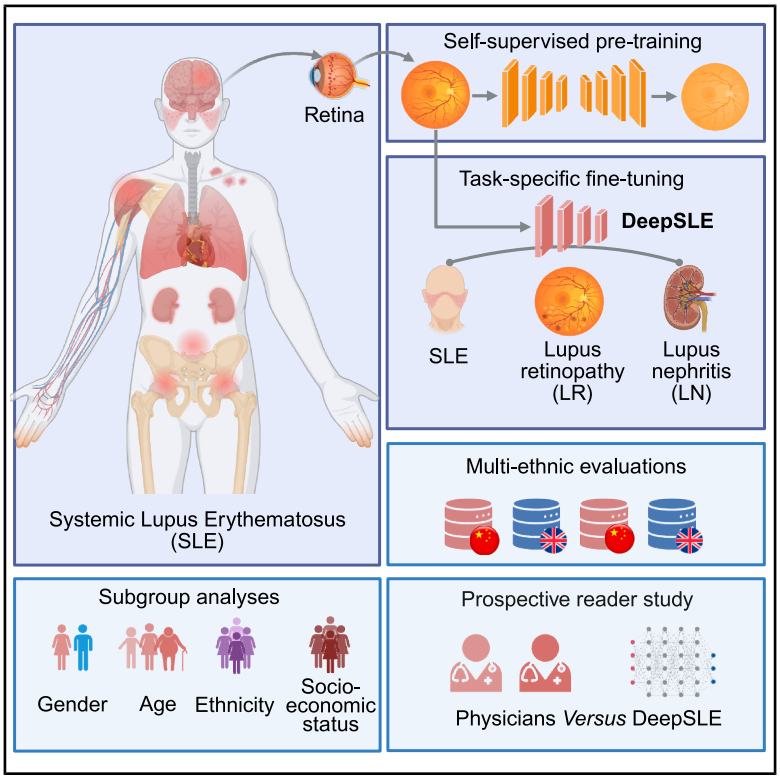
Cell Reports
Medicine

Article

A deep learning system for detecting systemic lupus erythematosus from retinal images

Graphical abstract



Highlights

- DeepSLE is a DL system to detect SLE and related complications from retinal images
- DeepSLE shows robustness and generalizability across multi-ethnic datasets
- DeepSLE can enable opportunistic SLE screening during routine retinal screening
- DeepSLE can be integrated into clinical workflows to assist SLE screening

Authors

Tingyao Li, Shiqun Lin, Zhouyu Guan, ..., Tien Yin Wong, Rongping Dai, Bin Sheng

Correspondence

derricka@sina.com (R.D.), shengbin@sjtu.edu.cn (B.S.)

In brief

Li et al. develop and validate a deep learning system (DeepSLE) to detect SLE and related complications using retinal fundus images solely among diverse multi-ethnic populations, offering a non-invasive and cost-effective digital solution for SLE screening in primary care and low-resource settings.

Li et al., 2025, Cell Reports Medicine 6, 102203
July 15, 2025
https://doi.org/10.1016/j.xcrm.2025.102203

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

A deep learning system for detecting systemic lupus erythematosus from retinal images

Tingyao Li,[1,2,19,22] Shiqun Lin,[3,20,22] Zhouyu Guan,[1,19,22] Yukun Zhou,[4,5,6,22] Dian Zeng,[7,22] Zheyuan Wang,[1,2] Yan Zhou,[3,20] Pinqi Fang,[8] Shujie Yu,[1] Ruhan Liu,[9] Xiang Chen,[1,2] Yan-Ran (Joyce) Wang,[10] Yuwei Lu,[1,2,19] Jia Shu,[1,2] Yiming Qin,[1,2,7] Yiting Wu,[8] Yilan Wu,[7] Chan Wu,[3,20] Shangzhu Zhang,[11] Jie Shen,[8] Huating Li,[1,19] Tingli Chen,[1,12] Jin Li,[13] Yih-Chung Tham,[14,15,16] Charumathi Sabanayagam,[14,16] Ying Feng Zheng,[17] Siegfried K. Wagner,[5,6] Pearse A. Keane,[5,6] Tien Yin Wong,[7,14,17,18,21] Rongping Dai,[3,20,*] and Bin Sheng[1,2,19,23,*]

[1]Shanghai International Joint Laboratory of Intelligent Prevention and Treatment for Metabolic Diseases, Department of Computer Science and Engineering, School of Electronic, Information, and Electrical Engineering, 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, China
[2]MOE Key Laboratory of AI, School of Electronic, Information, and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
[3]Department of Ophthalmology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China
[4]Centre for Medical Image Computing, University College London, London, UK
[5]NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust, London, UK
[6]Institute of Ophthalmology, University College London, London, UK
[7]School of Clinical Medicine, Beijing Tsinghua Changgung Hospital, Tsinghua Medicine, Tsinghua University, Beijing, China
[8]Medical Records and Statistics Office, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China
[9]Furong Laboratory, Central South University, Changsha, Hunan, China
[10]School of Medicine, Stanford University, Stanford, CA, USA
[11]Department of Rheumatology and Clinical Immunology, Chinese Academy of Medical Sciences & Peking Union Medical College, National Clinical Research Center for Dermatologic and Immunologic Diseases (NCRC-DID), Ministry of Science & Technology, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital (PUMCH), Key Laboratory of Rheumatology and Clinical Immunology, Ministry of Education, Beijing, China
[12]Department of Ophthalmology, Shanghai Health and Medical Center, Wuxi, China
[13]Department of Ophthalmology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[14]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore
[15]Centre for Innovation and Precision Eye Health, and Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[16]Ophthalmology and Visual Science Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore
[17]Zhongshan Ophthalmic Center, Guangzhou, China
[18]Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital, Beijing, China
[19]Institute for Proactive Healthcare, Shanghai Jiao Tong University, Shanghai, China
[20]Beijing Key Laboratory of Fundus Diseases Intelligent Diagnosis & Drug/Device Development and Translation, Beijing, China
[21]Beijing Key Laboratory of Intelligent Diagnostic Technology and Devices for Major Blinding Eye Diseases, Tsinghua Medicine, Tsinghua University, Beijing, China
[22]These authors contributed equally
[23]Lead contact
*Correspondence: derricka@sina.com (R.D.), shengbin@sjtu.edu.cn (B.S.)
https://doi.org/10.1016/j.xcrm.2025.102203

SUMMARY

Systemic lupus erythematosus (SLE) is a serious autoimmune disorder predominantly affecting women. However, screening for SLE and related complications poses significant challenges globally, due to complex diagnostic criteria and public unawareness. Since SLE-related retinal involvement could provide insights into disease activity and severity, we develop a deep learning system (DeepSLE) to detect SLE and its retinal and kidney complications from retinal images. In multi-ethnic validation datasets comprising 247,718 images from China and UK, DeepSLE achieves areas under the receiver operating characteristic curve of 0.822–0.969 for SLE. Additionally, DeepSLE demonstrates robust performance across subgroups stratified by gender, age, ethnicity, and socioeconomic status. To ensure DeepSLE's explainability, we conduct both qualitative and quantitative analyses. Furthermore, in a prospective reader study, DeepSLE demonstrates higher sensitivities compared with primary care physicians. Altogether, DeepSLE offers digital solutions for detecting SLE and related complications from retinal images, holding potential for future clinical deployment.

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

INTRODUCTION

Systemic lupus erythematosus (SLE) is a severe autoimmune disorder that affects many organs, including the skin, joints, the central nervous system, eyes, and the kidney.[1–4] It affects approximately 3.4 million people worldwide. Of these individuals, an estimated 3 million are female.[5] Previous studies[6–8] have shown that women are several times more likely to develop SLE compared to men, with the peak onset typically occurring between the ages of 15 and 45. This gender disparity in SLE prevalence underscores the importance of understanding the unique challenges faced by female patients in both disease diagnosis and treatment. Additionally, the delayed diagnosis of SLE is highly prevalent due to its complex diagnostic criteria and public unawareness. Early detection of SLE and further therapeutic measures are critical to increase the probability of SLE remission and improve patient prognosis. However, the screening of SLE remains a global public health challenge due to the lack of widely accepted, standardized, non-invasive, and cost-effective screening tools for early detection, particularly among asymptomatic individuals or those with mild symptoms.

In people with SLE, it is indispensable to screen for SLE-related complications, including lupus retinopathy (LR) and lupus nephritis (LN).[9] Early identification and management of these complications can enhance personalized care and improve long-term outcomes for individuals with SLE. However, challenges remain. Retinal changes associated with LR can be subtle and may require advanced imaging techniques,[10] such as optical coherence tomography (OCT) or fundus fluorescein angiography. Current LN screening relies on urinalysis and serum creatinine measurement, while kidney biopsy remains the gold standard for diagnosing LN. Consequently, significant gaps persist in the screening of LR and LN, as these procedures are not routinely implemented in primary care and low-resource settings.

Retinal changes are common among patients with SLE, including cotton-wool spots, retinal hemorrhages, vascular tortuosity, hard exudates, arteriolar narrowing, and arteriovenous crossing changes.[11] Additionally, retinal changes tend to coincide predominantly with the active phase of SLE[12] and can also provide valuable insights into the disease activity and severity.[13,14] Thus, the retina could potentially serve as a non-invasive, point-of-care, and cost-effective biomarker for detecting SLE and related complications. Furthermore, digital retinal photography is now widely available at a low cost in primary care optometry and community settings.

Deep learning (DL) has been widely applied to retinal photographs for detecting various retinal diseases (such as diabetic retinopathy and age-related macular degeneration) and systemic diseases (such as diabetes, chronic kidney disease, and cardiovascular diseases). However, applying DL to detect SLE and related complications from fundus images is relatively rare in previous studies. Liu et al.'s work[15] focused on the detection of LR from ultra-wide-field fundus photography (UWF-FP) images. Lin et al.[16] explored the feasibility of automatic SLE screening and LR diagnosis using OCT images. However, several challenges remain in this domain. First, the retinal imaging modalities (UWF-FP or OCT) used by previous studies to detect SLE and LR were not widely available outside specialized eye clinics due to their expensive cost, but color fundus photography (CFP) is one of the most commonly available eye examinations in community settings. Second, they did not explore the detection of LN from retinal images. Third, as a disease that primarily affects women of childbearing age, women with African background are disproportionately affected by SLE with more rapid progression and worse outcomes.[17] Previous studies did not investigate the performance and fairness of their models in detecting SLE across different gender, age, ethnicity, and socioeconomic subgroups, to democratize health access to underserved groups.

In this multi-center study, we developed, validated, and externally tested a vision transformer (ViT)-based DL system, termed DeepSLE, to detect SLE from retinal fundus images solely and further explore the feasibility and potential for detecting related retinal and kidney complications as a proof of concept. First, 666,383 fundus images from 173,346 participants were used to pre-train the DeepSLE system. Subsequently, we trained and validated our DeepSLE system using retinal fundus images from diverse multi-ethnic datasets, which comprise more than 254,246 images collected from 91,598 participants, from China and the UK. We further evaluated the model performance of detecting SLE in different subgroups stratified by gender, age, ethnicity, and socioeconomic status. To ensure the relevance and interpretability of DeepSLE, we also conducted saliency analysis and analysis of retinal vascular variables to gain insights into its diagnostic mechanisms of SLE detection. Furthermore, a prospective reader study was conducted to compare the performance of DeepSLE with primary care physicians and specialists.

RESULTS

The overall study design is shown in Figure 1. The DeepSLE system was first pre-trained using 666,383 fundus images of 173,346 participants from the Shanghai Integration Model (SIM) cohort through self-supervised learning. Subsequently, it was developed and validated using a total of 254,246 retinal fundus images from 91,598 participants from China and the UK on three disease detection tasks including SLE, LR, and LN. We included retinal images from the Peking Union Medical College Hospital (PUMCH) dataset for development and internal validation. Four independent multi-ethnic datasets were included for external validation: the Shanghai Six People's Hospital (SSPH) dataset, the Shanghai Diabetes Prevention Program (SDPP) dataset, the United Kingdom Biobank (UKB) dataset, and the Moorfields Eye Hospital (MEH) dataset. The detailed demographics information and retinal images of the included datasets are summarized in Tables 1 and S1 and STAR Methods section.

Performance of DeepSLE for detecting SLE and SLE-related retinal and kidney complications

For detecting SLE (Figure 2A; Table 2), the DeepSLE achieved an area under the receiver operating characteristic curve (AUROC) of 0.969 (95% confidence interval [CI], 0.950–0.984), with a high sensitivity of 92.1% (95% CI, 88.0%–95.8%) and a high specificity of 93.3% (95% CI, 90.8%–95.6%) in the internal validation

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

CellPress
OPEN ACCESS

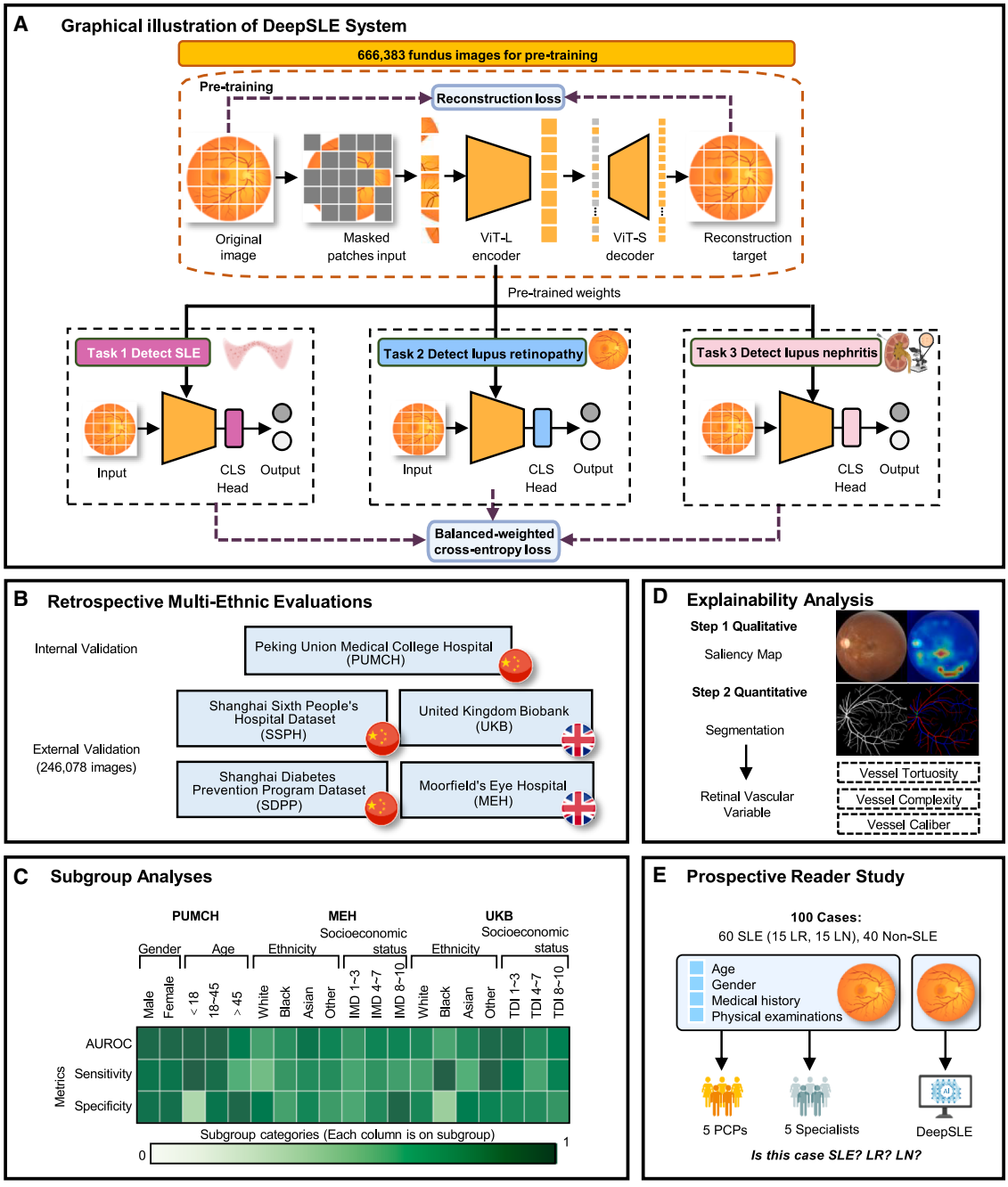


Figure 1. Overview study design of the DeepSLE system

(A) Graphical illustration of the DeepSLE system. The DeepSLE system could conduct three clinical tasks using retinal fundus images as inputs, including the detection of systemic lupus erythematosus (SLE), lupus retinopathy (LR), and lupus nephritis (LN). The DeepSLE system was first pre-trained in a self-supervised learning way to extract transferable visual representations from retinal fundus images and then adapted to these three clinical tasks.

(B) Retrospective multi-ethnic evaluations of the DeepSLE system in the internal and external validation datasets. Four external validation datasets from China and the UK were included.

(C) Subgroup analyses of the DeepSLE system for SLE detection. To ensure the fairness of the DeepSLE system, we conducted analyses on a wide range of patient subgroups, with respect to gender and age in the internal validation set and ethnicity and socioeconomic status in external test sets.

(D) Explainability analysis. To better understand how the DeepSLE system could detect SLE, LR, and LN, we conducted both qualitative and quantitative analyses to ensure the relevance and interpretability of the resulting features.

(E) Prospective reader study. We conducted a prospective reader study to compare the performance of the DeepSLE system with that of physicians for detecting SLE, LR, and LN in primary care settings. Five primary care physicians and five immunology specialists were recruited.

Figure 1 was created with https://BioRender.com.

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

Table 1. Characteristics of the participants for development and validation datasets

Cohorts	Pre-trained dataset (SIM)	Development (PUMCH)	Internal (PUMCH)	External-1 (SSPH)	External-2 (SDPP)	External-3 (UKB)	External-4 (MEH)
Images	666,383	6,528	1,640	960	133,646	111,101	371
Eyes	NA	4,435	1,133	480	66,823	107, 949	369
Participants	173,346	2,503	626	251	33,611	54,319	288
Gender							
Female, *n* (%)	55.40%	1,893 (78.59%)	492 (78.59%)	101 (40.24%)	12,709 (37.81%)	29,904 (55.05%)	187 (64.93%)
Male, *n* (%)	44.60%	582 (25.25%)	134 (21.41%)	144 (57.37%)	20,902 (62.19%)	24,415 (44.95%)	101 (35.07%)
Age (SD), years	66.22 (7.76)	44.45 (20.03)	43.40 (20.16)	50.42 (13.60)	46.94 (11.79)	57.21 (8.27)	66.19 (12.29)
Patients with SLE, *n* (%)	NA	756 (30.2%)	191 (30.5%)	29 (11.55%)	10 (0.03%)	54 (0.10%)	144 (50.0%)
Patients with LN in SLE, *n* (%)	NA	86 (11.38%)	26 (13.61%)	29 (100%)	NA	NA	NA
Eyes from participants with SLE, *n*	NA	1,471	370	53	20	102	NA
Eyes with LR in SLE, *n* (%)	NA	332 (21.89%)	83 (22.43%)	6 (10.91%)	3 (15%)	5 (4.90%)	NA
Patients with LR in SLE	NA	220 (29.10%)	60 (31.41%)	5 (17.24%)	2 (20%)	5 (9.26%)	NA
Patients with LR and LN in SLE	NA	35 (4.63%)	10 (5.24%)	5 (17.24%)	NA	NA	NA

SIM, Shanghai Integration Model; PUMCH, Peking Union Medical College Hospital; SSPH, Shanghai Six People's Hospital; SDPP, Shanghai Diabetes Prevention Program dataset; UKB, the United Kingdom Biobank; MEH, Moorfields Eye Hospital dataset; SLE, systemic lupus erythematosus; LR, lupus retinopathy; LN, lupus nephritis; NA, not available.

set. On the external validation datasets, our system achieved AUROCs of 0.850 (95% CI, 0.756–0.927) on external validation set 1, 0.836 (95% CI, 0.695–0.957) on external validation set 2, 0.861 (95% CI, 0.796–0.919) on external validation set 3, and 0.822 (95% CI, 0.766–0.869) on the external validation set 4. These results demonstrated that the DeepSLE system can accurately detect the presence of SLE using retinal fundus images only, across multi-ethnic datasets.

For detecting SLE-related retinal and kidney complications (Figure S1; Table S2), the DeepSLE system achieved an AUROC of 0.872 (95% CI, 0.825–0.911) for detecting LR in the internal set. Further validation on external datasets demonstrated its generalizability by achieving AUROCs of 0.962, 0.989, and 0.847 respectively, with superior sensitivities of 100% in external sets 1 and 2 and 95.0% in external set 3. For detecting LN, the system achieved an AUROC of 0.851 (95% CI, 0.759–0.926) in the internal dataset. On the external set 1, where all the patients with SLE had LN, DeepSLE achieved a sensitivity of 75.9% (95% CI, 62.0%–89.7%).

Subgroup analyses stratified by gender, age, ethnicity, and socioeconomic status

To ensure the fairness of the DeepSLE system, we conducted analyses on a wide range of patient subgroups, with respect to gender and age in the internal validation set and ethnicity and socioeconomic status in external validation sets. As illustrated in Figure 2B and Table 3, we evaluated performance for detecting SLE across different subgroups stratified by gender and age in the internal validation set. The AUROCs between the female group (0.970, 95% CI, 0.952–0.985) and the male group (0.962, 95% CI, 0.901–0.999) showed no significant differences. Additionally, there were no significant differences in sensitivities and specificities between these two subgroups. We also categorized the participants into three age subgroups: under 18 years (subgroup 1), 18–45 years (subgroup 2), and over 45 years (subgroup 3). DeepSLE achieved AUROCs of 0.979 (95% CI, 0.941–1.000), 0.977 (95% CI, 0.959–0.990), and 0.879 (95% CI, 0.787–0.957) in the three age subgroups.

We also compared DeepSLE's performance across different subgroups stratified by ethnicity and socioeconomic status in the UKB dataset (Figure 2C; Table 3) and the MEH dataset (Figure 2D; Table 3). The prevalence of SLE was the highest for the Black ethnicity subgroup (0.649% in the UKB dataset and 81.3% in the MEH dataset), while the White ethnicity subgroup had the lowest prevalence (0.073% in the UKB dataset and 33.6% in the MEH dataset). There were no significant differences in AUROCs among the Black, White, and Asian ethnicity subgroups. We used two different measurements of socioeconomic status for subgroup analyses: Townsend Deprivation Index (TDI) for the UKB dataset and Index of Multiple Deprivation (IMD) for the MEH dataset. Participants were categorized into three subgroups based on the deciles of these two measurements: decile 1 to 3, decile 4 to 7, and decile 8 to 10. A higher TDI indicated greater deprivation, whereas a lower IMD reflects greater deprivation. There were no noticeable differences in DeepSLE's performance across people with varying socioeconomic status levels.

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

CellPress
OPEN ACCESS

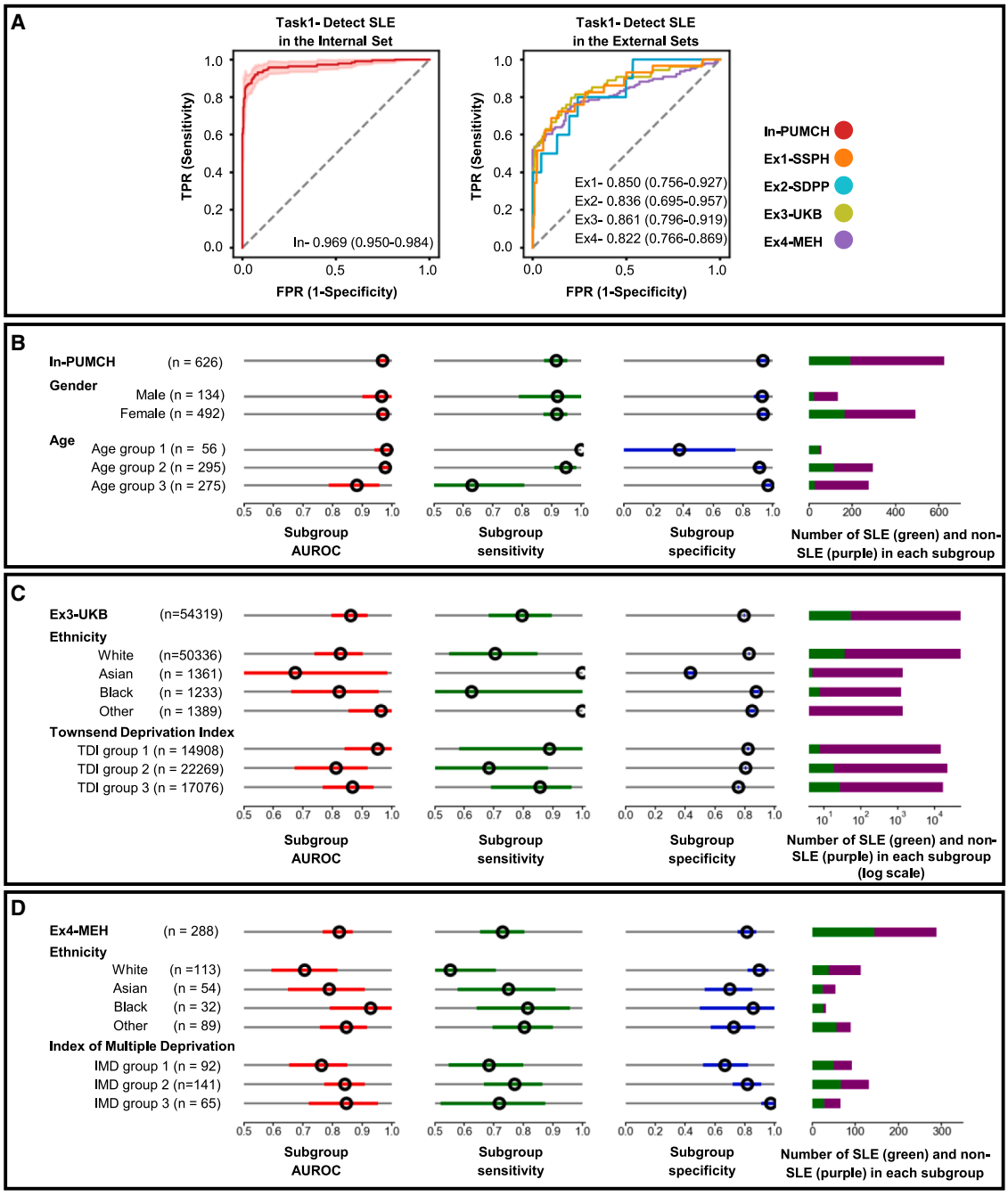


Figure 2. Performance of the DeepSLE system on validation sets and among different subgroups

Each subgroup was evaluated using three metrics: AUROC, sensitivity, and specificity. Green (SLE cases) and purple (non-SLE cases) bars represent the number of patients in each subgroup. Metrics are for all subgroups and are reported with 95% CIs calculated by bootstrapping with 1,000 replicates. PUMCH, Peking Union Medical College Hospital; SSPH, Shanghai Six People's Hospital dataset; SDPP, Shanghai Diabetes Prevention Program dataset; UKB, the United Kingdom Biobank; MEH, the Moorfields Eye Hospital dataset; SLE, systemic lupus erythematosus.

(A) Performance of the DeepSLE system on validation sets. Receiver operating characteristic (ROC) curves with area under the receiver operating characteristic curve (AUROC) of the DeepSLE system were shown for detecting SLE in the internal validation set and external validation sets.

(B) System performance for detecting SLE across demographic subgroups stratified by gender and age on the internal test set. Participants were categorized as under 18 years (subgroup 1), 18–45 years (subgroup 2), or over 45 years (subgroup 3).

(legend continued on next page)

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

Table 2. Performance of the DeepSLE system on detecting SLE across internal and external validation sets

	AUROC (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Internal set (PUMCH)	0.969 (0.950–0.984)	92.1 (88.0–95.8)	93.3 (90.8–95.6)
External set 1 (SSPH)	0.850 (0.756–0.927)	69.0 (50.0–85.2)	90.1 (86.0–94.1)
External set 2 (SDPP)	0.836 (0.695–0.957)	80.0 (50.0–100)	76.0 (75.5–76.3)
External set 3 (UKB)	0.861 (0.796–0.919)	79.6 (68.3–89.7)	79.5 (79.2–79.9)
External set 4 (MEH)	0.822 (0.766–0.869)	77.4 (72.6–82.3)	81.2 (75.2–87.6)

AUROC, area under the receiver operating characteristic curve. The performance was reported with 95% confidence interval using bootstrapping with 1,000 replicates. For analysis of sensitivity and specificity, we selected operating thresholds using the Youden index.
PUMCH, Peking Union Medical College Hospital; SSPH, Shanghai Six People's Hospital dataset; SDPP, Shanghai Diabetes Prevention Program dataset; UKB, the United Kingdom Biobank; MEH, Moorfields Eye Hospital dataset; SLE, systemic lupus erythematosus.

Explainability analyses of the DeepSLE system

The interpretability of the DeepSLE could shed insight into its diagnostic mechanism, which could enhance the trust of physicians in DeepSLE and thus enable broad adoption in real-world clinical practice. To better understand how the DeepSLE system could detect SLE, LR, and LN, we conducted both qualitative and quantitative analyses to ensure the relevance and interpretability of the resulting features.

We first conducted qualitative analyses by generating saliency maps to find out specific regions of retinal fundus images that were most important for DeepSLE's prediction. Representative examples of original images of SLE, LR, and LN, along with their corresponding saliency maps, were shown in Figure 3A. The average retinal images and corresponding average heatmaps for SLE were shown in Figure S2. These results showed that our DeepSLE system focused on the retinal vessels, the macula, and retinal lesions for disease detection. These patterns align with the retinal changes of people with SLE and related complications.

Considering SLE as a systemic disease may impact the clinically relevant morphological variables, and inspired by the findings from saliency maps, we further conducted a detailed quantitative analysis of widely accepted retinal vascular variables to evaluate their association with SLE, using an open-source automated retinal vascular morphology quantification tool.[18] First, we performed vessel segmentation on CFPs in various regions to get vascular contours (Figure 3B). Using the segmented images, we quantified a range of retinal vascular variables, including fractal dimension related to vessel complexity, distance tortuosity and squared curvature tortuosity related to vessel tortuosity, and central retinal arteriolar equivalent (CRAE) and central retinal venular equivalent (CRVE) related to vessel caliber.[19] As shown in Figure 3B and Table S3, for CFPs with SLE and without SLE, fractal dimension values showed significant differences between these two groups. Vessel tortuosity in Zone C also demonstrated significant differences between groups. Additionally, both CRAE and CRVE in Zone B and Zone C exhibited significant differences. These results showed that retinal vascular geometry might be predictive patterns for SLE detection.

Comparison of DeepSLE with physicians in a prospective reader study

To simulate the scenario of screening SLE and related complications in primary care settings, where specialists in rheumatology are not readily available, we conducted a prospective reader study to compare the performance of the DeepSLE system with that of physicians for detecting SLE, LR, and LN in primary care settings. Five primary care physicians (PCPs) and five rheumatology specialists were recruited and invited to make diagnosis of SLE, LR, and LN, based on age, gender, medical history, findings from physical examinations, and retinal fundus images (Figures 4A and S4), while the DeepSLE system made the diagnosis using retinal fundus images only as inputs. We included 60 cases with SLE (15 cases with LR and 15 cases with LN) and 40 non-SLE cases. As shown in Figure 4B and Table S4, PCPs achieved sensitivities from 50.00% to 55.00%, while rheumatology specialists showed higher sensitivities ranging from 65.00% to 73.33%. The DeepSLE system demonstrated a significantly higher sensitivity (98.33%, all $p < 0.001$) for detecting SLE from retinal fundus images only, compared with PCPs and rheumatology specialists. For LR detection (Figure S3A; Table S4), the DeepSLE system also showed a better sensitivity (93.33%) than PCPs (40.00%–66.67%) and rheumatology specialists (53.33%–80.00%). For LN detection (Figure S3B; Table S4), PCPs achieved sensitivities from 33.33% to 53.33%, while rheumatology specialists showed variations in sensitivities from 13.33% to 86.67%. The sensitivity of the DeepSLE system outperformed all PCPs and four rheumatology specialists.

DISCUSSION

SLE and related complications, such as LR and LN, are severe autoimmune diseases, affecting women of childbearing age disproportionally. Early detection and timely interventions of SLE and related complications are critical to increase the probability of SLE remission and improve patient prognosis. To address the gaps, we developed and validated a retinal image-based DL system (termed DeepSLE), to discriminate

(C) System performance for detecting SLE across demographic subgroups stratified by ethnicity and socioeconomic status on the UKB dataset. The socioeconomic status was measured by Townsend deprivation index deciles. Participants were categorized as least deprived (decile 1 to decile 3), moderately deprived (decile 4 to decile 7), or most deprived (decile 8 to decile 10).
(D) System performance for detecting SLE across demographic subgroups stratified by ethnicity and socioeconomic status on the MEH dataset. The socioeconomic status was measured by the Index of Multiple Deprivation deciles. Participants were categorized as most deprived (decile 1 to decile 3), moderately deprived (decile 4 to decile 7), and least deprived (decile 8 to decile 10).

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

Table 3. Subgroup analyses stratified by gender, age, ethnicity, and socioeconomic status for detecting SLE

Subgroup analyses	Number of SLE (%)	AUROC (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Gender (*n* = 626, PUMCH)				
Male (*n* = 134)	24 (17.8%)	0.962 (0.901–0.999)	91.7 (78.9–100)	92.7 (87.4–97.2)
Female (*n* = 492)	167 (33.9%)	0.970 (0.952–0.985)	91.6 (87.3–95.5)	93.5 (90.7–96.0)
Age (*n* = 626, PUMCH)				
Age ≤ 18 (*n* = 56)	48 (85.7%)	0.979 (0.941–1.000)	100.0 (100.0–100.0)	37.5 (0–75.0)
Age 18–45 (*n* = 295)	116 (39.3%)	0.977 (0.959–0.990)	94.8 (90.9–98.4)	91.1 (86.7–94.8)
Age > 45 (*n* = 275)	27 (9.8%)	0.879 (0.787–0.957)	63.0 (43.7–80.0)	96.8 (94.3–98.8)
Ethnicity (*n* = 54,319, UKB)				
White (*n* = 50,336)	37 (0.073%)	0.827 (0.739–0.902)	70.3 (55.0–84.8)	82.9 (82.6–83.3)
Asian (*n* = 1,361)	5 (0.367%)	0.674 (0.436–0.986)	100 (100.0–100.0)	43.7 (40.8–46.4)
Black (*n* = 1,233)	8 (0.649%)	0.820 (0.660–0.956)	62.5 (25.0–100)	87.7 (85.6–89.5)
Other (*n* = 1,389)	4 (0.288%)	0.962 (0.854–1.000)	100 (100.0–100.0)	84.8 (83.0–86.6)
Ethnicity (*n*=288, MEH)				
White (*n* = 113)	38 (33.6%)	0.706 (0.593–0.817)	55.3 (38.7–70.7)	89.3 (82.0–95.9)
Asian (*n* = 54)	24 (44.4%)	0.789 (0.650–0.910)	75.0 (57.7–90.9)	70.0 (53.1–85.2
Black (*n* = 32)	26 (81.3%)	0.917 (0.791–1.000)	80.0 (64.3–95.7)	83.3 (50.0–100)
Other (*n* = 89)	56 (62.9%)	0.845 (0.757–0.917)	80.4 (69.6–90.0)	72.7 (57.1–87.1)
Socioeconomic status (*n* = 54,253, UKB)[a]				
TDI 1–3 (*n* = 14,908)	8 (0.053%)	0.948 (0.841–1.000)	87.5 (58.3–100)	82.2 (81.6–82.9)
TDI 4–7 (*n* = 22,269)	19 (0.085%)	0.808 (0.672–0.919)	68.4 (46.7–88.3)	80.6 (80.0–81.1)
TDI 8–10 (*n* = 17,076)	27 (0.158%)	0.864 (0.767–0.940)	85.2 (69.0–96.4)	75.8 (75.2–76.4)
Socioeconomic status (*n*=288, MEH)				
IMD 1–3 (*n* = 92)	50 (54.3%)	0.760 (0.653–0.850)	68.0 (54.8–80.0)	66.7 (52.0–82.1)
IMD 4–7 (*n* = 141)	66 (50.4%)	0.843 (0.772–0.910)	77.3 (66.7–86.5)	81.5 (71.8–91.0)
IMD 8–10 (*n* = 65)	28 (43.1%)	0.846 (0.720–0.955)	71.4 (52.0–87.5)	97.3 (91.2–100)

The performance was reported with 95% confidence interval using bootstrapping with 1,000 replicates. For analysis of sensitivity and specificity, we selected operating thresholds using the Youden index.
PUMCH, Peking Union Medical College Hospital; UKB, the United Kingdom Biobank; MEH, Moorfields Eye Hospital dataset; SLE, systemic lupus erythematosus; TDI, Townsend deprivation index; IMD, Index of Multiple Deprivation.
[a]There were 66 participants among the included 54,319 participants in the UKB dataset without the Townsend deprivation index at recruitment.

people with SLE from people without SLE and further explore feasibility and potential for the detection of LR and LN among patients with SLE as a proof of concept, using retinal fundus images only. We demonstrated that the DeepSLE system could accurately detect SLE across datasets from China and the UK. Furthermore, we conducted a prospective reader study to demonstrate the potential of integrating DeepSLE into primary care workflows to assist PCPs in the screening of SLE and its complications.

Our proposed retinal image-based DL system provides a proof-of-concept digital solution to address the current gap in the detection of SLE and related complications, in which delayed diagnosis is highly prevalent.[20] Our proposed DeepSLE system could provide a non-invasive, low-cost, low-labor-dependent approach to detect SLE and related complications with good accuracy and sensitivity, serving as an auxiliary opportunistic screening tool that assists early detection of SLE, particularly in primary care settings where access to rheumatologic expertise or extensive laboratory testing is limited. The identified patients can then be referred to and followed up at tertiary medical centers with comprehensive diagnostic evaluations and subsequent multidisciplinary management. The detection of SLE and related complications based on retinal images could also leverage existing community eye care infrastructure that enables opportunistic SLE screening during routine screening for common eye diseases. Furthermore, with the growing popularity and adoption of digital retinal cameras and smartphone-based cameras in primary care worldwide, access to retinal photography is anticipated to increase,[21] which lays a solid foundation for the application of our DeepSLE system.

Several DL algorithms have been proposed for detecting SLE and related retinal complications using advanced imaging modalities, such as UWF-FP for LR and OCT for SLE. Our study improves upon previous work in several ways. First, we developed the DL algorithm to detect SLE using CFPs alone. Previous algorithms require expensive and specialized imaging modalities, which might not be feasible in community settings. By contrast, our algorithm could detect SLE based on retinal photographs only, thus improving the efficiency and potential cost-effectiveness of the algorithm. Second, we first explored the feasibility of using retinal fundus images alone to detect LN with promising

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

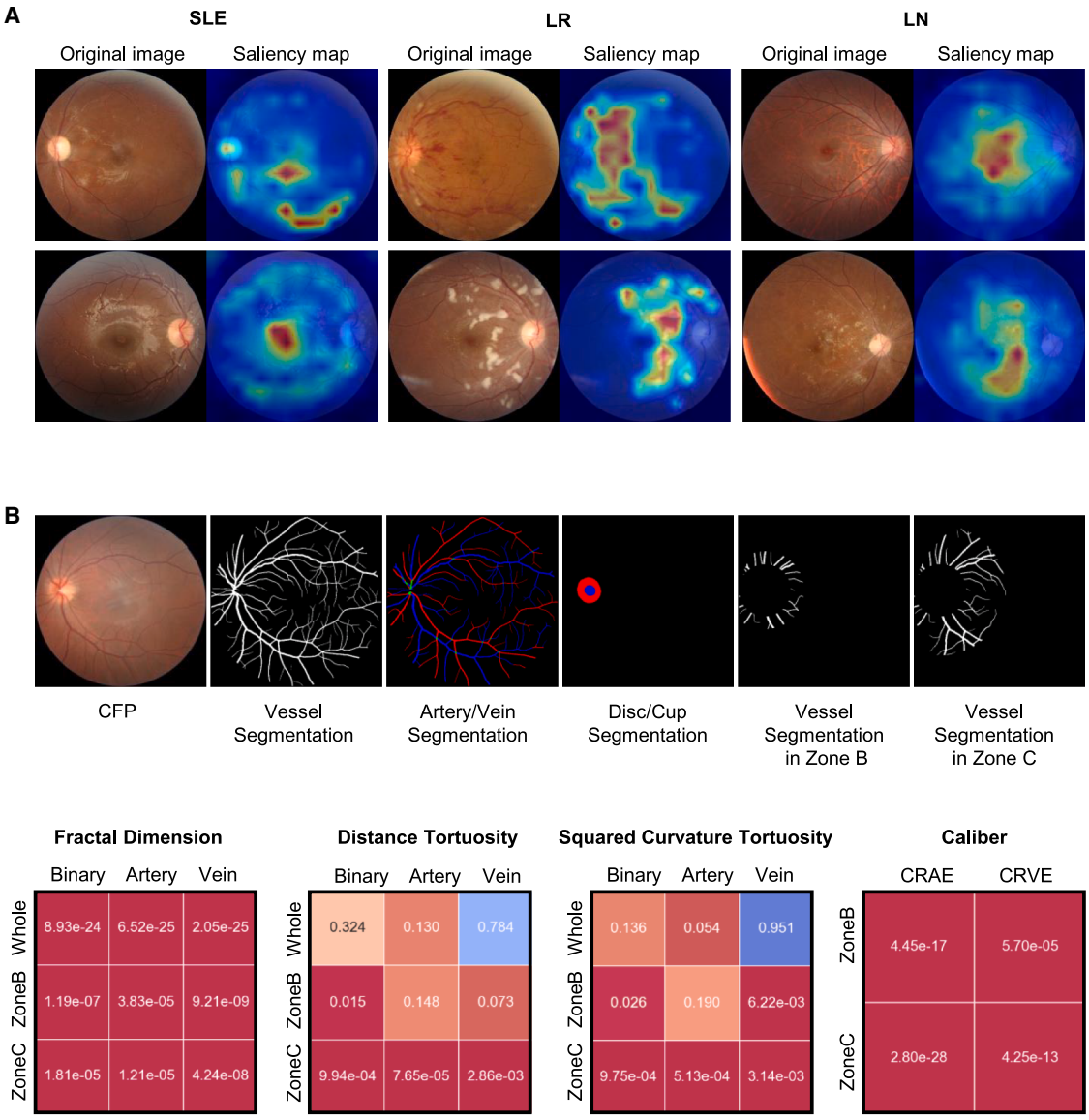


Figure 3. Explainability analyses of the DeepSLE system

(A) Qualitative analysis using saliency maps. The results showed that our DeepSLE system focused on the retinal vessels, the macula, and retinal lesions for disease detections.

(B) Quantitative analysis of retinal vascular variables for color fundus photographs with SLE and without SLE. First, we performed vessel segmentation on CFPs in various regions to get vascular contours. Using the segmented images, we quantified a range of retinal vascular variables, including fractal dimension related to vessel complexity, distance tortuosity and squared curvature tortuosity related to vessel tortuosity, and central retinal arteriolar equivalent (CRAE) and central retinal venular equivalent (CRVE) related to vessel caliber. *p* values for comparing the retinal vascular variables between CFPs with and without SLE were shown. *p* values were calculated using the Mann-Whitney U test.

performance. Third, we conducted comprehensive subgroup and explainability analyses to ensure the fairness and interpretability of our DeepSLE system, paving the way for future deployment in real-world clinical settings.

To warrant the model fairness of the DeepSLE system, we conducted analyses on a wide range of participant subgroups, dividing them with respect to gender, age, ethnicity, and socioeconomic status. As DL models are increasingly deployed in real-world clinical settings, it is indispensable to assess not only model performance but also potential biases toward specific demographic subgroups.[22] These subgroup analyses were designed to establish whether the DeepSLE system was underperforming in any of the evaluated categories and further investigate whether the system performed well for subgroups of participants who are underserved or at higher risk of SLE and related complications. The results of subgroup analyses indicated the DeepSLE system remained predictive within different subgroups stratified by

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

CellPress
OPEN ACCESS

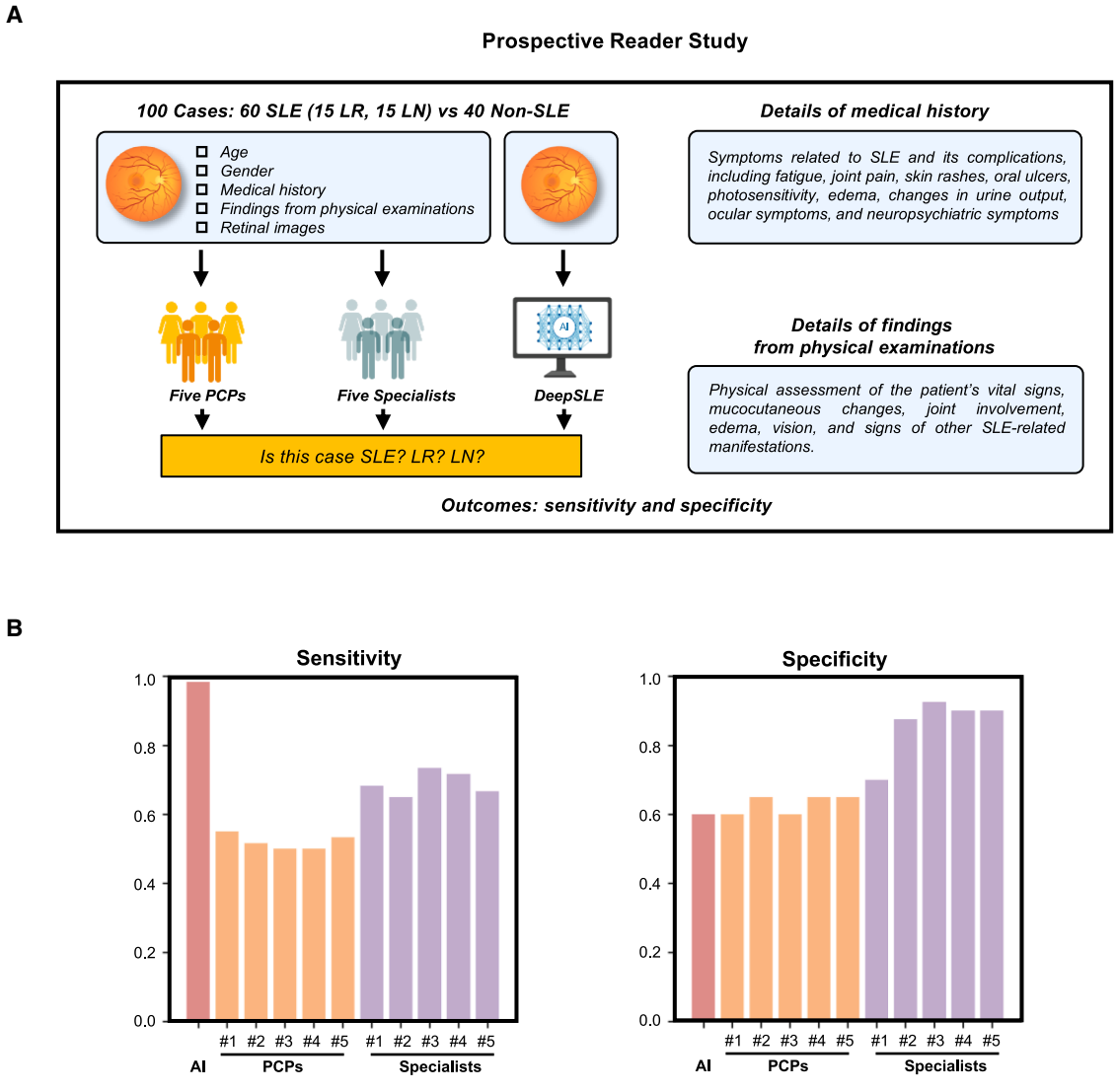


Figure 4. Comparison of DeepSLE with physicians in a prospective reader study

(A) Reader study design. Five primary care physicians and five immunology specialists were recruited to make diagnosis of SLE, LR, and LN, based on age, gender, medical history, findings from physical examinations, and retinal fundus images. The AI model (DeepSLE) was provided with only the retinal fundus photographs (created with https://BioRender.com).

(B) Comparison of DeepSLE's performance with that of PCPs and specialists for detecting SLE. The sensitivity of DeepSLE outperformed all the PCPs for detecting SLE, while no significant differences in specificity were observed between DeepSLE and the PCPs.

gender, age, ethnicity, or socioeconomic status. These findings demonstrate the reliability and equitable performance of the DeepSLE system, supporting its potential for broad clinical application globally.

The interpretability of the DeepSLE system offers valuable insights into its diagnostic mechanisms, which could facilitate its broader adoption in real-world clinical settings. Our qualitative analyses revealed that the DeepSLE system predominantly focuses on retinal vessels, the macula, and retinal lesions to make predictions. These retinal changes could be observed in people with SLE and related complications, demonstrating that the system aligns well with the established clinical knowledge.[10,23–25] Quantitative analyses further supported these findings, showing significant differences in retinal vascular variables such as fractal dimension, vessel tortuosity, CRAE, and CRVE between SLE and non-SLE groups. These results collectively suggest that the DeepSLE system is capable of detecting subtle retinal changes associated with SLE. The combination of qualitative and quantitative insights underscores the system's potential for accurate and explainable disease detection.

Furthermore, the prospective reader study underscores the potential of integrating the DeepSLE system into clinical workflows to assist PCPs in the screening of SLE and its related complications, in primary care and low-resource settings. Due to the variations in clinical presentations and lack of well-trained PCPs for autoimmune diseases, the delayed diagnosis of SLE and its

complications is highly prevalent worldwide. We demonstrated that, using retinal image only, the sensitivities of our DeepSLE system outperformed all the PCPs for detecting SLE, LR, and LN. Thus, the integration of DeepSLE into primary care workflows could potentially lead to fewer missed diagnoses of SLE and related complications. These findings suggest that our DeepSLE system could be utilized as a valuable, non-invasive, point-of-care, and cost-effective screening tool for physicians to screen SLE and related complications, particularly in challenging cases involving subtle retinal and systemic manifestations of SLE.

Limitations of the study

Our study had several limitations. First, our DeepSLE system was trained in the Chinese population and tested using datasets from China and the UK. Further validations in more diverse multiethnic populations will demonstrate its generalization and robustness. Second, the relatively limited sample size of LR and LN cases highlights the need for further evaluation to establish the model's generalizability and robustness. Third, some intrinsic biases cannot be eliminated in the present study design, such as data distribution and selection bias. Future large multicenter prospective studies[26] are needed to evaluate the patient outcomes after the integration of DeepSLE into real-world clinical settings.

In conclusion, we developed a single, non-invasive retinal image-based DL system (DeepSLE), to detect SLE and related complications that primarily affect women's health. DeepSLE could provide a rapid, non-invasive, and cost-effective screening tool for SLE and related complications, using retinal images solely.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Bin Sheng (shengbin@sjtu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- The code for the development of the DeepSLE system can be accessed via https://github.com/DeepSLE/deepsle.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2022YFC2407000), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0509202 and 2023ZD0509201), the National Natural Science Foundation of China (62272298), and Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102) to B.S.; the National High Level Hospital Clinical Research Funding (2022-PUMCH-B-101) to R.D.; and the National Key R&D Program of China (2022YFC2502800), the National Natural Science Foundation of China (82388101), and the Beijing Natural Science Foundation (IS23096) to T.Y.W. We would like to thank Prof Carol Y. Cheung for the fruitful discussions. The authors acknowledge the contributions of all the participants and investigators. This research has been conducted using the UK Biobank Resource under Application Number 104443.

AUTHOR CONTRIBUTIONS

B.S., R.D., T.Y.W., and P.A.K. conceived the idea for this study. T.L., S.L., Z.G., Yukun Zhou, and D.Z. collected and analyzed the data. All authors contributed to drafting and review of the paper. All authors agreed to the final version of this manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Study design
  - Study datasets and diagnostic criteria
  - Ethics statement
- METHOD DETAILS
  - Development and implementation of DeepSLE system
  - Explainability analyses
  - A prospective reader study to compare the performance of DeepSLE and physicians
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2025.102203.

Received: October 9, 2024
Revised: February 13, 2025
Accepted: May 30, 2025

REFERENCES

1. Kaul, A., Gordon, C., Crow, M.K., Touma, Z., Urowitz, M.B., Van Vollenhoven, R., Ruiz-Irastorza, G., and Hughes, G. (2016). Systemic lupus erythematosus. Nat. Rev. Dis. Primers *2*, 16039.
2. Gómez-Bañuelos, E., Goldman, D.W., Andrade, V., Darrah, E., Petri, M., and Andrade, F. (2024). Uncoupling interferons and the interferon signature explains clinical and transcriptional subsets in SLE. Cell Rep. Med. *5*, 101569. https://doi.org/10.1016/j.xcrm.2024.101569.
3. Kain, J., Owen, K.A., Marion, M.C., Langefeld, C.D., Grammer, A.C., and Lipsky, P.E. (2022). Mendelian randomization and pathway analysis demonstrate shared genetic associations between lupus and coronary artery disease. Cell Rep. Med. *3*, 100805. https://doi.org/10.1016/j.xcrm.2022.100805.
4. Jiang, S.H., Mercan, S., Papa, I., Moldovan, M., Walters, G.D., Koina, M., Fadia, M., Stanley, M., Lea-Henry, T., Cook, A., et al. (2021). Deletions in VANGL1 are a risk factor for antibody-mediated kidney disease. Cell Rep. Med. *2*, 100475. https://doi.org/10.1016/j.xcrm.2021.100475.
5. Siegel, C.H., and Sammaritano, L.R. (2024). Systemic Lupus Erythematosus: A Review. JAMA *331*, 1480–1491. https://doi.org/10.1001/jama.2024.2315.
6. Pons-Estel, G.J., Alarcón, G.S., Scofield, L., Reinlib, L., and Cooper, G.S. (2010). Understanding the Epidemiology and Progression of Systemic Lupus Erythematosus. Semin. Arthritis Rheum. *39*, 257–268.

7. Johnson, A.E., Gordon, C., Palmer, R.G., and Bacon, P.A. (1995). The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Arthritis Rheum. *38*, 551–558.
8. Yee, C.S., Su, L., Toescu, V., Hickman, R., Situnayake, D., Bowman, S., Farewell, V., and Gordon, C. (2015). Birmingham SLE cohort: Outcomes of a large inception cohort followed for up to 21 years. Rheumatology *54*, 836–843.
9. Fanouriakis, A., Kostopoulou, M., Andersen, J., Aringer, M., Arnaud, L., Bae, S.-C., Boletis, J., Bruce, I.N., Cervera, R., Doria, A., et al. (2024). EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. Ann. Rheum. Dis. *83*, 15–29. https://doi.org/10.1136/ard-2023-224762.
10. Dias-Santos, A., Tavares Ferreira, J., Pinheiro, S., Cunha, J.P., Alves, M., Papoila, A.L., Moraes-Fontes, M.F., and Proença, R. (2021). Retinal and choroidal thickness changes in systemic lupus erythematosus patients: a longitudinal study. Eye *35*, 2771–2780. https://doi.org/10.1038/s41433-020-01292-1.
11. Shoughy, S.S., and Tabbara, K.F. (2016). Ocular findings in systemic lupus erythematosus. Saudi J. Ophthalmol. *30*, 117–121. https://doi.org/10.1016/j.sjopt.2016.02.001.
12. Dammacco, R. (2018). Systemic lupus erythematosus and ocular involvement: an overview. Clin. Exp. Med. *18*, 135–149.
13. Gao, N., Li, M.T., Li, Y.H., Zhang, S.H., Dai, R.P., Zhang, S.Z., Zhao, L.D., Wang, L., Zhang, F.C., Zhao, Y., and Zeng, X.F. (2017). Retinal vasculopathy in patients with systemic lupus erythematosus. Lupus *26*, 1182–1189.
14. Hsiao, Y.P., Tsai, J.D., Muo, C.H., Tsai, C.H., Sung, F.C., Liao, Y.T., Chang, Y.J., and Yang, J.H. (2014). Atopic diseases and systemic lupus erythematosus: An epidemiological study of the risks and correlations. Int. J. Environ. Res. Public Health *11*, 8112–8122.
15. Liu, R., Wang, T., Li, H., Zhang, P., Li, J., Yang, X., Shen, D., and Sheng, B. (2023). TMM-Nets: Transferred Multi- to Mono-Modal Generation for Lupus Retinopathy Diagnosis. IEEE Trans. Med. Imag. *42*, 1083–1094.
16. Lin, S., Masood, A., Li, T., Huang, G., and Dai, R. (2023). Deep learning-enabled automatic screening of SLE diseases and LR using OCT images. Vis. Comput. *39*, 3259–3269.
17. Chae, D.H., Martz, C.D., Fuller-Rowell, T.E., Spears, E.C., Smith, T.T.G., Hunter, E.A., Drenkard, C., and Lim, S.S. (2019). Racial Discrimination, Disease Activity, and Organ Damage: The Black Women's Experiences Living With Lupus (BeWELL) Study. Am. J. Epidemiol. *188*, 1434–1443. https://doi.org/10.1093/aje/kwz105.
18. Zhou, Y., Wagner, S.K., Chia, M.A., Zhao, A., Woodward-Court, P., Xu, M., Struyven, R., Alexander, D.C., and Keane, P.A. (2022). AutoMorph: Automated Retinal Vascular Morphology Quantification Via a Deep Learning Pipeline. Transl. Vis. Sci. Technol. *11*, 12.
19. Parr, J.C., and Spears, G.F. (1974). General caliber of the retinal arteries expressed as the equivalent width of the central retinal artery. Am. J. Ophthalmol. *77*, 472–477. https://doi.org/10.1016/0002-9394(74)90457-7.
20. Kernder, A., Richter, J.G., Fischer-Betz, R., Winkler-Rohlfing, B., Brinks, R., Aringer, M., Schneider, M., and Chehab, G. (2021). Delayed diagnosis adversely affects outcome in systemic lupus erythematosus: Cross sectional analysis of the LuLa cohort. Lupus *30*, 431–438. https://doi.org/10.1177/0961203320983445.
21. Cheung, C.Y., Ran, A.R., Wang, S., Chan, V.T.T., Sham, K., Hilal, S., Venketasubramanian, N., Cheng, C.-Y., Sabanayagam, C., Tham, Y.C., et al. (2022). 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. https://doi.org/10.1016/S2589-7500(22)00169-8.
22. Yang, Y., Zhang, H., Gichoya, J.W., Katabi, D., and Ghassemi, M. (2024). The limits of fair medical imaging AI in real-world generalization. Nat. Med. *30*, 2838–2848.
23. Liu, G.Y., Utset, T.O., and Bernard, J.T. (2015). Retinal nerve fiber layer and macular thinning in systemic lupus erythematosus: an optical coherence tomography study comparing SLE and neuropsychiatric SLE. Lupus *24*, 1169–1176. https://doi.org/10.1177/0961203315582285.
24. Ushiyama, O., Ushiyama, K., Koarada, S., Tada, Y., Suzuki, N., Ohta, A., Oono, S., and Nagasawa, K. (2000). Retinal disease in patients with systemic lupus erythematosus. Ann. Rheum. Dis. *59*, 705–708.
25. Pelegrín, L., Morató, M., Araújo, O., Figueras-Roca, M., Zarranz-Ventura, J., Adán, A., Cervera, R., Casaroli-Marano, R.P., Budi, V., Barrera-López, L., et al. (2023). Preclinical ocular changes in systemic lupus erythematosus patients by optical coherence tomography. Rheumatology *62*, 2475–2482.
26. Li, S.W., Zhang, L.H., Cai, Y., Zhou, X.B., Fu, X.Y., Song, Y.Q., Xu, S.W., Tang, S.P., Luo, R.Q., Huang, Q., et al. (2024). Deep learning assists detection of esophageal cancer and precursor lesions in a prospective, randomized controlled study. Sci. Transl. Med. *16*, eadk5395. https://doi.org/10.1126/scitranslmed.adk5395.
27. Wagner, S.K., Hughes, F., Cortina-Borja, M., Pontikos, N., Struyven, R., Liu, X., Montgomery, H., Alexander, D.C., Topol, E., Petersen, S.E., et al. (2022). AlzEye: longitudinal record-level linkage of ophthalmic imaging and hospital admissions of 353 157 patients in London, UK. BMJ Open *12*, e058552. https://doi.org/10.1136/bmjopen-2021-058552.
28. Seth, G., Chengappa, K.G., Misra, D.P., Babu, R., Belani, P., Shanoj, K.C., Kumar, G., and Negi, V.S. (2018). Lupus retinopathy: a marker of active systemic lupus erythematosus. Rheumatol. Int. *38*, 1495–1501. https://doi.org/10.1007/s00296-018-4083-4.
29. Wilkinson, C.P., Ferris, F.L., 3rd, Klein, R.E., Lee, P.P., Agardh, C.D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., and Verdaguer, J.T.; Global Diabetic Retinopathy Project Group (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology *110*, 1677–1682. https://doi.org/10.1016/S0161-6420(03)00475-5.
30. He, K., Chen, X., Xie, S., Li, Y., Dollar, P., and Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 15979–15988. https://doi.org/10.1109/CVPR52688.2022.01553.
31. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., et al. (2023). A foundation model for generalizable disease detection from retinal images. Nature *622*, 156–163.
32. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2021). An Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale. ICLR 2021 - 9th Int. Conf. Learn. Represent. https://openreview.net/forum?id=YicbFdNTTy.
33. Chefer, H., Gur, S., and Wolf, L. (2021). Transformer Interpretability Beyond Attention Visualization. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 782–791. https://doi.org/10.1109/CVPR46437.2021.00084.
34. Cheung, C.Y., Tay, W.T., Mitchell, P., Wang, J.J., Hsu, W., Lee, M.L., Lau, Q.P., Zhu, A.L., Klein, R., Saw, S.M., and Wong, T.Y. (2011). Quantitative and qualitative retinal microvascular characteristics and blood pressure. J. Hypertens. *29*, 1380–1391. https://doi.org/10.1097/HJH.0b013e328347266c.

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
DeepSLE	This paper	https://github.com/DeepSLE/deepsle
Python 3.7.12	Python	http://www.python.org/
Torch 1.8.1+cu111	PyTorch	https://pytorch.org/
Torchvision 0.9.1+cu111	PyTorch	https://pytorch.org/vision/stable
Opencv-python 4.5.3.56	OpenCV	https://pypi.org/project/opencv-python/4.5.3.56/
Matplotlib 3.5.3	Matplotlib	https://matplotlib.org
Pillow 8.3.1	Pillow	https://pillow.readthedocs.io
Numpy 1.21.6	Numpy	https://numpy.org
Pandas 0.25.0	Pandas	https://pandas.pydata.org/
RETFound-MAE	Github	https://github.com/rmaphoh/RETFound_MAE
Automorph	Github	https://github.com/rmaphoh/AutoMorph
RELPROP	Github	https://github.com/hila-chefer/Transformer-Explainability

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study design

The purpose of this study was to develop and evaluate a DL system for predicting SLE and its complications LR and LN from retinal image solely. We first pre-trained the DL system using 666,383 retinal fundus images from 173,346 participants from the Shanghai Integrated Diabetes Prevention and Care System (Shanghai Integration Model, SIM). Then, we used retinal images from the Peking Union Medical College Hospital to develop and internally validate the DeepSLE system. We also included four independent multi-ethnic datasets (246,078 images in total) for external validation, two from China and two from the UK. We further assessed DeepSLE's performance in different subgroups stratified by gender, age, ethnicity and socioeconomic status. To ensure the interpretability and explainability of DeepSLE, we also conducted saliency analysis and analysis of retinal vascular variables to gain insights into its diagnostic mechanisms. Furthermore, a prospective reader study was conducted to compare the performance of DeepSLE with primary care physicians and specialists.

Study datasets and diagnostic criteria

To develop and internally validate the DeepSLE system for detecting SLE, we collected retinal fundus images and corresponding clinical data from the electronic records of Peking Union Medical College Hospital (PUMCH). We also included four independent external datasets to demonstrate the generalizability of the DeepSLE system. The Shanghai Six People's Hospital (SSPH) dataset was extracted from the Hospital Information System of Shanghai Sixth People's Hospital. All participants in the SSPH dataset underwent kidney biopsy and retinal imaging. The Shanghai Diabetes Prevention Program (SDPP) dataset is a community-based dataset comprising 33,611 participants who underwent comprehensive examinations in Shanghai Health and Medical Center. Data on demographic information, anthropometric indices, biochemical measurements and retinal images were recorded. The United Kingdom Biobank (UKB) was a large-scale biomedical database and research resource, from half a million UK participants. We included 54,319 participants from UKB for external validation. The Moorfields Eye Hospital (MEH) dataset was derived from a large retrospective cohort dataset AlyEye.[27] Among them, 144 participants with SLE and 144 participants without SLE were included. The label of SLE was extracted from ICD codes for these datasets.

To develop and validate the DeepSLE system for detecting lupus retinopathy (LR), two ophthalmologists labeled retinal fundus images of patients with SLE from PUMCH, SSPH, SDPP, and UKB datasets. Inconsistent annotations were checked by a senior ophthalmologist to give final labels. LR was diagnosed if cotton-wool spots, hemorrhages, vasculitis, retinal detachment or optic disc changes as papilledema, optic atrophy were present.[28] Images with DR[29] were excluded in the development and validation datasets for LR detection.

To develop the DeepSLE system for detecting lupus nephritis (LN), 947 patients with SLE were enrolled for development and internal validation. Among them, 112 patients were diagnosed with LN after kidney biopsies and confirmation by pathologists from PUMCH. We included 29 participants with SLE from SSPH dataset for external validation, who were all with LN diagnosed by kidney biopsy.

Cell Reports Medicine
Article

CellPress
OPEN ACCESS

All retinal fundus photographs of SLE patients were acquired after a confirmed diagnosis of SLE. Acquiring images prior to diagnosis would introduce ambiguity regarding the disease status at the time of image capture, potentially resulting in incorrect labeling and underlying bias for the development of DeepSLE. The retinal images were acquired from multiple sources and camera devices, including DRI OCT Triton (Topcon), Topcon TRC-NW300 (Topcon), Topcon TRC-NW400 (Topcon), Topcon 3D OCT1000 Mark II (Topcon), and Topcon 3DOCT-2000SA (Topcon). Additional details regarding the characteristics of participants with SLE included in this study (duration of SLE, the places of retinal imaging, disease stage, reasons for retinal imaging, and Hydroxychloroquine treatment) were provided in Table S1.

Ethics statement

The study was approved by the Institutional Review Board of Peking Union Medical College Hospital of the Chinese Academy of Medical Sciences (Approval No. K22C2366), and the Ethics Committee of Shanghai Sixth People's Hospital (Approval No. 2019-087). All images were retrospectively collected and de-identified prior to the analysis. All procedures followed the tenets of the Declaration of Helsinki.

METHOD DETAILS

Development and implementation of DeepSLE system

Pre-training the DeepSLE system using generative self-supervised learning

Inspired by the success of generative self-supervised learning (SSL) approach in both natural imaging[30] and retinal imaging,[31] we also employed an encoder-decoder architecture to reconstruct retinal images from the highly masked versions to pre-train the DeepSLE system. Same as the figurations of masked autoencoder and RETFound, the large vision Transformer (ViT-large)[32] encoder has 24 Transformer blocks, each of which comprises alternating layers of multi-head self-attention and multilayer perceptron (MLP) blocks. The lightweight decoder is a small vision Transformer (ViT-small) with eight Transformer blocks. An unmasked input image is patched with size of 16×16 and subsequently projected through linear embeddings as the embedding vector of size 1024. These vectors are fed into the encoder to generate high-level features. The decoder then inserts masked dummy tokens into the extracted high-level encoded patches to perform image reconstruction.

The mask ratio is of 0.75. The total training epoch is 800. We selected the checkpoint after the final epoch to form a well-trained SSL model, which could be utilized for further training of the DeepSLE system. We used four NVIDIA A800 (80G) for pretraining, with the batch size as 1,792 (4 GPUs×448 per GPU).

Training of DeepSLE system for detecting SLE, LR and LN

For detecting SLE, LR and LN respectively, the development dataset was divided into training set and tuning set (8: 2) independently, where the training set was used for model optimization and the tuning set for model selection. We only utilized the encoder (ViT-large) of the SSL model to extract high-level features from retinal fundus images. The decoder was replaced with a multilayer perceptron (MLP) as the classifier to output the predicted probability of disease categories. The encoder's initial weights were set using the above pre-trained checkpoint, while the weights of MLP was initialized using a truncated normal distribution with a standard deviation of 2×10^{-5}.

Taking both imbalanced data and imbalanced diagnosis difficulty into consideration, we used a curriculum three-stage training schedule and utilized a balanced-weighted cross-entropy loss with asymmetric label smoothing. In the first stage, a general/standard classifier was trained. In the second stage, we assign larger weight to minority classes for "imbalanced data". In the last stage, we utilized the per-class performance on the tuning set as the indicator for diagnosis difficulty, and updated the weights accordingly. Hence, the training loss was given by:

$$L = -\frac{1}{B}\sum_{i=1}^{B} w_i CE(\tilde{y}_i, y_i), w_i = \begin{cases} 1 & e < E_1 \\ \left(\frac{C/N_c}{\sum_{c\in C} 1/N_c}\right)^{\frac{e-E_1}{E_2-E_1}} & E_1 < e < E_2 \\ \left(\frac{C/f_c^e}{\sum_{c\in C} 1/f_c^e}\right)^{\frac{e-E_2}{E-E_2}} & E_2 < e < E \end{cases}.$$

where B denotes the batch size, C is the total number of classes, w_i is the weight that adjusts the contribution of each sample to the loss. y_i is the ground truth label, $\tilde{y}_i$ is the model prediction. Nc denotes the number of training samples in class C. The diagnosis difficulty indicator is denoted by f_c^e, which means the F1-score of class c on the tuning set after epoch e. The network is trained for total E epochs, where E_1 and E_2 are hyperparameters for stages.

We employed a combination of geometric transformations (e.g., random rotations, flipping, and scaling (resizing cropped patches to 224×224)), color augmentations (e.g., brightness and contrast adjustments), Gaussian noise injection, and image normalization to enhance model generalizability. The training objective is to generate the same categorical output as the label. The batch size is 16. The total training epoch is 50 and the first 10 epochs are for learning rate warming up (from 0 to learning rate 5e-4), followed by a

Please cite this article in press as: Li et al., A deep learning system for detecting systemic lupus erythematosus from retinal images, Cell Reports Medicine (2025), https://doi.org/10.1016/j.xcrm.2025.102203

cosine annealing schedule (from learning rate 5e-4 to 1e-6 in the rest 40 epochs). E_1 and E_2 are set as 10 and 25 respectively. After each epoch training, the model will be evaluated on the tuning set. The model weights with the highest AUROC on the tuning set will be saved as the final model for internal and external validations. We use NVIDIA RTX 3090 (24 GB) for training and testing.

Explainability analyses

RELPROP[33] algorithm specifically designed for explaining Transformer-based networks, was employed to visualize the saliency regions of images. This technique employs layer-wise relevance propagation to compute relevancy scores for each attention head within each layer, which are then integrated across the attention graph by combining relevancy and gradient information. Consequently, this process visualizes the regions of input images contributing to specific classifications.

For retinal vascular variables analysis, we first performed vessel segmentation on CFPs in various regions to get vascular contours. As shown in Figure 3B, we carried out binary vessel segmentation, artery/vein segmentation and optic/disc segmentation on the entire CFP image. Subsequently, binary and artery/vein segmentation were also performed in Zone B (the annulus 0.5–1 optic disc diameter from the disc margin) and in Zone C (the annulus 0.5–2 optic disc diameter from the disc margin), the two standard observation areas.[34] Using the segmented images, we quantified vascular variables associated with the tortuosity and complexity of the entire image, Zone B, and Zone C. Additionally, we measured the caliber of arteries and veins in Zone B and Zone C. We then compared the vascular variables of participants with SLE to those without SLE in the internal test set. The variables included fractal dimension related to vessel complexity, distance tortuosity and squared curvature tortuosity related to vessel tortuosity, and central retinal arteriolar equivalent (CRAE) and central retinal venular equivalent (CRVE)[19] related to vessel caliber.

A prospective reader study to compare the performance of DeepSLE and physicians

We conducted a prospective reader study to compare the performance of DeepSLE with physicians, for detecting SLE, LR and LN. The goal of this reader study was to simulate the scenario of screening SLE and related complications in primary care settings, where rheumatology specialists are not readily available. Five PCPs and five rheumatology specialists were recruited. We curated a dataset for this reader study from PUMCH, comprising 40 non-SLE cases, 15 SLE cases with LR, 15 SLE cases with LN, and 30 SLE cases without LR or LN. The DeepSLE system made the diagnosis of SLE, LR and LN, using retinal fundus images only as inputs. To ensure a reasonable comparison, physicians were asked to make the diagnosis based on age, gender, medical history, findings from physical examinations, and retinal fundus images, which are routinely used in clinical practice. For medical history, we ensured that the physicians were provided with symptoms related to SLE and its complications, including fatigue, joint pain, skin rashes, oral ulcers, photosensitivity, edema, changes in urine output, ocular symptoms, and neuropsychiatric symptoms. This historical context was made available to the primary care physicians to support their clinical decision-making during the comparison. For clinical examinations, we included physical assessment of the patient's vital signs, mucocutaneous changes, joint involvement, edema, vision, and signs of other SLE-related manifestations (Figures 4 and S4). Additionally, we clarify that the primary care physicians and specialists were not expected to perform specialized diagnostic tests (such as advanced imaging or laboratory biomarkers) during this comparison. For outcome analysis, we primarily compared the detection sensitivities of DeepSLE and physicians, since higher sensitivities could minimize the number of false negatives.

QUANTIFICATION AND STATISTICAL ANALYSIS

All task performances were evaluated using AUROCs, sensitivities and specificities. Sensitivity and specificity for SLE detection were reported using the Youden index, while those for LR and LN detection were reported at high-sensitivity operating points (85% for LR and LN) selected from the internal validation set. We also conducted subgroup analyses stratified by gender, age, ethnicity and socioeconomic status. The 95% confidence intervals were estimated with the non-parametric bootstrap method using 1,000 random resampling with replacement. For quantitative analysis of retinal vascular variables, P values were calculated using the Mann-Whitney U tests. For the prospective reader study, McNemar's tests were used to calculate p-values comparing the DeepSLE system's sensitivities and specificities with those of the physicians.