^{-1})	1.3±0.4	1.3±0.4	1.3±0.4	1.3±0.4	1.3±0.3	1.2±0.3	1.2±0.4	1.4±0.5	1.2±0.4	1.2±0.3	1.5±0.4	1.6±0.4	1.6±0.3	–
LDL-C (mmol l^{-1})	3.0±0.8	3.0±0.8	3.0±0.8	3.0±0.8	3.1±0.8	2.9±0.8	2.7±1.0	2.3±0.7	3.4±0.9	2.6±0.9	3.5±0.9	3.4±1.1	2.6±0.9	–
Hypertension, *n* (%)	5,983 (29.2)	839 (28.9)	1,668 (28.7)	1,721 (28.8)	2,945 (55.7)	768 (33.8)	163 (56.0)	199 (39.7)	3,890 (58.9)	681 (11.9)	20,869 (53.7)	1,035 (34.9)	41 (31.5)	1,681 (37.9)
Antihypertensives, *n* (%)	3,875 (18.9)	525 (18.1)	1,047 (18.0)	1,094 (18.3)	1,780 (33.7)	416 (18.3)	142 (48.8)	–	2,102 (31.9)	665 (11.6)	7,954 (20.5)	1,083 (36.5)	51 (39.2)	3,693 (83.3)
Diabetes, *n* (%)	1,972 (9.6)	312 (10.8)	507 (8.7)	577 (9.7)	1,346 (25.5)	233 (10.2)	291 (100.0)	501 (100.0)	1,697 (25.7)	5,722 (100.0)	2,431 (6.3)	220 (7.4)	9 (6.9)	973 (21.9)
Carotid atherosclerosis, *n* (%)	1,335 (6.5)	200 (6.9)	448 (7.7)	399 (6.7)	–	210 (9.2)	133 (45.7)	–	–	–	–	–	–	1,548 (34.9)
New-onset stroke, *n* (%)	125 (0.6)	12 (0.4)	39 (0.7)	42 (0.7)	177 (3.3)	19 (0.8)	10 (3.4)	20 (4.0)	84 (1.3)	247 (4.3)	786 (2.0)	14 (0.5)	12 (9.2)	357 (8.0)
Follow-up (years)	2.7±1.7	2.7±1.7	2.7±1.7	2.7±1.7	4.5±0.6	3.2±2.1	3.4±1.4	4.1±2.0	6.1±1.0	2.5±1.5	12.8±1.8	3.5±0.7	6.9±3.8	9.2±2.7

Data are presented as mean±s.d. or *n* (%) as appropriate. The en dash means not available. CUHK-STDR, The Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy; SEED, the Singapore Epidemiology of Eye Diseases study; MeLODY, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia Diabetes Registry in Malaysia cohort; UKB, UK Biobank; NICOLA, The Northern Ireland Cohort for the Longitudinal study of Ageing; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD; FPG, fasting plasma glucose; BP, blood pressure; TC, total cholesterol; TG, triglyceride; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

https://doi.org/10.1038/s41551-025-01413-9

https://doi.org/10.1038/s41551-025-01413-9

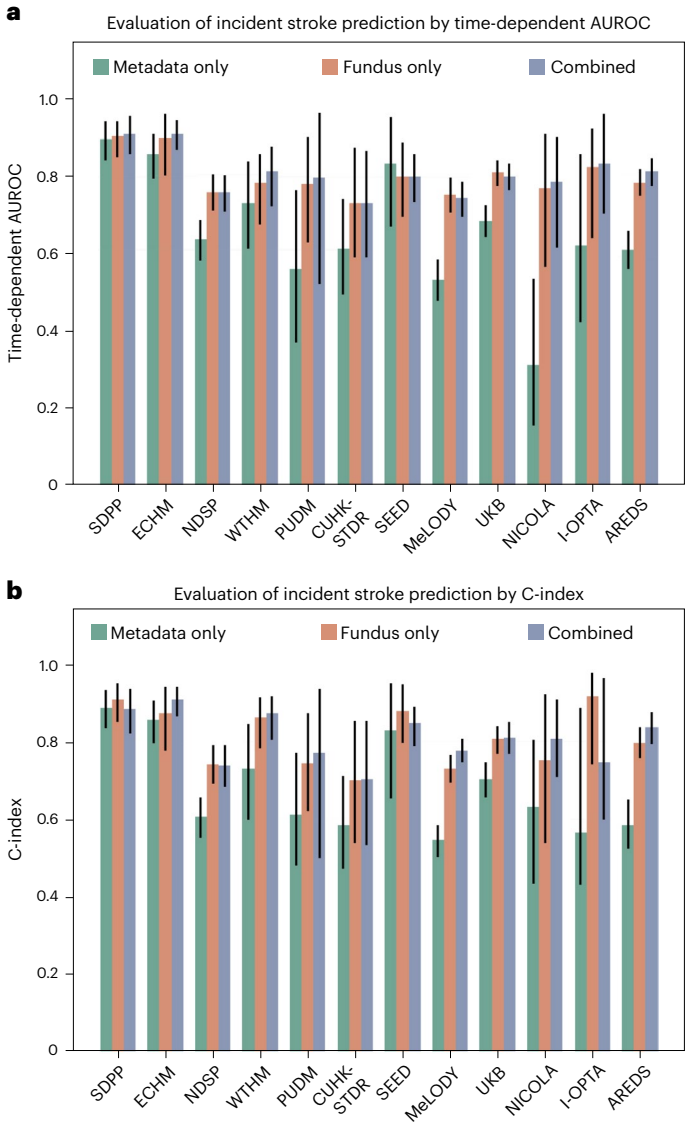


Fig. 3 | Performance of the DeepRETStroke system for the prediction of incident stroke. a, The time-dependent AUROC for the fifth year of the DeepRETStroke system for the prediction of incident stroke. **b**, The C-index of the DeepRETStroke system for prediction of incident stroke. The error bars represent bootstrapped (n = 1,000) 95% CIs. CUHK-STDR, The Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy; SEED, the Singapore Epidemiology of Eye Diseases study; MeLODY, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia Diabetes Registry in Malaysia cohort; UKB, UK Biobank; NICOLA, The Northern Ireland Cohort for the Longitudinal study of Ageing; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD.

incidence between the DeepRETStroke fundus model and the metadata model (Table 3). After adjustment for covariates including demographics, anthropometric indices and biochemical measurements, in the IM group, the difference in the incident stroke events between the fundus model and metadata model was not statistically significant in both the low-risk group (aRR −38.86%, 95% CI −91.5% to 297.93%) and the high-risk group (aRR 48.09%, 95% CI −62.42% to 598.57%). However, patients from the fundus high-risk group of the non-IM group had more incident stroke events compared with the metadata high-risk group (202.17 versus 53.93 per 1,000 person-years, aRR 543.61%, 95% CI 53.68% to 2,572.37%), while the fundus low-risk group had fewer incident stroke events compared with the metadata low-risk group (27.14 versus 136.94 per 1,000 person-years, aRR −97.14%, 95% CI −99.49% to −90.81%). Under comprehensive interventions (that is, intensive intervention for the high-risk group and non-intensive intervention for the low-risk group, stratified by median risk), compared with the metadata model based on clinical traits, the fundus model was associated with 82.44% (95% CI 1.58% to 324.47%) fewer recurrent stroke events (Table 3).

Explainability analysis

We then visualized the interpretability of the DeepRETStroke system. We utilized GradientShap[24] and the occlusion method[25] for visualizing the interpretability of the output prediction. The interpretability summary plot of fundus images for SBI detection and incident stroke prediction are shown in Extended Data Figs. 2 and 3, respectively, highlighting the key anatomical structures associated with SBI lesions and stroke, such as retinal vasculature.

Discussion

Improved precision in stroke risk prediction with simpler, practical and cost-effective methods will lead to reduced morbidity, disability and mortality related to stroke in the general population. To address this important public health and clinical question, we developed a retinal image-based DL system—DeepRETStroke—to detect SBI, a subclinical cerebrovascular disease phenotype, and to use this information to augment stroke risk prediction. Using a large diverse multicountry, multiethnic dataset from China, Singapore, Malaysia, the USA, the UK and Denmark, we trained, validated and externally tested DeepRETStroke. Our main findings are, first, that DeepRETStroke was able to precisely detect the presence of SBI with an AUC of 0.797. Second, by incorporating SBI detection, DeepRETStroke could predict up to 5-year risk of stroke, with an AUC of 0.901 for incident stroke and an AUC of 0.769 for recurrent stroke. We showed largely consistent results in external validation cohorts, in which DeepRETStroke was also able to effectively detect SBI and predict stroke with improved discrimination compared with clinical traits (that is, metadata). Finally, compared with the current assessment based on clinical traits, we showed in a prospective proof-of-concept study that information from DeepRETStroke was able to stratify stroke risk, which was associated with 82.44% (95% CI 1.58% to 324.47%) less recurrent stroke events with appropriate comprehensive interventions.

While some previous studies have investigated the potential of using retinal imaging and artificial intelligence (AI) techniques to assess conventional risk factors and to predict vascular risk including stroke[26–30], our study was unique in at least four important aspects. First, we focused on detecting preclinical and subclinical cerebrovascular disease in the form of SBI, which may affect up to 20% of the population. Detection of SBI and the information from its detection allow DeepRETStroke to more directly and precisely predict future clinical stroke events. This contrasts with other studies using traditional DL algorithms to use relatively simplistic binary outcome labels (presence versus absence of stroke), which may lead to an underutilization of relevant subclinical cerebrovascular phenotypes and imaging characteristics. As a result, the effectiveness of some of these algorithms in predicting stroke outcomes has been, at best, moderate[31]. By leveraging information from brain imaging features such as SBI, our DeepRETStroke showed superior performance in stroke risk prediction. Second, our study incorporated large-scale, international representative multiethnic cohorts and had notably longer follow-up of up to 5 years to predict incident stroke. Third, DeepRETStroke could also predict recurrent stroke in patients with prior stroke, a clinical gap that is useful to stroke neurologists. Finally, we validated findings in an independent prospective cohort, showing that using DeepRETStroke algorithms may help to identify high-risk populations for implementing preventive strategies that may lead to lower risk of future stroke.

Growing evidence suggests compelling biological connections between the brain and the eye that warrant consideration in our study.

https://doi.org/10.1038/s41551-025-01413-9

Table 3 | Associations between risk identification model and recurrent stoke events

		Stroke events Incidence per 1,000 person-years, number of cases/number of people		aRR(95% CI)
Integrated community management programme	IM group (*n*=56)	DeepRETStroke low risk (AI-Low)	Metadata-low risk (Meta-Low)	−38.86 (−91.5 to 297.93)
		73.95 (2/28)	82.05 (2/26)	
		DeepRETStroke high risk (AI-High)	Metadata-high risk (Meta-High)	48.09 (−62.42 to 598.57)
		324.12 (8/28)	292.47 (8/30)	
	Non-IM group (*n*=162)	DeepRETStroke low risk (AI-Low)	Metadata-low risk (Meta-Low)	−97.14 (−99.49 to −90.81)
		27.14 (3/112)	136.94 (6/48)	
		DeepRETStroke high risk (AI-High)	Metadata-high risk (Meta-High)	543.61 (53.68 to 2,572.37)
		202.17 (9/50)	53.93 (6/114)	
	Comprehensive interventions: [(AI-High + AI-Low) − (Meta-High + Meta-Low)] in IM group − [(AI-High + AI-Low) − (Meta-High + Meta-Low)] in Non-IM group			82.44 (1.58 to 324.47)

Intensive intervention for the high-risk group and non-intensive intervention for the low-risk group, stratified by median risk. IM group were provided regular clinical and metabolic measurements, were advised by specialists in comprehensive hospitals and received lifestyle guidance and peer support at community health service centres. Details of biochemical measurements and anthropometric data collection included body weight, waist circumference, blood pressure, lipid profile and related factors of cardiometabolic diseases.

First, the retina, as an extension of the central nervous system, offers a unique opportunity for non-invasive assessment of cerebrovascular health[32–34]. Second, retinal imaging provides valuable insights into microvascular pathology[35], reflecting systemic vascular conditions such as hypertension[36], diabetes[37] and atherosclerosis[36,38]. Third, SBIs and strokes share common underlying vascular aetiologies, including small vessel disease and cerebral microinfarcts[39]. Therefore, given the anatomical and physiological similarities between retinal and cerebral vasculature, alterations observed in retinal microvasculature, such as arteriolar narrowing, venular dilation and microaneurysms, may serve as biomarkers for the early detection of SBI and stroke[40–44].

SBI, in particular, which has been reported to affect nearly 20% of individuals in the community[9–12], frequently remains unidentified until an incidental brain MRI discovery[10]. Because of its prevalence, it is common to misclassify individuals with undetected SBIs as 'healthy' and 'low risk', which could underestimate their stroke risk[45–47]. However, given the high cost and limited availability of MRI, screening the large-scale asymptomatic general population with MRI to detect SBI is prohibitively expensive and impractical[48]. Therefore, our DeepRETStroke system, which uses a non-invasive retinal imaging modality[35], may serve as an assistive and initial screening tool for SBI detection to augment stroke risk prediction in the community without the need for brain imaging.

Strengths of our study include the following. Our study had one of the largest pretrained dataset of retinal photographs and used multiethnic, international cohorts for external validation. Moreover, our algorithm, by learning from small-scale imaging data and using knowledge transfer enhances stroke risk prediction through the recognition of SBI features. However, there are also several limitations that may merit further considerations. First, although we have tested its generalizability in multiple datasets, the dataset used for model development was a solely Chinese cohort, due to difficulties in data sharing of primary retina and MRI imaging data between countries. Second, some inherent biases, including selection bias, unbalanced training data, bias in human labelling, racial and ethnic bias, and unknown confounders, cannot be eliminated and evaluated in our data. Third, as the labelling of our training dataset was based on clinician-derived diagnosis, potential variability in labelling definitions and protocols across cohorts exists, which may have the adverse effect on the development and validation of the DL algorithm. Nevertheless, the semi-supervised learning strategy adopted in DeepRETStroke development can enhance the training set with more easily diagnosable samples, which to some extent mitigate this concern. Fourth, our prospective cohort may be limited by its sample size. Further prospective studies are needed to validate the outcome of the integration with the DeepRETStroke system into clinical practice.

In conclusion, we trained, validated and externally tested a retinal image-based DeepRETStroke system to detect SBI, a common subclinical cerebrovascular disease phenotype, and used this information to augment a prediction algorithm of incident stroke as well as recurrent stroke. We showed the potential of DeepRETStroke in a proof-of-concept prospective cohort study to enhance stroke risk stratification that could be used to guide stroke prevention strategy, without the need for brain imaging.

Methods

Ethical approval

The study received approval from the Ethics Committee of Shanghai Sixth People's Hospital (approval no. 2023-KY-023 [K]). For the development and validation of the DeepRETStroke system, deidentified retrospective data were used, without the active involvement of participants. For the real-world prospective study, all participants provided informed consent before their involvement. All included studies adhered to the tenets of the Declaration of Helsinki and had respective local ethical committee approval.

Study sample

The primary objective of this DeepRETStroke system is to utilize retinal images for both the detection of prevalent SBI and the prediction of incident stroke. The model is further fine-tuned to enable the prediction of recurrent stroke. For SBI detection, we used retinal photographs from six independent datasets for model development and validation. We included participants who underwent retinal photography, brain MRI or CT and had no history of overt stroke. Similarly, for incident stroke prediction, we used retinal photographs from 12 independent datasets for model development and validation. For this task, we included participants who underwent retinal photography, had no history of stroke and had been followed for certain period of time. Furthermore, for recurrent stroke prediction, we used retinal photographs from two independent datasets for model development and validation. For this task, we included participants who underwent retinal photography and had prior stroke history. Image-level data were further filtered via image quality control, and there is no participant-level or image-level overlap between the developmental and validation sets.

Details of these datasets are demonstrated in Supplementary Methods and Supplementary Table 2.

Image quality control

The retinal images were captured using a variety of standard fundus cameras, including Topcon TRC-NW6 (Topcon) and Canon CR1–Mark II (Canon). Following the exclusion criteria proposed by Carol et al.[21], we selected the fundus images as follows: if more than 25% of the peripheral area of the retina was unobservable or the central region of the retina had substantial artefacts that would affect analysis, the photograph was excluded from the dataset. After the image quality control, fundus images were transferred to the AI team to develop and validate our DeepRETStroke system.

Fundus image enhancement

To extract non-specific vascular features that are highly related to our target vascular-related systemic diseases on the fundus photo, a series of image enhancements were proposed to improve the performance of our deep model. In the first step, contrast-limited adaptive histogram equalization was used to enhance the contrast of the image while suppressing noise[49]. We first transformed the input fundus photos from RGB (red, green, blue) to LAB (lightness, green–red, blue–yellow) colour space and divided the images into fixed-size pieces. Then, contrast-limited adaptive histogram equalization was applied on the lightness channel of each piece with its own unique distribution, respectively. After that, the processed images were converted back to RGB space. In the second step, colour normalization was utilized to reduce the colour changes between fundus images caused by different conditions, including the models of the fundus camera, the photography settings, the exposure changes and so on. Following the method proposed by the Kaggle Diabetic Retinopathy Detection Competition Report[50], each fundus image was processed with a Gaussian filter ($P_c = \alpha \times P + \beta \times \mathrm{Gauss}(P, s) + \delta$), where P represents the fundus image, $\mathrm{Gauss}(P, s)$ represents applying a Gaussian filter with a standard deviation of s on P, and α, β and δ represent the parameters. We used $\alpha = 4, \beta = -4$, $s = 5$ and $\delta = 128$. After that, every enhanced fundus image was resized to 512 × 512 for the following training and validation.

Definition and criteria for disease diagnosis

In our study, diagnoses of SBI were made on the basis of clinical diagnostic criteria as follows: physician-diagnosed cerebral infarcts based on brain CT or MRI without any corresponding stroke episode (that is, self-reported history of stroke)[51] (Supplementary Table 3). Diagnoses of stroke were made on the basis of American Heart Association and American Stroke Association criteria[52] and were ascertained as follows: any physician-diagnosed fatal or non-fatal stroke reported by self-report or documented in hospital records during the follow-up period, provided there was no history of overt stroke at baseline. Diagnoses of recurrent stroke were defined as follows: any physician-diagnosed stroke occurring during the follow-up period in individuals with a history of stroke diagnosed at baseline.

Development of the DeepRETStroke system

The DeepRETStroke system encodes a domain-specific foundation model representing eye–brain connections, which was built on a three-stage pretraining strategy. The development of the system was completed according to steps shown in Extended Data Fig. 4. The network architecture is shown in Supplementary Fig. 1. For the first stage, RETFound[31] was used as the primitive encoder of our system, and Masked AutoEncoder[53]—an unsupervised learning algorithm—was used to improve the model's generalizability across different racial groups with two large Chinese datasets—the Shanghai Integration Model cohort (173,346 participants with 693,384 images) and the China National Diabetic Complications Study cohort (50,564 participants with 202,256 images)[54].

For the second stage, we adopted a model initialization step to address the cold start challenge in the next phase of semi-supervised learning. Here, the SDPP dataset (specifically refered to the participants without MRI scans from the SDPP cohort) was used to train the encoder and Stroke Predictor for incident stroke prediction by predicting the risk of onset for the next 5 years. This allows the model to acquire prior knowledge about eye–brain connections before progressing to the subsequent stage.

Based on this foundation, we initiated the third stage to further enhance the representation capabilities of our domain-specific foundation model. Specifically, there existed training iterations to continuously optimize the Encoder of the system, with each iteration consisting of two steps: 'semi-supervised learning' and 'knowledge transfer'. At the same time, the 'semi-supervised learning' step in an iteration consists of several rounds, with each round consisting of two steps: 'update of labelled database' and 'update of SBI Detector'. For a training iteration, the first step was to perform semi-supervised learning with several training rounds. We adopted the collaborative training strategy in semi-supervised learning[55] by treating the left and right eye images of all participants as two sufficiently redundant and conditionally independent views. Before the start of the semi-supervised learning step, the labelled database contained only samples from the SDPP-MRI dataset (specifically refered to the participants with MRI scans from the SDPP cross-sectional study), while the unlabelled database contained all samples from the SDPP dataset. After the start of the semi-supervised learning step, for one training round, we first used the labelled database to update SBI Detector and then used this model to make predictions on the unlabelled database. After prediction, we selected samples with high prediction confidence from the results and labelled them with 'pseudo-labels'. Here, based on our internal validation database, we chose the minimum predictive probability score that can maintain the predictive precision of positive samples (the proportion of true positive samples in the samples predicted as positive) above 0.75 as the high prediction confidence standard for positive samples, and the maximum predictive probability score that could maintain the prediction precision of negative samples (the proportion of true negative samples in the samples predicted as negative) above 0.75 as the high prediction confidence standard for negative samples. These samples were removed from the unlabelled database and added to the labelled database. Then, we used this expanded labelled database to further update SBI Detector. Therefore, along with the training rounds, samples from the unlabelled database were continuously added to the labelled database until there are no samples in the unlabelled database or no samples that could be assigned pseudo-labels, which also indicates the end of the semi-supervised learning step. Afterwards, the updated SBI Detector made predictions for the entire SDPP dataset, directly assigning the predicted 'disease probability distribution' as a 'soft label' to serve as the information on possible SBI at baseline for each sample, and starting the 'knowledge transfer' step. In this step, we once again used the SDPP dataset for incident stroke prediction. But, unlike the model initialization step, here we used the soft labels to construct a new auxiliary task along with the primary prediction task for joint training. To be specific, SBI Learner was trained to fit the probability distribution of SBI at baseline indicated by the soft label, while Stroke Predictor was trained to predict the future incident stroke. Through this approach, the encoder could simultaneously learn the cerebrovascular conditions of a certain participant both at baseline and during the follow-up period, thereby improving the prediction of incident stroke. Afterwards, the encoder optimized in this knowledge transfer step could be regarded as an improved feature extractor for the next training iteration before the entire system development was ended. It is worth noting that we did not directly let the encoder learn labelled data in SBI detection but instead used the soft labels output by SBI Predictor to update the encoder in stroke risk prediction. This is because, in the early stage of semi-supervised learning in SBI detection,

only a small amount of labelled data can be used as training samples, and updating the encoder at this time poses a substantial risk of overfitting. For incident stroke risk prediction, the synchronized prediction task of incident stroke could be regarded as a constraint for soft label learning, thus reducing the possibility of overfitting of the encoder.

After this three-stage pretraining, the domain-specific foundation model that represents eye–brain connections has been successfully developed. Therefore, for the other cerebrovascular related diseases, the encoder of our system can also be used as a pretrained encoder with strong prior knowledge, thereby playing a certain gain role in the training of other detection or prediction tasks. On this basis, when it comes to recurrent stroke prediction, we selected the developed encoder and Stroke Predictor as a pretrained model to perform fine-tuning on the participants from Nicheng Diabetes Screening Project (NDSP) who had stroke history at baseline, thus developing a specific model for recurrent stroke prediction.

Development of the metadata model and combined model

To evaluate the performance of the DeepRETStroke system, two other models with different sets of input data were developed, the metadata model and the combined model. For the SBI detection, the metadata model was a logistic-regression classifier with a series of conventional cardiovascular risk factors at baseline, including age, gender, smoking status (yes/no), body mass index (BMI), systolic blood pressure, total cholesterol, high-density lipoprotein cholesterol, baseline hypertension (yes/no) and baseline diabetes (yes/no). For the incident/recurrent stroke prediction, the metadata model was a Cox-proportional hazards model with the same risk factors as those used in the SBI detection task. The metadata features were normalized and standardized before model training.

For the development of combined models, we first froze the developed encoder and used it to extract high-dimensional retinal features. Then, the extracted features were concatenated to the risk factors used in the metadata model to form new combined features. After that, classifiers (with more input dimensions for the addition of the risk factors) of SBI Detector and Stroke Predictor were used and trained as the combined model of SBI detection and incident/recurrent stroke prediction, respectively.

Implementation details of DeepRETStroke system developmental process

We adopted the retinal feature encoder of RETFound[31] as the prototype of our encoder for subsequent pretraining and fine-tuning. It is a large vision transformer[56] (ViT-large) with 24 transformer blocks and an embedding vector size of 1,024. Each transformer block is composed of multiheaded self-attention and multilayer perceptrons (MLPs), taking feature vectors as input and generating high-level features. For SBI Detector, we used two logistic-regression models with L2 regularization. SBI Learner was composed of a two-layer MLP with a two-dimensional output while Stroke Predictor used the same two-layer MLP with a five-dimensional output instead. For the first stage of pretraining, the objective is the classical MAE algorithm to reconstruct retinal images from the highly masked version, with a mask ratio of 0.75. For the second stage of pretraining, the objective function was as follows:

$$\mathrm{Loss}_{\mathrm{stage2}} = \frac{1}{N}\frac{1}{N_{\mathrm{T}}}\sum_{i,t} -\left[y_{i,t}\log\left[p_{\theta,t}(x_i)\right] + (1-y_{i,t})\log\left[1-p_{\theta,t}(x_i)\right]\right],$$

where $t = 1, 2, 3, 4, 5$ represents years; $N_{\mathrm{T}} = 5$ is for participants with stroke occurring within 5 years or without stroke over 5 years, while $N_{\mathrm{T}} < 5$ is for participants censored within 5 years; $i = 1, 2, 3, \ldots, N$ is the index of each participant; x_i is the input fundus image of ith participant; $p_{\theta,t}(\cdot)$ is the estimated probability of occurring the incident stroke before the timepoint t; $y_{i,t} = 1$ is for incident stroke occurrence before timepoint t; and $y_{i,t} = 0$ is for no incident stroke within a t-year time window. For the third stage of pretraining, the objective function of SBI detection was the classical cross-entropy loss for binary classification. The objective function of incident stroke prediction was just the loss function in the second stage plus a Kullback–Leibler divergence loss by the soft label:

$$\mathrm{Loss}_{\mathrm{stage3}} = \mathrm{Loss}_{\mathrm{stage2}} + \alpha \times \frac{1}{N}\sum_{i} -\left[s^{\mathrm{T}}_{i,0}\log\left[q^{\mathrm{T}}_{i,0}(x_i)\right] + s^{\mathrm{T}}_{i,1}\log\left[q^{\mathrm{T}}_{i,1}(x_i)\right]\right],$$

where $i = 1, 2, 3, \ldots, N$ is the index of each participant; x_i is the input fundus image of the ith participant; $s^{\mathrm{T}}_{i,j}(\cdot)$ is the value of soft label output by SBI Detector on class j; $q^{\mathrm{T}}_{i,j}(\cdot)$ is the value of softmax output by SBI Learner on class j; and α is the parameter to control the effects of the soft label on the learning of this stage. For the fine-tuning of recurrent stroke prediction, the objective function was the loss function in the second stage with the replacement of the predicted event from incident stroke to recurrent stroke.

For the developmental details of the system, we first trained the encoder and the Stroke Predictor for 30 epochs in model initialization step and selected the model with the best C-index on the validation set of incident stroke prediction to be used in the SBI detection. Then, we trained SBI Detector with semi-supervised learning strategy until there were no unlabelled samples remained or SBI Detector could not find any samples with results of high confidence. After that, we trained the encoder, SBI Learner and Stroke Predictor with knowledge transfer strategy for 30 epochs. In this step, if the validation result of incident stroke prediction had surpassed the best result from the previous iterations, we would select the corresponding best model to repeat the training process of the third stage; otherwise, the entire system development would be ended. In recurrent stroke prediction, we performed the fine-tuning on the copy of the encoder and Stroke Predictor for recurrent stroke prediction. The total training epoch is 30, and the model with the best C-index on its validation set was selected.

The entire system was implemented using PyTorch. During the developmental process, training was performed in batches of 128 images after data augmentation, including random horizontal and vertical flips, random rotation and random Gaussian noise addition. We used the Adam optimizer with a learning rate warming up (from 0 to a learning rate of 1×10^{-3}) for ten epochs. The parameter α of the objective function of the third stage was set to 0.3. No samples were overlapping at the patient level in training and validation sets.

Data from UK Biobank and Age-Related Eye Disease Study (AREDS) were accessed under application numbers 104443 and 26125 by Shanghai Jiao Tong University and the Ohio State University, respectively. For domestic datasets from Shanghai (SDPP, NDSP), Beijing (Peking Union Diabetes Management, PUDM), Wuhan (Wuhan Tongji Health Management, WTHM) and Wuxi (The Eastern China Health Management, ECHM), the principal investigator of each study provided data and supervised the data analyses, ensuring the data were appropriately analysed within research team, ensuring that no external entities had access to the data. For other datasets, we delivered a docker program and an instruction guide to principal investigators and researchers at each study site, who conducted the external validation and the related analyses within each cohort and locally following the same instruction guide (https://docs.docker.com/get-started/overview/). After this, the analyses were performed, and the summary statistics and performance matrices (for example, AUC) were then sent back to the requesting team at Shanghai Sixth People's Hospital. No raw data transfer occurred across countries.

Real-world study of the DeepRETStroke system

For the real-world study within a community-based prospective cohort study of Chinese adults, 215 participants with prior stroke and 3 participants with SBI were screened in November 2022 (Extended Data Fig. 5).

Among these patients, 56 received IM, while 162 did not. IM group were provided regular clinical and metabolic measurements, advised by specialists in comprehensive hospitals, received lifestyle guidance and peer support at community health service centres. Details of biochemical measurements and anthropometric data collection included body weight, waist circumference, blood pressure, lipid profile and related factors of cardiometabolic diseases.

Explainability analysis of the DeepRETStroke system

We utilized GradientShap[24] and the occlusion method[25] for visualizing the interpretability of the output predictions from the DeepRETStroke system. GradientShap approximates SHapley Additive exPlanations values by computing the expectations of gradients by randomly sampling from the distribution of baselines or references. It adds white noise to each input sample n times, selects a random baseline from baselines' distribution and a random point along the path between the baseline and the input, and computes the gradient of outputs with respect to those selected random points. The occlusion method is a perturbation-based approach to compute attribution, involving replacing each contiguous rectangular region with a given baseline or reference and computing the difference in output. For features located in multiple regions (hyperrectangles), the corresponding output differences are averaged to compute the attribution for that feature. For the output results of GradientShap and occlusion method, we selected green and red colour to show the region where the models pay more attention, respectively.

Statistical analysis

To evaluate the performance of DeepRETStroke system, for SBI detection, we used ROC curves of sensitivity versus 1 − specificity. The area under the receiver operating characteristic (AUROC) was calculated with 95% CIs by the non-parametric bootstrap method (1,000 random resampling with replacement). Sensitivity and specificity were estimated by the best cut-off value of the output scores (Youden index for this evaluation) on the validation set. The significant differences between AUROC were computed using Delong methods[57]. For incident/recurrent stroke prediction, we used Harrell's C-index and time-dependent AUROC[58]. Each method was adjusted for censoring by weighing with the inverse probability of censoring and calculated for data before a given cut-off time τ with 95% CIs by the non-parametric bootstrap method. A two-sided significance level of 5% was considered as statistically significant.

Reporting summary

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

Data availability

The main data supporting the results in this study are available within the article and its Supplementary Information. Individual-level patient data are protected because of patient privacy; they are accessible with the consent of the data management committee from institutions and are not publicly available. Requests for the non-profit use of the retinal fundus images and related clinical information should be sent to T.Y.W. (wongtienyin@tsinghua.edu.cn). The data management committee will then review all the requests and grant permission (if successful). A formal data transfer agreement will be required upon approval. Generally, all these requests for access to the data will be responded to within 1 month. All data shared will be deidentified. Source data are provided with this paper.

Code availability

The code being used in the current study for developing the algorithm is available via GitHub at https://github.com/njiang2024/DeepRETStroke (ref. 59).

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol.* **20**, 795–820 (2021).
2. Zhang, K. et al. Utility of China-PAR stroke equations for predicting 10-year stroke risk in the rural Inner Mongolian population in China. *Neurol. Res.* **44**, 989–994 (2022).
3. Judd, S. E. et al. Self-report of stroke, transient ischemic attack, or stroke symptoms and risk of future stroke in the REasons for Geographic And Racial Differences in Stroke (REGARDS) study. *Stroke* **44**, 55–60 (2013).
4. Reitz, C. et al. Validity of self-reported stroke in elderly African Americans, Caribbean Hispanics, and Whites. *Arch. Neurol.* **66**, 834–840 (2009).
5. Kleindorfer, D. O. et al. 2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline from the American Heart Association/American Stroke Association. *Stroke* **52**, e364–e467 (2021).
6. Hong, C. et al. Predictive accuracy of stroke risk prediction models across black and white race, sex, and age groups. *JAMA* **329**, 306–317 (2023).
7. Kelly, P. J. et al. Validation and comparison of imaging-based scores for prediction of early stroke risk after transient ischaemic attack: a pooled analysis of individual-patient data from cohort studies. *Lancet Neurol.* **15**, 1238–1247 (2016).
8. Kang, D. W. et al. Silent new ischemic lesions after index stroke and the risk of future clinical recurrent stroke. *Neurology* **86**, 277–285 (2016).
9. Prabhakaran, S. et al. Prevalence and determinants of subclinical brain infarction: the Northern Manhattan Study. *Neurology* **70**, 425–430 (2008).
10. Fanning, J. P., Wesley, A. J., Wong, A. A. & Fraser, J. F. Emerging spectra of silent brain infarction. *Stroke* **45**, 3461–3471 (2014).
11. Fanning, J. P., Wong, A. A. & Fraser, J. F. The epidemiology of silent brain infarction: a systematic review of population-based cohorts. *BMC Med.* **12**, 119 (2014).
12. Vermeer, S. E., Longstreth, W. T. Jr. & Koudstaal, P. J. Silent brain infarcts: a systematic review. *Lancet Neurol.* **6**, 611–619 (2007).
13. Gupta, A. et al. Silent brain infarction and risk of future stroke: a systematic review and meta-analysis. *Stroke* **47**, 719–725 (2016).
14. Smith, E. E. et al. Prevention of stroke in patients with silent cerebrovascular disease: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* **48**, e44–e71 (2017).
15. Meschia, J. F. et al. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* **45**, 3754–3832 (2014).
16. Burke, J. F., Gelb, D. J., Quint, D. J., Morgenstern, L. B. & Kerber, K. A. The impact of MRI on stroke management and outcomes: a systematic review. *J. Eval. Clin. Pr.* **19**, 987–993 (2013).
17. Huang, Y. et al. AI-integrated ocular imaging for predicting cardiovascular disease: advancements and future outlook. *Eye* https://doi.org/10.1038/s41433-023-02724-4 (2023).
18. Ting, D. S. W. & Wong, T. Y. Eyeing cardiovascular risk factors. *Nat. Biomed. Eng.* **2**, 140–141 (2018).
19. Guan, Z. et al. Artificial intelligence in diabetes management: advancements, opportunities, and challenges. *Cell Rep. Med.* **4**, 101213 (2023).
20. He, G. et al. Fundoscopy use in neurology departments and the utility of smartphone photography: a prospective prevalence and crossover diagnostic accuracy study amongst neurology inpatients. *Eur. J. Neurol.* **29**, 2463–2472 (2022).

21. Cheung, C. Y. et al. A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case–control study. *Lancet Digit. Health* **4**, e806–e815 (2022).
22. 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).
23. Dai, L. et al. A deep learning system for predicting time to progression of diabetic retinopathy. *Nat. Med.* https://doi.org/10.1038/s41591-023-02702-z (2024).
24. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Proc. 30th Conference on Advances in Neural Information Processing Systems* (eds Isabelle, G. et al.) 4765–4774 (Neural Information Processing Systems, 2017).
25. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *Proc. 13th European Conference on Computer Vision* (eds David, J. F. et al.) 818–833 (Springer, 2014).
26. Gerrits, N. et al. Age and sex affect deep learning prediction of cardiometabolic risk factors from retinal images. *Sci. Rep.* **10**, 9432 (2020).
27. Zhang, L. et al. Prediction of hypertension, hyperglycemia and dyslipidemia from retinal fundus photographs via deep learning: a cross-sectional study of chronic diseases in central China. *PLoS ONE* **15**, e0233166 (2020).
28. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
29. Zhu, Z. et al. Retinal age gap as a predictive biomarker of stroke risk. *BMC Med.* **20**, 466 (2022).
30. 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).
31. Zhou, Y. et al. A foundation model for generalizable disease detection from retinal images. *Nature* **622**, 156–163 (2023).
32. Cheung, C. Y., Ikram, M. K., Chen, C. & Wong, T. Y. Imaging retina to study dementia and stroke. *Prog. Retin Eye Res.* **57**, 89–107 (2017).
33. Cheung, C. Y. et al. Retinal microvascular changes and risk of stroke: the Singapore Malay Eye Study. *Stroke* **44**, 2402–2408 (2013).
34. Baker, M. L., Hand, P. J., Wang, J. J. & Wong, T. Y. Retinal signs and stroke: revisiting the link between the eye and brain. *Stroke* **39**, 1371–1379 (2008).
35. Cheung, N. et al. Retinal microvascular abnormalities and subclinical magnetic resonance imaging brain infarct: a prospective study. *Brain* **133**, 1987–1993 (2010).
36. Rim, T. H., Teo, A. W. J., Yang, H. H. S., Cheung, C. Y. & Wong, T. Y. Retinal vascular signs and cerebrovascular diseases. *J. Neuroophthalmol.* **40**, 44–59 (2020).
37. Tan, T. E. et al. Imaging modalities for assessing the vascular component of diabetic retinal disease: review and consensus for an updated staging system. *Ophthalmol. Sci.* **4**, 100449 (2024).
38. Wang, S. B. et al. A spectrum of retinal vasculature measures and coronary artery disease. *Atherosclerosis* **268**, 215–224 (2018).
39. Meinel, T. R., Kaesmacher, J., Roten, L. & Fischer, U. Covert brain infarction: towards precision medicine in research, diagnosis, and therapy for a silent pandemic. *Stroke* **51**, 2597–2606 (2020).
40. Wong, T. Y. Is retinal photography useful in the measurement of stroke risk? *Lancet Neurol.* **3**, 179–183 (2004).
41. Lindley, R. I. et al. Retinal microvasculature in acute lacunar stroke: a cross-sectional study. *Lancet Neurol.* **8**, 628–634 (2009).
42. McGeechan, K. et al. Prediction of incident stroke events based on retinal vessel caliber: a systematic review and individual-participant meta-analysis. *Am. J. Epidemiol.* **170**, 1323–1332 (2009).
43. Wong, T. Y. et al. Cerebral white matter lesions, retinopathy, and incident clinical stroke. *JAMA* **288**, 67–74 (2002).
44. Wong, T. Y. et al. Retinal microvascular abnormalities and incident stroke: the Atherosclerosis Risk in Communities Study. *Lancet* **358**, 1134–1140 (2001).
45. Hobeanu, C. et al. Risk of subsequent disabling or fatal stroke in patients with transient ischaemic attack or minor ischaemic stroke: an international, prospective cohort study. *Lancet Neurol.* **21**, 889–898 (2022).
46. Weber, R. et al. Risk of recurrent stroke in patients with silent brain infarction in the Prevention Regimen for Effectively Avoiding Second Strokes (PRoFESS) imaging substudy. *Stroke* **43**, 350–355 (2012).
47. Gioia, L. C. et al. Silent ischemic lesions in young adults with first stroke are associated with recurrent stroke. *Neurology* **79**, 1208–1214 (2012).
48. Hassell, M. E. et al. Silent cerebral infarcts associated with cardiac disease and procedures. *Nat. Rev. Cardiol.* **10**, 696–706 (2013).
49. Pisano, E. D. et al. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging* **11**, 193–200 (1998).
50. Graham, B. *Kaggle Diabetic Retinopathy Detection Competition Report* (University of Warwick, 2015).
51. Zhu, Y. C., Dufouil, C., Tzourio, C. & Chabriat, H. Silent brain infarcts: a review of MRI diagnostic criteria. *Stroke* **42**, 1140–1145 (2011).
52. Sacco, R. L. et al. An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* **44**, 2064–2089 (2013).
53. He, K. et al. Masked autoencoders are scalable vision learners. In *Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 15979–15988 (IEEE, 2022).
54. Dai, L. et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* **12**, 3242 (2021).
55. Engelen, J. E. V. & Hoos, H. H. A survey on semi-supervised learning. *Mach. Learn.* **109**, 373–440 (2020).
56. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. In *Proc. 9th International Conference on Learning Representations* (OpenReview.net, 2021).
57. Harrell, F. E. Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *JAMA* **247**, 2543–2546 (1982).
58. Kamarudin, A. N., Cox, T. & Kolamunnage-Dona, R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med. Res. Methodol.* **17**, 53 (2017).
59. Jiang, N. et al. DeepRETStroke. *GitHub* https://github.com/njiang2024/DeepRETStroke (2025).

Acknowledgements

This study was supported by the National Key R&D Program of China (grant no. 2022YFC2502800), the National Natural Science Foundation of China (grant no. 82388101) and the Beijing Natural Science Foundation (grant no. IS23096) to T.Y.W.; the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0523200), the Excellent Young Scientists Fund of the National Natural Science Foundation of China (grant no. 82022012) and the Innovative research team of high-level local universities in Shanghai (grant no. SHSMU-ZDCX20212700) and the Shanghai Municipal Key Clinical Specialty to H.L.; the National Key R&D Program of China (grant no. 2022YFC2407000), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (grant nos. 2023ZD0509202

and 2023ZD0509201) and the General Program of the National Natural Science Foundation of China (grant no. 62272298) to B.S.; National Postdoctoral Innovative Talent Support Program (grant no. BX20230189), the Shuimu Scholar Program of Tsinghua University (grant no. 2023SM196) and the China Postdoctoral Science Foundation (grant no. 2024M751733) to H.J. This research has been conducted using the UK Biobank Resource under application number 104443. The computation resources in this article 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.

Author contributions

Y. Wang, Q.D., W.J., H.L., B.S. and T.Y.W. participated in conceptualization, methodology, resources, writing of the original draft, supervision and funding acquisition. N.J. participated in the conceptualization, methodology, formal data analysis and writing of the original draft. H.J., Z.G., D.L., S.W., W.J., H.L., B.S. and T.Y.W. participated in the conceptualization, methodology and writing of the original draft. N.J., H.J., Z.G., Y.P., C.D., D.L., T.C., S.W., D.Y., A.R.R., H.H., M.L.C., C.Y., B.S.T., F.N.P., Q.P., T.C.Q., J.H.L.G., S.S., A.S.A.R., J.W.S.L.-B., Y.L., S.H., J. Shu, S.Y., T.L., J.W., Xiaolong Yang, X.W., H.Z., Y.L., Y. Zhou, S. Lin, C.W., R.D., L.R., R.E.H., D.W., P.Z., G.J.M., J.G., L.-L.L., Z.L., C.Y.C., Y.C.T., C.-Y.C., Y. Wang, W.J., H.L., B.S. and T.Y.W. collected and organized data. Y.G., Y. Wu, Y.J., Y.Q., T.H., Z.W., Y. Zhao, S. Lee, J. Shen, Y.X.W., Y. Zheng, G.S.W.T., C.S., Y.B., C.Z., W.Z., M.G., Xiaokang Yang and Q.D. provided critical comments and reviewed the paper. All authors contributed to the research, editing and approval of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41551-025-01413-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41551-025-01413-9.

Correspondence and requests for materials should be addressed to Yongjun Wang, Qionghai Dai, Weiping Jia, Huating Li, Bin Sheng or Tien Yin Wong.

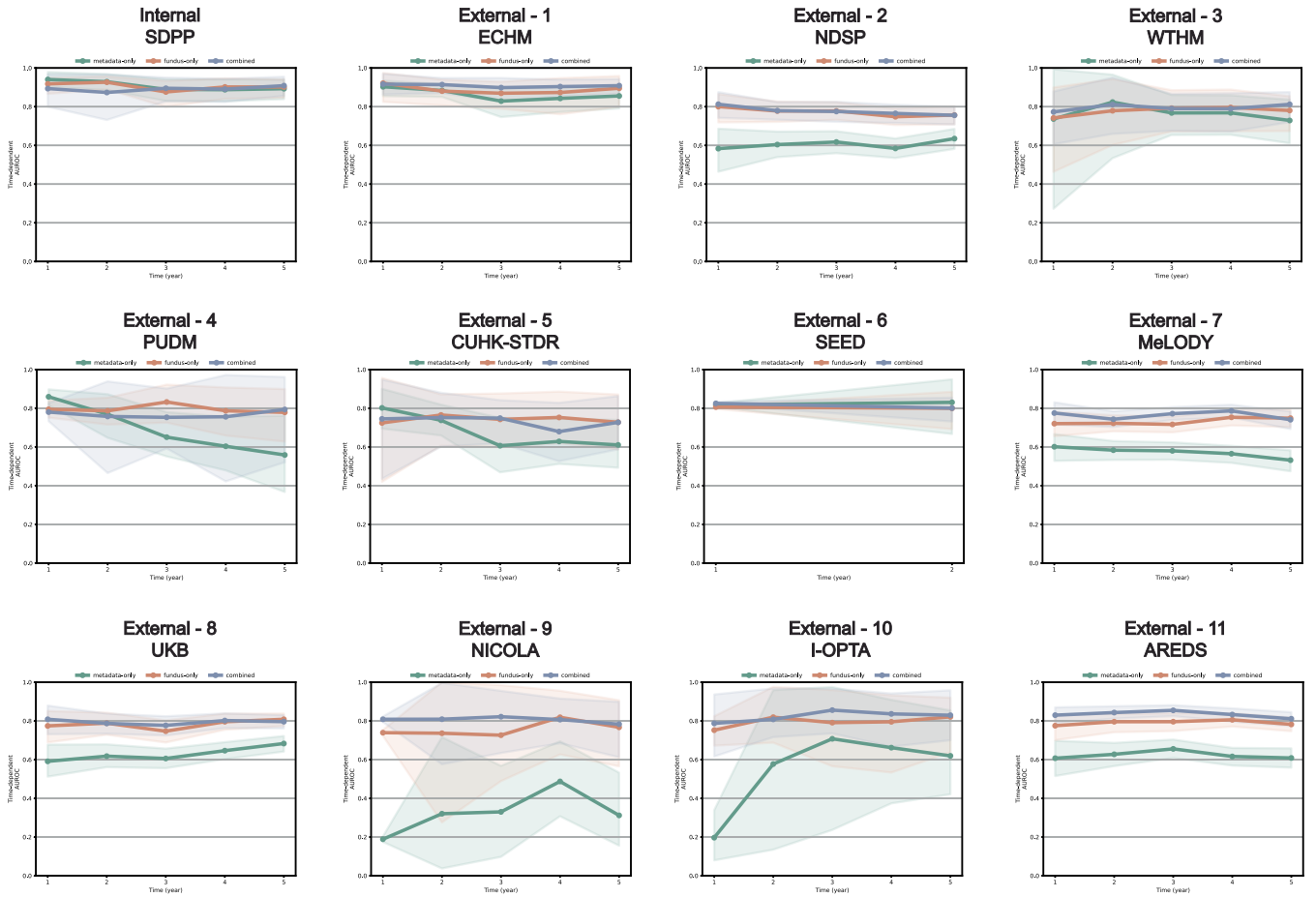
Peer review information *Nature Biomedical Engineering* thanks the anonymous reviewers for their contribution to the peer review of this work.

Nan Jiang[1,2,3,33], Hongwei Ji[3,33], Zhouyu Guan[1,33], Yuesong Pan[4,33], Chenxin Deng[5,6,33], Yuchen Guo[7,33], Dan Liu[1,33], Tingli Chen[8,33], Shiyu Wang[1], Yilan Wu[3], Dawei Yang[9], An Ran Ran[9], Haslina Hamzah[10], Miao Li Chee[10], Changchang Yin[11,12], Benjamin Sommer Thinggaard[13,14], Frederik N. Pedersen[13], Qingsheng Peng[10,15], Ten Cheer Quek[10], Jocelyn Hui Lin Goh[10], Sarkaaj Singh[16], Anis Syazwani Abd Raof[16], Jian Wen Samuel Lee-Boey[16], Yuwei Lu[1], Shan Huang[1,2], Jia Shu[1,2], Shujie Yu[1], Yixiao Jin[3], Tingyao Li[1,2], Yiming Qin[1,2,3], Jing Wang[8], Xiaolong Yang[8], Tingting Hu[1], Zheyuan Wang[1,2], Yaoning Zhao[3], Seungmin Lee[17], Xiaoer Wei[18], Haotian Zheng[1,2], Yuehua Li[18], Jie Shen[19], Yan Zhou[20], Shiqun Lin[20], Chan Wu[20], Rongping Dai[20], Lei Ruan[5,6], Ruth E. Hogg[21], David Wright[21], Ya Xing Wang[22,23,24], Yingfeng Zheng[25], Gavin Siew Wei Tan[10,26], Charumathi Sabanayagam[10,26], Yuqian Bao[1], Cuntai Zhang[5,6], Ping Zhang[11,12], Weiwen Zou[17], Minyi Guo[1], Xiaokang Yang[2], Gareth J. McKay[21], Jakob Grauslund[27], Lee-Ling Lim[16,28,29], Zixiao Li[4], Carol Y. Cheung[9], Yih Chung Tham[10,26,30,31], Ching-Yu Cheng[10,26,30,31], Yongjun Wang[4] ✉, Qionghai Dai[32] ✉, Weiping Jia[1] ✉, Huating Li[1] ✉, Bin Sheng[1,2] ✉ & Tien Yin Wong[3,10,23,24] ✉

[1]Shanghai Belt and Road International Joint Laboratory for Intelligent Prevention and Treatment of Metabolic Disorders, Department of Computer Science and Engineering, School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, 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 Center for Diabetes, Shanghai Key Laboratory of Diabetes Mellitus, Shanghai, China. [2]MOE Key Laboratory of AI, School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. [3]Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital Eye Center, Tsinghua Medicine, Tsinghua University, Beijing, China. [4]Department of Neurology, China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, China. [5]Department of Geriatrics, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [6]Key Laboratory of Vascular Aging, Ministry of Education, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. [7]Institute for Brain and Cognitive Sciences, BNRist, Tsinghua University, Beijing, China. [8]Department of Ophthalmology, Shanghai Health and Medical Center, Wuxi, China. [9]Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, China. [10]Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore. [11]Department of Computer Science and Engineering, The Ohio State University,

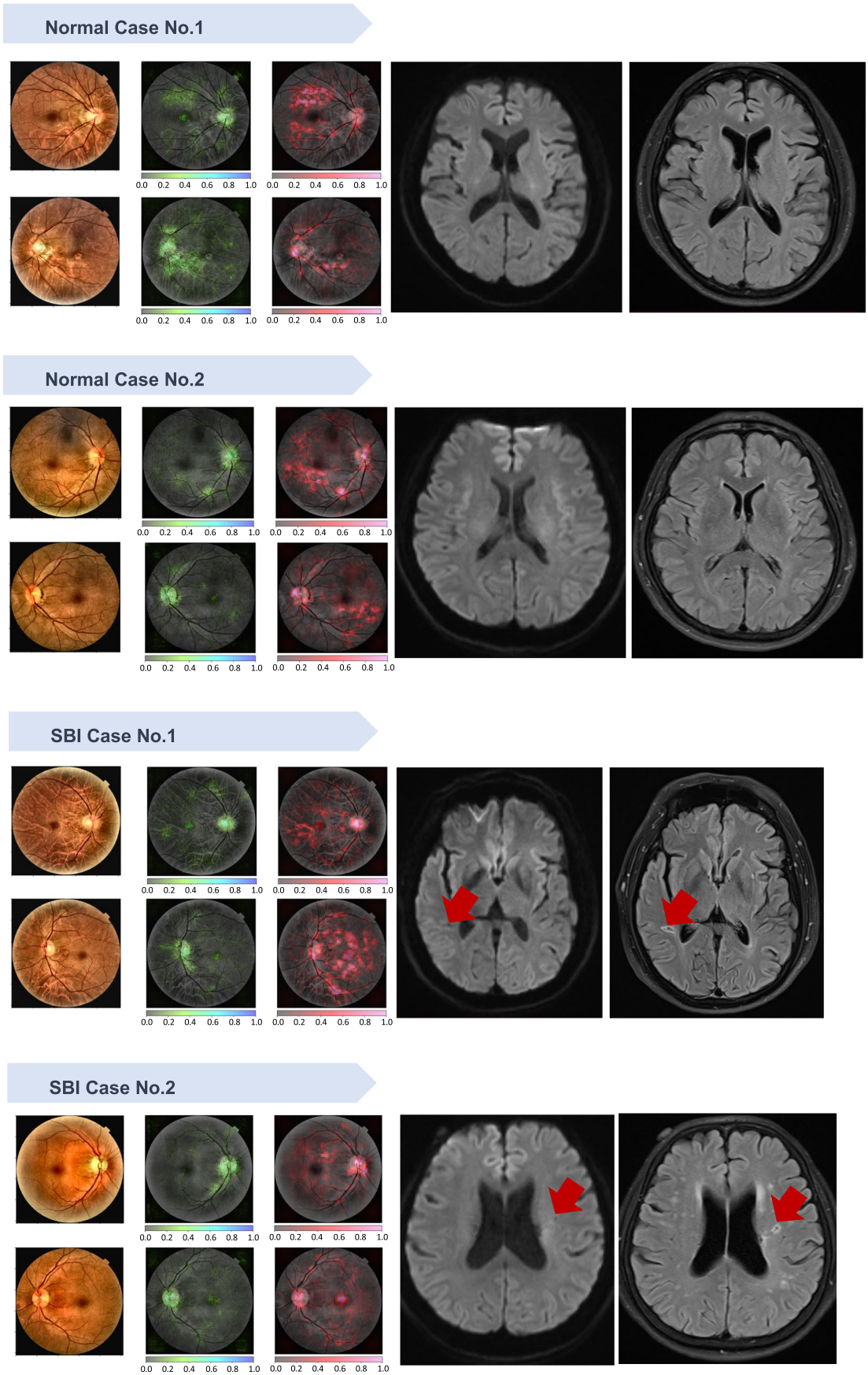
Columbus, OH, USA. [12]Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA. [13]Department of Ophthalmology, Odense University Hospital, Odense, Denmark. [14]OPEN, Open Patient data Explorative Network, Odense University Hospital, Odense, Denmark. [15]Clinical and Translational Sciences Program, Duke-NUS Medical School, Singapore, Singapore. [16]Department of Medicine, Faculty of Medicine, Universiti Malaya, Kuala Lumpur, Malaysia. [17]State Key Laboratory of Micro-Nano Engineering Science, Intelligent Microwave Lightwave Integration Innovation Center, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. [18]Institute of Diagnostic and Interventional Radiology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China. [19]Medical Records and Statistics Office, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China. [20]Department of Ophthalmology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China. [21]Centre for Public Health, Faculty of Medicine, Health and Life Sciences, Queen's University Belfast, Belfast, UK. [22]Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Medical Artificial Intelligence Research and Verification Laboratory of the Ministry of Industry and Information Technology, Beijing Tongren Hospital, Capital Medical University, Beijing, China. [23]School of Clinical Medicine, Beijing Tsinghua Changgung Hospital, Tsinghua Medicine, Tsinghua University, Beijing, China. [24]Beijing Key Laboratory of Intelligent Diagnostic Technology and Devices for Major Blinding Eye Diseases, Tsinghua Medicine, Tsinghua University, Beijing, China. [25]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China. [26]Ophthalmology and Visual Science Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore. [27]Department of Ophthalmology, Steno Diabetes Center Odense, Odense University Hospital, Odense, Denmark. [28]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. [29]Asia Diabetes Foundation, Hong Kong, China. [30]Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [31]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [32]Department of Automation, Tsinghua University, Beijing, China. [33]These authors contributed equally: Nan Jiang, Hongwei Ji, Zhouyu Guan, Yuesong Pan, Chenxin Deng, Yuchen Guo, Dan Liu, Tingli Chen. ✉e-mail: yongjunwang@ncrcnd.org.cn; daiqh@tsinghua.edu.cn; wpjia@sjtu.edu.cn; huarting99@sjtu.edu.cn; shengbin@sjtu.edu.cn; wongtienyin@tsinghua.edu.cn

https://doi.org/10.1038/s41551-025-01413-9



Extended Data Fig. 1 | Time-Dependent Analysis for Predicting Incident Stroke Event by Validation Cohorts. Shaded areas represent bootstrapped (n = 1,000) 95% confidence intervals. SDPP: Shanghai Diabetes Prevention Program; ECHM, The Eastern China Health Management; NDSP, Nicheng Diabetes Screening Project; WTHM, Wuhan Tongji Health Management; PUDM: Peking Union Diabetes Management; CUHK-STDR, The Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy; SEED, the Singapore Epidemiology of Eye Diseases study; MeLODY, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia Diabetes Registry in Malaysia cohort; UKB, UK biobank; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD; AREDS, Age-Related Eye Disease Study; AUC: area under the curve. CI: confidence interval.

https://doi.org/10.1038/s41551-025-01413-9

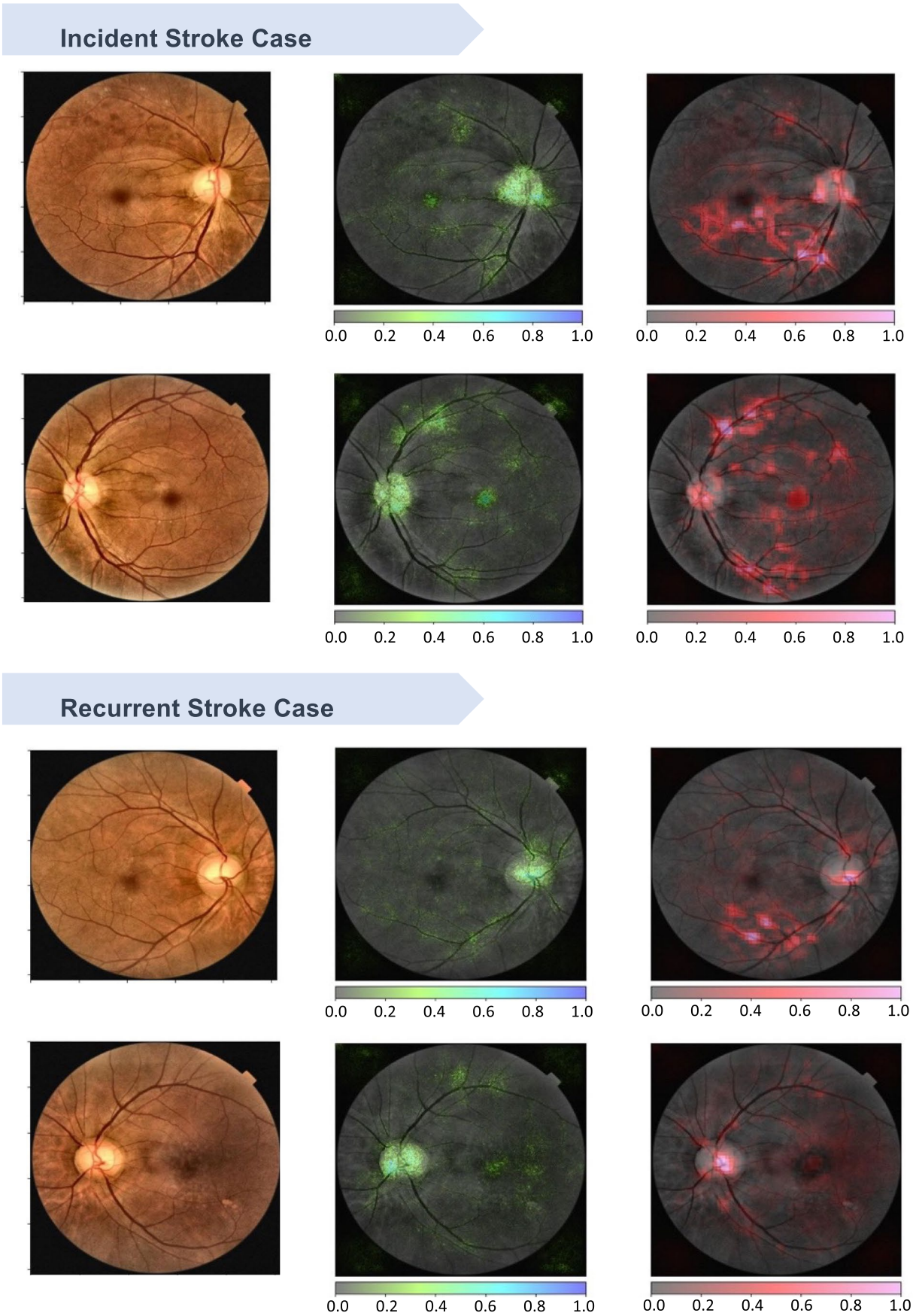


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41551-025-01413-9

Extended Data Fig. 2 | Explainability Analysis of DeepRETStroke System in Detecting Silent Brain Infarction. Left Panel: Original color fundus images of participants. Middle Panel: SHAP (Shapley Additive Explanations) summary plot of fundus images. The pixel color from green to blue indicates the increasing expectation of gradients for SBI detection. Right Panel: Occlusion-based attribution plot of fundus images. Red pixels reveal the areas with positive attribution highlighting possible pathological changes. For the detection of SBI (silent brain infarct), some anatomical structures associated with specific diseases, for example retinal vasculature, are highlighted. MRI Image: MRI lesion of SBI was highlighted.

https://doi.org/10.1038/s41551-025-01413-9

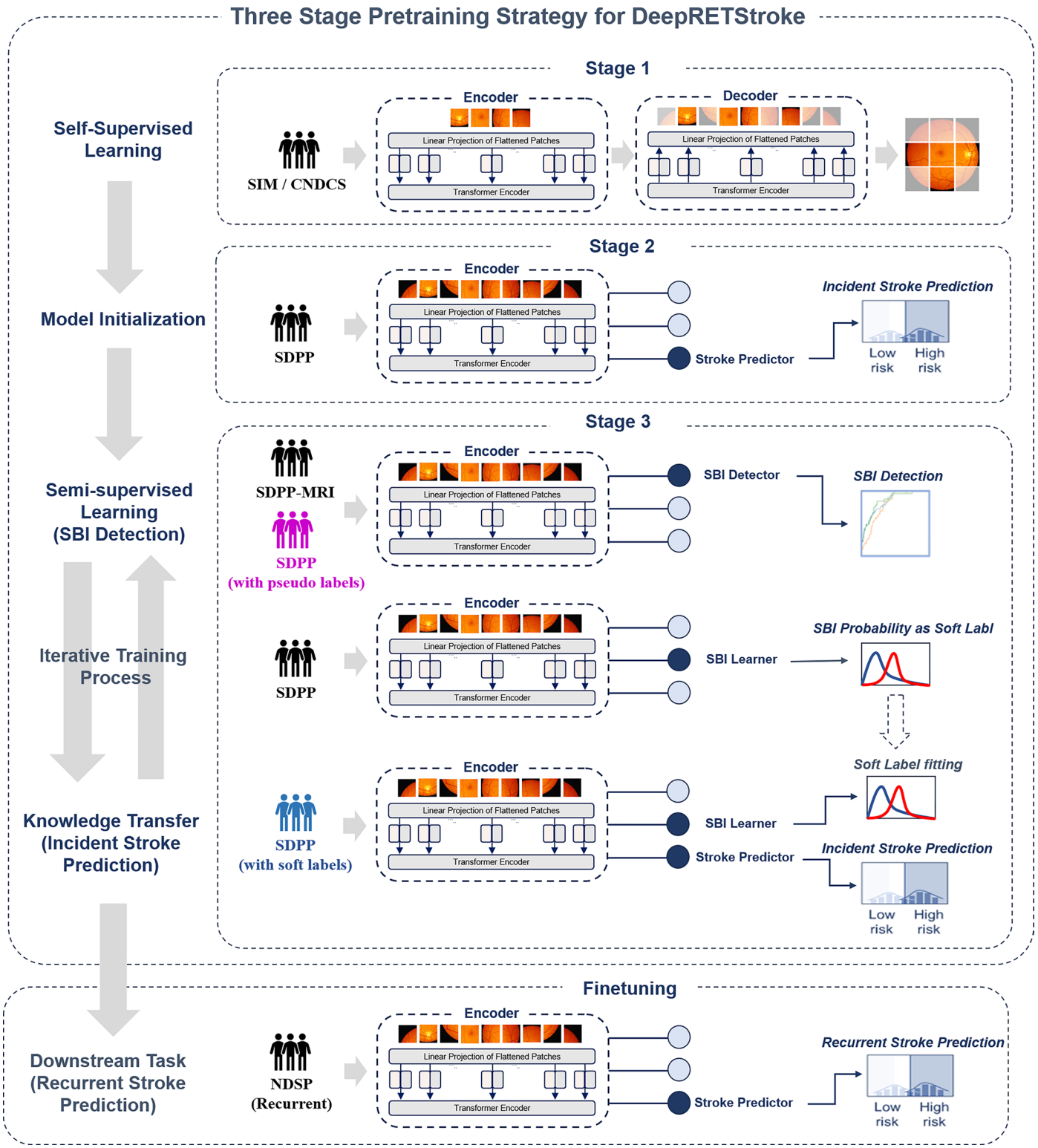


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41551-025-01413-9

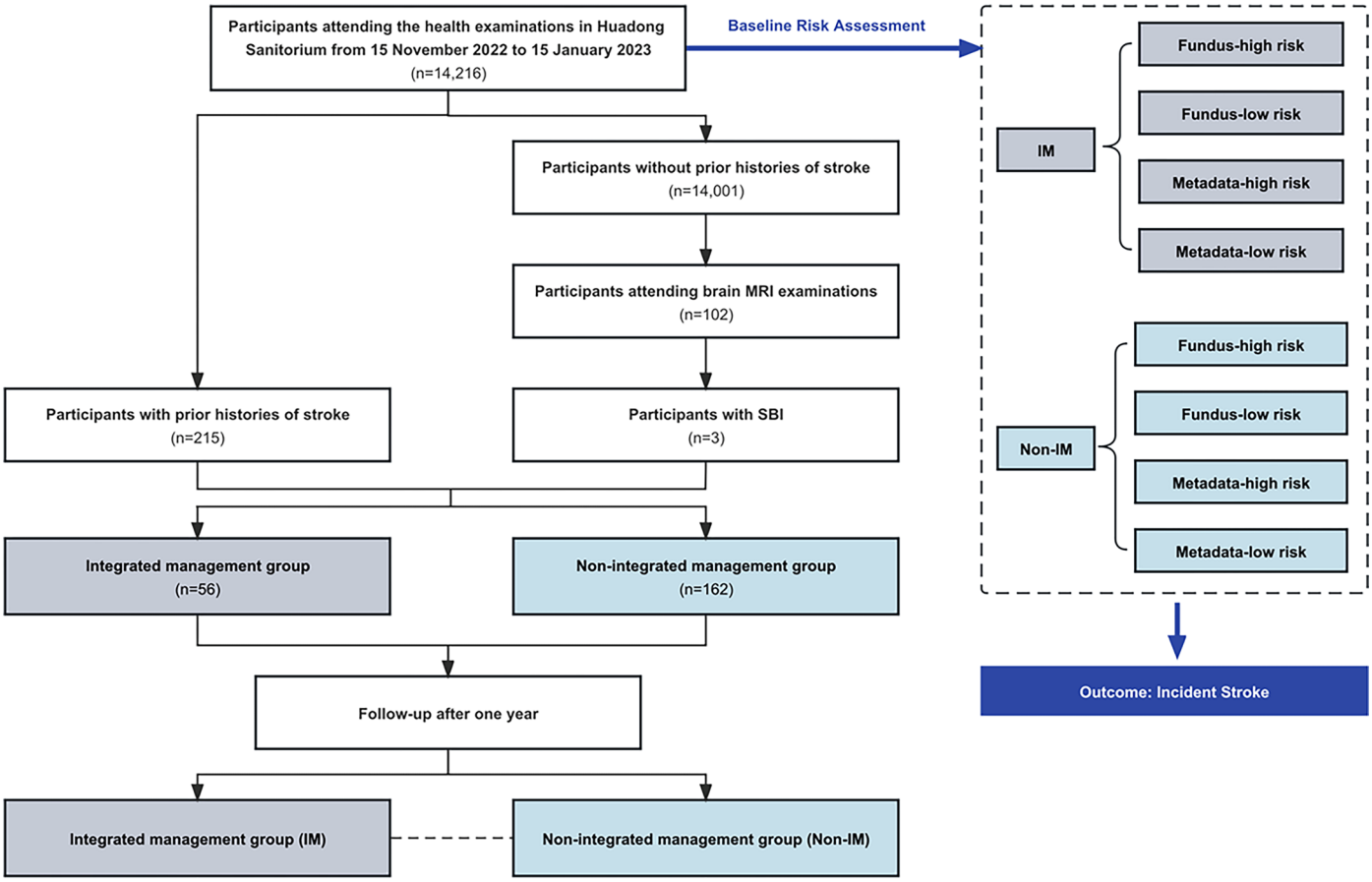
Extended Data Fig. 3 | Explainability Analysis of DeepRETStroke System in Predicting Future Stroke. Left Panel: Original color fundus images of participants. Middle Panel: SHAP (Shapley Additive Explanations) summary plot of fundus images. The pixel color from green to blue indicates the increasing expectation of gradients for stroke prediction. Right Panel: Occlusion-based attribution plot of fundus images. Red pixels reveal the areas with positive attribution highlighting possible pathological changes. For the prediction of new onset stroke, some anatomical structures associated with specific diseases, for example retinal vasculature, are highlighted.

https://doi.org/10.1038/s41551-025-01413-9



Extended Data Fig. 4 | The Workflow of the Development of DeepRETStroke System. The DeepRETStroke system, which was built with a three-stage pretraining strategy, includes an encoder, an SBI Detector, an SBI Learner and a Stroke Predictor. For the first stage, a retinal feature encoder from RETFound were employed to perform self-supervised pretraining on SIM and CNDCS dataset. For the second stage, SDPP dataset was employed to roughly train the encoder and Stroke Predictor. For the third stage, SDPP-MRI dataset was employed along with SDPP dataset to train SBI Detector by semi-supervised learning strategy. The developed SBI Detector was then used to generate the 'soft label' for each sample in SDPP dataset. Then, the encoder, SBI Learner and Stroke Predictor were jointly trained on samples in SDPP dataset with their 'soft label' of SBI and ground truth label of future stroke. Finally, the developed encoder and Stroke Predictor were fine-tuned on population of NDSP dataset with stroke history to develop a specific model for recurrent stroke prediction. Of note, SBI detection and incident stroke prediction can be iteratively repeated for further optimization of the entire system. Here we highlighted the modules trained in each stage to provide a clearer explanation of the development process.

https://doi.org/10.1038/s41551-025-01413-9



Extended Data Fig. 5 | Study Design of the Prospective Real-world Study. For the real-world study within a community-based prospective cohort study of Chinese adults, 215 participants with prior stroke and 3 participants with SBI were screened in November 2022. Among these patients, 56 received integrated management (IM), while 162 did not. IM group were provided regular clinical and metabolic measurements, advised by specialists in comprehensive hospitals, received lifestyle guidance and peer support at community health service centers.

https://doi.org/10.1038/s41551-025-01413-9

Extended Data Table 1 | Performance of DeepRETStroke System for Detection of Silent Brain Infarction

Datasets	Positive/ Negative cases	Model	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV (95% CI)	NPV (95% CI)	F1-Score (95% CI)
Internal	5 / 237	Metadata only	0.633(0.533-0.748)	1.000(1.000-1.000)	0.523(0.462-0.586)	0.042(0.009-0.084)	1.000(1.000-1.000)	0.081(0.019-0.155)
SDPP		Fundus only	0.797(0.500-0.995)	0.800(0.333-1.000)	0.781(0.730-0.830)	0.071(0.018-0.146)	0.995(0.983-1.000)	0.131(0.034-0.254)
		Combined	0.828(0.694-0.983)	1.000(1.000-1.000)	0.679(0.620-0.737)	0.062(0.014-0.123)	1.000(1.000-1.000)	0.116(0.029-0.220)
External-1	40 / 398	Metadata only	0.565(0.463-0.663)	0.500(0.333-0.649)	0.648(0.601-0.695)	0.125(0.075-0.175)	0.928(0.896-0.956)	0.200(0.122-0.271)
WTHM		Fundus only	0.751(0.676-0.817)	0.775(0.648-0.895)	0.691(0.644-0.739)	0.201(0.137-0.266)	0.968(0.947-0.986)	0.320(0.230-0.404)
		Combined	0.788(0.712-0.859)	0.800(0.667-0.918)	0.766(0.724-0.808)	0.256(0.180-0.333)	0.974(0.955-0.990)	0.388(0.292-0.477)
External-2	295 / 2493	Metadata only	0.647(0.611-0.681)	0.733(0.680-0.785)	0.509(0.489-0.528)	0.136(0.119-0.155)	0.948(0.936-0.960)	0.229(0.203-0.257)
PRECISE		Fundus only	0.778(0.749-0.805)	0.767(0.710-0.816)	0.717(0.698-0.735)	0.222(0.197-0.250)	0.967(0.958-0.975)	0.344(0.311-0.379)
		Combined	0.802(0.775-0.829)	0.783(0.733-0.835)	0.722(0.705-0.741)	0.229(0.200-0.257)	0.969(0.961-0.977)	0.354(0.318-0.389)
External-3	4 / 34	Metadata only*	0.537(0.000-0.917)	0.500(0.000-1.000)	0.853(0.730-0.969)	0.286(0.000-0.667)	0.935(0.833-1.000)	0.364(0.000-0.667)
MeLODY		Fundus only	0.757(0.486-0.946)	0.750(0.000-1.000)	0.824(0.686-0.939)	0.333(0.000-0.667)	0.966(0.885-1.000)	0.462(0.000-0.750)
		Combined	0.798(0.513-1.000)	1.000(1.000-1.000)	0.500(0.324-0.667)	0.190(0.048-0.368)	1.000(1.000-1.000)	0.320(0.091-0.538)
External-4	18 / 138	Metadata only	0.726(0.600-0.830)	0.722(0.500-0.923)	0.761(0.692-0.831)	0.283(0.161-0.415)	0.955(0.913-0.991)	0.406(0.245-0.535)
UKB		Fundus only	0.792(0.707-0.879)	0.889(0.733-1.000)	0.703(0.629-0.771)	0.281(0.164-0.400)	0.980(0.949-1.000)	0.427(0.277-0.556)
		Combined	0.811(0.733-0.881)	0.944(0.818-1.000)	0.667(0.586-0.746)	0.270(0.161-0.381)	0.989(0.966-1.000)	0.420(0.276-0.544)
External-5	1 / 126	Metadata only	0.603(0.516-0.686)	1.000(1.000-1.000)	0.603(0.516-0.686)	0.020(0.017-0.074)	1.000(1.000-1.000)	0.038(0.033-0.138)
I-OPTA		Fundus only	0.770(0.696-0.841)	1.000(1.000-1.000)	0.770(0.696-0.841)	0.033(0.026-0.120)	1.000(1.000-1.000)	0.065(0.051-0.214)
		Combined	0.810(0.738-0.873)	1.000(1.000-1.000)	0.810(0.738-0.873)	0.040(0.031-0.138)	1.000(1.000-1.000)	0.077(0.061-0.243)

Abbreviations: SDPP, Shanghai Diabetes Prevention Program; WTHM, Wuhan Tongji Health Management; PRECISE, Poly-vascular Evaluation for Cognitive Impairment and Vascular Events study; SEED, the Singapore Epidemiology of Eye Diseases study; MeLODY, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia Diabetes Registry in Malaysia cohort; UKB, UK biobank; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD; CI, confidence interval. PPV: positive predictive value; NPV: negative predictive value. *This metadata model used a constant 0 to represent smoking status as the smoking data is missing.

https://doi.org/10.1038/s41551-025-01413-9

Extended Data Table 2 | Performance of DeepRETStroke System for Prediction of Incident Stroke

Datasets	Events / No Event	Model	Prediction of Incident Stroke	
			AUC (95%CI)	C-index (95%CI)
Internal	39 / 5766	Metadata only	0.892(0.839-0.939)	0.887(0.836-0.933)
SDPP		Fundus only	0.901(0.846-0.940)	0.910(0.853-0.950)
		Combined	0.908(0.856-0.954)	0.885(0.821-0.937)
External-1	42 / 5928	Metadata only	0.855(0.791-0.907)	0.859(0.798-0.907)
ECHM		Fundus only	0.895(0.799-0.958)	0.875(0.777-0.942)
		Combined	0.908(0.866-0.942)	0.909(0.866-0.942)
External-2	177 / 5109	Metadata only	0.635(0.581-0.685)	0.608(0.552-0.657)
NDSP		Fundus only	0.756(0.708-0.802)	0.742(0.694-0.792)
		Combined	0.755(0.708-0.800)	0.740(0.685-0.790)
External-3	19 / 2255	Metadata only	0.728(0.611-0.836)	0.731(0.601-0.847)
WTHM		Fundus only	0.780(0.674-0.854)	0.864(0.785-0.917)
		Combined	0.812(0.720-0.875)	0.874(0.805-0.919)
External-4	10 / 281	Metadata only	0.559(0.368-0.761)	0.612(0.481-0.772)
PUDM		Fundus only	0.779(0.628-0.900)	0.745(0.620-0.875)
		Combined	0.794(0.521-0.962)	0.771(0.500-0.937)
External-5	20 / 481	Metadata only	0.611(0.493-0.740)	0.585(0.473-0.711)
CUHK-STDR		Fundus only	0.728(0.588-0.871)	0.701(0.540-0.854)
		Combined	0.727(0.588-0.863)	0.703(0.534-0.855)
External-6	84 / 6515	Metadata only	0.831(0.667-0.950)	0.831(0.656-0.950)
SEED		Fundus only	0.800(0.693-0.885)	0.879(0.797-0.947)
		Combined	0.801(0.732-0.855)	0.848(0.788-0.891)
External-7	247 / 5475	Metadata only*	0.532(0.475-0.583)	0.548(0.503-0.587)
MeLODY		Fundus only	0.749(0.703-0.794)	0.731(0.695-0.766)
		Combined	0.741(0.693-0.784)	0.779(0.747-0.808)
External-8	786 / 38055	Metadata only	0.683(0.641-0.722)	0.703(0.656-0.749)
UKB		Fundus only	0.808(0.773-0.838)	0.807(0.769-0.840)
		Combined	0.796(0.761-0.830)	0.812(0.770-0.851)
External-9	14 / 2950	Metadata only	0.312(0.155-0.533)	0.633(0.434-0.806)
NICOLA		Fundus only	0.767(0.565-0.907)	0.754(0.538-0.925)
		Combined	0.783(0.612-0.898)	0.807(0.708-0.911)
External-10	12 / 118	Metadata only	0.619(0.421-0.854)	0.567(0.431-0.888)
I-OPTA		Fundus only	0.821(0.638-0.920)	0.918(0.743-0.978)
		Combined	0.829(0.700-0.958)	0.748(0.598-0.964)
External-11	357 / 4078	Metadata only #	0.608(0.558-0.658)	0.585(0.526-0.651)
AREDS		Fundus only	0.782(0.747-0.817)	0.800(0.760-0.838)
		Combined	0.811(0.773-0.845)	0.838(0.796-0.876)

Abbreviations: SDPP: Shanghai Diabetes Prevention Program; ECHM, The Eastern China Health Management; NDSP, Nicheng Diabetes Screening Project; WTHM, Wuhan Tongji Health Management; PUDM: Peking Union Diabetes Management; CUHK-STDR, The Chinese University of Hong Kong-Sight-Threatening Diabetic Retinopathy; SEED, the Singapore Epidemiology of Eye Diseases study; MeLODY, the Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia Diabetes Registry in Malaysia cohort; UKB, UK biobank; I-OPTA, Identification of patient-reported barriers to treatment with anti-VEGF for neovascular AMD; AREDS, Age-Related Eye Disease Study; AUC: area uder the curve. CI: confidence interval. *This metadata model used a constant 0 to represent smoking status, as the smoking data is missing. #This metadata model used the mean high-density lipoprotein cholesterol value of our training population, as the data is missing.

https://doi.org/10.1038/s41551-025-01413-9

Extended Data Table 3 | Internal and External Validation of the DeepRETStroke system in the Prediction of Incident Stroke in Diabetes cohort, in Hypertension cohort, and Carotid Atherosclerosis cohort

Datasets	Model	Diabetes		Hypertension		Carotid atherosclerosis	
		With	Without	With	Without	With	Without
Internal	Metadata only	0.781(0.577-0.952)	0.900(0.839-0.949)	0.779(0.659-0.894)	0.954(0.909-0.988)	0.733(0.348-1.000)	0.924(0.885-0.957)
SDPP	Fundus only	0.910(0.808-0.983)	0.903(0.831-0.947)	0.879(0.797-0.939)	0.960(0.934-0.983)	0.905(0.797-0.984)	0.901(0.834-0.944)
	Combined	0.872(0.770-0.989)	0.913(0.857-0.963)	0.905(0.832-0.961)	0.893(0.774-0.983)	0.921(0.816-0.994)	0.908(0.842-0.962)
External-1	Metadata only	0.717(0.575-0.841)	0.855(0.763-0.921)	0.766(0.650-0.861)	0.882(0.788-0.938)	0.752(0.536-0.940)	0.702(0.573-0.828)
ECHM	Fundus only	0.823(0.559-0.990)	0.935(0.901-0.960)	0.876(0.727-0.965)	0.946(0.879-0.979)	0.909(0.773-1.000)	0.742(0.594-0.831)
	Combined	0.864(0.783-0.931)	0.913(0.856-0.955)	0.896(0.832-0.943)	0.917(0.869-0.957)	0.935(0.829-1.000)	0.783(0.660-0.871)
External-2	Metadata only	0.613(0.521-0.704)	0.626(0.553-0.689)	0.598(0.526-0.666)	0.581(0.479-0.676)	/	/
NDSP	Fundus only	0.738(0.655-0.816)	0.760(0.695-0.818)	0.756(0.691-0.809)	0.769(0.684-0.842)	/	/
	Combined	0.755(0.666-0.836)	0.755(0.699-0.811)	0.749(0.686-0.810)	0.769(0.694-0.837)	/	/
External-3	Metadata only	0.718(0.103-0.915)	0.731(0.616-0.852)	0.660(0.545-0.775)	0.649(0.437-0.859)	0.806(0.648-0.922)	0.844(0.763-0.908)
WTHM	Fundus only	0.725(0.000-0.933)	0.792(0.683-0.866)	0.758(0.622-0.865)	0.794(0.511-0.927)	0.812(0.504-0.997)	0.935(0.888-0.966)
	Combined	0.725(0.443-0.862)	0.829(0.721-0.909)	0.792(0.668-0.889)	0.833(0.672-0.938)	0.899(0.817-0.963)	0.911(0.854-0.953)
External-4	Metadata only	0.653(0.610-0.699)	0.630(0.537-0.719)	0.641(0.543-0.753)	0.670(0.626-0.713)	/	/
UKB	Fundus only	0.833(0.804-0.863)	0.702(0.595-0.792)	0.810(0.734-0.878)	0.807(0.771-0.839)	/	/
	Combined	0.788(0.743-0.829)	0.809(0.755-0.862)	0.848(0.781-0.899)	0.785(0.744-0.826)	/	/

Area under the curves were presented in the table. ***Abbreviations***: SDPP: Shanghai Diabetes Prevention Program; ECHM, The Eastern China Health Management; NDSP, Nicheng Diabetes Screening Project; WTHM, Wuhan Tongji Health Management; UKB: UK biobank.

https://doi.org/10.1038/s41551-025-01413-9

Extended Data Table 4 | Characteristics of the Developmental, Internal and External Validation Datasets for the Prediction of Recurrent Stroke

	Developmental	Internal	External-1
		NDSP	MeLODY
Fundus images	280	94	384
Participants	140	47	192
Race, n (%)			
Chinese	140(100.0)	47(100.0)	55(28.6)
Indian			62(32.3)
Malay			74(38.5)
Other*			1(0.5)
Age (years)	63.2±4.2	63.3±3.6	61.3±10.3
Female, n (%)	60(42.9)	25(53.2)	115(59.9)
Smoking, n (%)	24(17.1)	10(21.3)	-
BMI (kg/m2)	25.5±3.4	26.5±3.7	28.4±3.8
Systolic BP (mmHg)	138.3±16.4	136.0±16.1	140.6±17.8
Diastolic BP (mmHg)	83.5±8.9	83.7±7.0	76.4±8.5
FPG (mmol/L)	6.8±2.0	6.4±1.2	9.5±6.0
HbA1c (%)	6.2±1.1	5.9±0.6	8.2±2.1
TC (mmol/L)	5.3±0.8	5.0±0.9	4.6±1.0
TG (mmol/L)	1.8±2.3	1.6±0.9	1.7±0.9
HDL-C (mmol/L)	1.3±0.3	1.3±0.3	1.2±0.3
LDL-C (mmol/L)	3.2±0.7	3.0±0.7	2.6±1.0
Hypertension, n (%)	110(78.6)	37(78.7)	46(24.0)
Antihypertensives, n (%)	85(60.7)	30(63.8)	45(23.4)
Diabetes, n (%)	53(37.9)	17(36.2)	192(100.0)
Recurrent stroke, n (%)	33(23.6)	10(21.3)	11(5.7)
Follow-up, years	4.0±1.3	4.2±0.9	1.8±1.3

Data are presented as mean± SD or n (%) as appropriate. `-` means not available. ***Abbreviations***: NDSP, Nicheng Diabetes Screening Project; MeLODY, Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia. BMI: body mass index; FPG: fasting plasma glucose; BP: blood pressure; TC: total cholesterol; TG: triglyceride; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol

https://doi.org/10.1038/s41551-025-01413-9

Extended Data Table 5 | Performance of DeepRETStroke system for Predicting Recurrent Stroke

Datasets	Model	Prediction of recurrent stroke	
		AUC (95%CI)	C-index (95%CI)
Internal	Metadata only	0.568(0.143-0.924)	0.524(0.293-0.730)
NDSP	Fundus only	0.769(0.375-1.000)	0.782(0.671-0.889)
	Combined	0.833(0.500-1.000)	0.802(0.667-0.922)
External-1	Metadata only*	0.705(0.348-1.000)	0.384(0.204-0.575)
MeLODY	Fundus only	0.727(0.333-1.000)	0.797(0.683-0.914)
	Combined	0.773(0.444-1.000)	0.763(0.570-0.904)

Abbreviations: NDSP, Nicheng Diabetes Screening Project; MeLODY: Multiethnic Lifestyle, Obesity, and Diabetes Registry in Malaysia, AUC: area under the curve. *This metadata model used a constant 0 to represent smoking status as the smoking data is missing.

nature portfolio

Corresponding author(s): Tien Yin Wong

Last updated by author(s): Mar 29, 2025

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	No special software was used for data collection
Data analysis	Python version 3.7.5 (Python Software foundation, Delaware, United States) was used for all statistical analyses in the study. The code being used in the current study for developing the algorithm is provided via GitHub at https://github.com/njiang2024/DeepRETStroke.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The main data supporting the results in this study are available within the paper and its Supplementary Information. Individual-level patient data are protected

because of patient privacy, which can be accessible with the consent of the data management committee from institutions and are not publicly available. Requests for the non-profit use of the retinal fundus images and related clinical information should be sent to T.Y.W. (wongtienyin@tsinghua.edu.cn). The data management committee will then review all the requests and grant (if successful). A formal data transfer agreement will be required upon approval. Generally, all these requests for access to the data will be responded to within 1 month. All data shared will be de-identified. Source data for Figs. 2 & 3 and Extended Data Fig. 1 are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	The sex information of all the internal and external cohorts was described in Table 1, Table 2, Extended Data Table 4, and Supplementary Table 1.
Reporting on race, ethnicity, or other socially relevant groupings	For the development and validation of DeepRETStroke's performance, a multi-ethnic dataset was used. The race and ethnicity information of all the internal and external cohorts was shown in Table 1, Table 2, Extended Data Table 4, and Supplementary Table 1.
Population characteristics	Detailed cohort characteristics were given in Table 1, Table 2, Extended Data Table 4, and Supplementary Table1.
Recruitment	The internal cohort, SDPP, was a longitudinal cohort, composed of 58,320 fundus images from 29,160 participants who had no stroke history at their first visit and underwent physical examination between 2015 and 2023. The data for the model training collected from Chinese subjects, might not be representative for the generalized population, potentially introducing biases.
Ethics oversight	The study was approved by the Ethics Committee of Shanghai Sixth People's Hospital 2023-KY-023 (K) and conducted in accordance with the Declaration of Helsinki.

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	A specific sample size calculation was not done. The internal cohort, SDPP, was a longitudinal cohort, composed of 58,320 fundus images from 29,160 participants who had no stroke history at their first visit and underwent physical examination between 2015 and 2023. The sample size was determined by the data availability.
Data exclusions	We selected the fundus images as follows: If more than 25% of the peripheral area of the retina was unobservable or the central region of the retina had significant artefacts that would affect analysis, the photograph will be excluded from the dataset.
Replication	DeepSTROKE system was validated in 5 external cohorts for SBI detection, and 11 external cohorts for stroke risk prediction.
Randomization	Samples were randomly allocated to the development and validation datasets.
Blinding	During the data processing, all data was first de-identified to remove any patient related information.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks	Not applicable.
Novel plant genotypes	Not applicable.
Authentication	Not applicable.