



# Translation of Color Fundus Photography into Fluorescein Angiography Using Deep Learning for Enhanced Diabetic Retinopathy Screening

Danli Shi, MD, PhD,<sup>1,2,\*</sup> Weiji Zhang, MS,<sup>1,\*</sup> Shuang He, MD,<sup>3,\*</sup> Yanxian Chen, MD, PhD,<sup>1</sup> Fan Song, MD,<sup>1</sup> Shunming Liu, MS,<sup>4</sup> Ruobing Wang, MD, PhD,<sup>5</sup> Yingfeng Zheng, PhD,<sup>3</sup> Mingguang He, MD, PhD<sup>1,2,4</sup>

**Purpose:** To develop and validate a deep learning model that can transform color fundus (CF) photography into corresponding venous and late-phase fundus fluorescein angiography (FFA) images.

**Design:** Cross-sectional study.

**Participants:** We included 51 370 CF-venous FFA pairs and 14 644 CF-late FFA pairs from 4438 patients for model development. External testing involved 50 eyes with CF-FFA pairs and 2 public datasets for diabetic retinopathy (DR) classification, with 86 952 CF from EyePACs, and 1744 CF from MESSIDOR2.

**Methods:** We trained a deep-learning model to transform CF into corresponding venous and late-phase FFA images. The translated FFA images' quality was evaluated quantitatively on the internal test set and subjectively on 100 eyes with CF-FFA paired images (50 from external), based on the realism of the global image, anatomical landmarks (macula, optic disc, and vessels), and lesions. Moreover, we validated the clinical utility of the translated FFA for classifying 5-class DR and diabetic macular edema (DME) in the EyePACs and MESSIDOR2 datasets.

**Main Outcome Measures:** Image generation was quantitatively assessed by structural similarity measures (SSIM), and subjectively by 2 clinical experts on a 5-point scale (1 refers real FFA); intragrader agreement was assessed by kappa. The DR classification accuracy was assessed by area under the receiver operating characteristic curve.

**Results:** The SSIM of the translated FFA images were  $> 0.6$ , and the subjective quality scores ranged from 1.37 to 2.60. Both experts reported similar quality scores with substantial agreement (all kappas  $> 0.8$ ). Adding the generated FFA on top of CF improved DR classification in the EyePACs and MESSIDOR2 datasets, with the area under the receiver operating characteristic curve increased from 0.912 to 0.939 on the EyePACs dataset and from 0.952 to 0.972 on the MESSIDOR2 dataset. The DME area under the receiver operating characteristic curve also increased from 0.927 to 0.974 in the MESSIDOR2 dataset.

**Conclusions:** Our CF-to-FFA framework produced realistic FFA images. Moreover, adding the translated FFA images on top of CF improved the accuracy of DR screening. These results suggest that CF-to-FFA translation could be used as a surrogate method when FFA examination is not feasible and as a simple add-on to improve DR screening.

**Financial Disclosure(s):** Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science* 2023;3:100401 © 2023 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Supplemental material available at [www.aaojournal.org](http://www.aaojournal.org).

Diabetic retinopathy (DR) is a retinal microvascular complication of diabetes and a leading cause of vision loss and blindness.<sup>1</sup> The condition is characterized by structural and functional abnormalities in the retinal microvasculature, resulting in capillary occlusion, leakage, retinal ischemia, and neovascularization. An accurate screening is essential for the early detection and monitoring of DR.

Fluorescein angiography (FFA) is a crucial procedure for detecting lesions of the blood-retinal barrier, guiding intervention strategies, and monitoring the treatment response of DR. However, FFA is invasive, requiring the intravenous

injection of a fluorescent dye, and carries a risk of severe side effects including nausea, heart attack, and anaphylactic shock.<sup>2</sup> As a result, FFA is not suitable for routine community screening for DR, and should only be performed in settings with close monitoring and trained technicians. Therefore, developing a noninvasive, safe, and low-cost alternative to FFA is imperative, particularly in regions where DR prevalence is increasing.<sup>3</sup>

Color fundus (CF) photography is a commonly used method to screen and monitor DR. Color fundus images share important anatomical features with FFA, making them

a potential source for generating realistic FFA images. Generative adversarial networks (GANs) offer a promising framework for generating realistic images in this process.<sup>4</sup> Multiple GAN-based models have been implemented in ophthalmic imaging for tasks such as treatment prediction, domain transfer, image denoising, super-resolution, data augmentation, and image segmentation.<sup>5,6</sup> Several exploratory studies have demonstrated the feasibility of using GANs for CF-to-FFA translation.<sup>7–12</sup> Recently, Huang et al reported a GAN model for branch retinal vein occlusion-lesion augmented CF-to-FFA translation with a mean structural similarity measure (SSIM) of 0.536. However, these exploratory studies were limited by small sample sizes (usually < 100 patients) and the inability to generate various pathological changes, such as microaneurysms, neovascularization, and ischemia. To optimize their clinical utility, further research is required. The recent proposal of Gradient Variance loss<sup>13</sup> could potentially enhance image generation by paying closer attention to the high-frequency components, thereby improving the integrity of structural and lesional generation.

The main objective of our study was to create and validate a GAN model that could generate realistic FFA images from CF images by using a substantial clinical dataset. To test our model's reliability, we used external CF-FFA pairs. Additionally, we aimed to evaluate its potential for improving DR screening by using EyePACs and MESSIDOR2 datasets. Our method of translating CF images to FFA images via a GAN model represents a unique and innovative alternative to invasive FFA procedures, and has the potential to enhance retinal disease screening.

## Methods

### Data

We retrospectively collected a total of 13 594 CF images and 396 232 FFA images from 4829 patients who underwent FFA examination between 2016 and 2019 from Zhongshan Ophthalmic Center, Sun Yat-sen University. All patient data were anonymized and deidentified. Color fundus images were obtained using Topcon TRC-50XF and Zeiss FF450 Plus (Carl Zeiss, Inc) cameras, with resolutions ranging from 1110 × 1467 to 2600 × 3200. Fundus fluorescein angiography images were obtained using Zeiss FF450 Plus and Heidelberg Spectralis (Heidelberg Engineering) cameras, with a resolution of 768 × 768.

To validate the generalizability of our model, we retrospectively collected 50 paired CF and FFA images with DR, from Renji Hospital affiliated with Shanghai Jiao Tong University School of Medicine. All images were captured using a Zeiss FF450 Plus camera.

To validate the clinical utility of the translated FFA images, we utilized 2 publicly available datasets commonly used in the development and testing of DR detection: the EyePACs and MESSIDOR2.<sup>14</sup> These datasets were chosen to evaluate the effectiveness of our approach in improving DR screening. The EyePACs dataset consists of 88 702 fundus images captured under various conditions and with various devices at multiple primary care sites throughout California and elsewhere. The MESSIDOR2 dataset includes bilateral images from 874 patients (comprising 1748 images) captured using a Topcon TRC NW6 nonmydriatic fundus camera with a 45-degree field of view.

Diabetic retinopathy was graded according to the ETDRS scale on a scale of 0 to 4 (0 = no DR, 1 = mild nonproliferative DR, 2 = moderate nonproliferative DR, 3 = severe nonproliferative DR, and 4 = proliferative DR, based on ETDRS grading<sup>15</sup>) in both datasets, with MESSIDOR2 also containing labels for diabetic macular edema (DME). Our analysis was restricted to gradable images from both datasets. Table 1 presents the dataset characteristics.

The study adheres to the tenets of the Declaration of Helsinki. The institutional review board approved the study (No.2021KYPJ164-3) and individual consent for this retrospective analysis was waived.

### CF and FFA Matching

We performed a matching of CF and FFA images from the same eye and visit. To achieve pixel-level image matching, we extracted retinal vessels from CF images using the retina-based microvascular health assessment system<sup>16</sup> and vessels from FFA images using a deep learning model.<sup>17</sup> First, FFA images from the same eye were registered, and then they were registered with CF images. Key points were detected from the corresponding vessel map using the AKAZE (Accelerated KAZE) key point detector<sup>18</sup> for feature matching, and random sample consensus<sup>19</sup> was used for generating homography matrices and outlier rejection. To exclude erroneously registered pairs, we added a validity restriction that restricted the value of the rotation scale to 0.8 to 1.3 and the absolute value of the rotation radian to < 2 before the warping transformation. We also filtered out image pairs with poor registration performance, i.e., dice coefficient < 0.5, which was set empirically based on the dataset in our experiments.

### CF to FFA Translation

We used CF images as the input and corresponding real venous and late-phase FFA images as the ground truth to train the model. To reduce variation, we used the cut-off range of 40 seconds to 1.5 minutes for venous-phase and 5 to 6 minutes for late-phase FFA. The images were split at a ratio of 8:1:1 for training, validation, and testing on a patient level. During training, the images were resized to 512 × 512 and fed into pix2pixHD,<sup>20</sup> which is a popular GAN model that can translate images into different domains using a minimax game during training, where the generator G tries to generate a realistic-looking FFA image to fool the discriminator D, while the discriminator D aims to distinguish the generated image from the real one. The generator was constructed with a sequence of stacked transposed convolution layers, facilitating incremental enhancement of image resolution. It was enriched by integrating skip connections, and amalgamating low-level and high-level features to uphold details and contextual information. This strategy enabled the generator to progressively yield intricate images resembling real FFA images. The discriminator was a multiscale convolutional neural network that divided the image into multiple patches and judged the fidelity patch by patch, thus helping to generate high-resolution FFA images closer to the real one. To enhance the generation of high-frequency components, including retinal structure and lesions, we added Gradient Variance loss<sup>13</sup> as a supplementary term within the existing loss functions of the pix2pixHD generator. As a result, the comprehensive loss within the modified Pix2pixHD framework encompasses GAN loss, Feature Matching loss, Content loss, and Gradient Variance loss. To prevent overfitting, we conducted extensive data augmentations during training, including random resized crops (at a scale of 0.3–3.5), random horizontal or vertical flipping, and a random rotation (0–45 degrees). We trained the models with a batch size of 4 and a learning rate of 0.0002, and a total

Table 1. Dataset Characteristics

	N	No DR	Mild	Moderate NPDR	Severe NPDR	PDR
Model Development						
CF-FFA pairs	66 014	28 628 (43.4%)	12 386 (18.8%)	9 486 (14.4%)	4 044 (6.1%)	11 470 (17.4%)
External validation						
CF and FFA ex	50	2 (4.0%)	5 (10.0%)	14 (28.0%)	22 (44.0%)	7 (14.0%)
EyePACs	86 952	64 762 (74.5%)	7 551 (8.7%)	10 955 (12.6%)	1 943 (2.2%)	1 741 (2.0%)
MESSIDOR2	1744	1 017 (58.3%)	270 (15.5%)	347 (19.9%)	75 (4.3%)	35 (2.0%)
		DME				
		yes	no			
		1 593 (91.3%)	151 (8.7%)			

CF = color fundus photography; DME = diabetic macular edema; DR = diabetic retinopathy; FFA = fundus fluorescein angiography; N = number of images; NPDR = nonproliferative diabetic retinopathy; PDR = proliferative diabetic retinopathy.

of 50 epochs were preset for each training. The training employed the Adam optimizer with parameters beta1 set at 0.5 and beta2 at 0.999.

### Assessment of CF-to-FFA Translation Performance

**Objective Evaluation.** To assess the quality of the generated images, we used 4 standard objective measures commonly utilized in image generation for our internal test set: Fréchet inception distance (FID),<sup>21</sup> SSIM,<sup>22</sup> mean absolute error (MAE), and the peak signal-to-noise ratio (PSNR). The FID measures the distance between feature vectors calculated for real and generated images using the Inception V3 model. The coding layer of this model is used to capture computer-vision-specific features of an input image, then these activation functions are calculated for a group of original and generated images and summarized as multivariate Gaussian by calculating the mean and covariance of the images. The distances between the original and generated images are then calculated as the Fréchet distance. A lower FID score indicates higher quality of generated images and a score of 0.0 means the generated images are perfectly consistent with the original ones. Structural similarity measures is an objective image quality metric under the assumption of human visual perception; it compares local patterns of pixel intensities that have been normalized for luminance and contrast. The value of SSIM ranges from 0 to 1, where 1 represents complete similarity and 0 indicates no similarity.<sup>23</sup> The MAE is an intensity-based metric that measures the mean absolute pixel difference for the generated image compared to the original one. A lower MAE demonstrates better image quality, as it reflects a smaller average difference between synthetic and real images.<sup>24</sup> Peak signal-to-noise ratio is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Peak signal-to-noise ratio is defined via the mean squared error, describing the distortion and noise between generated and original images. Higher PSNR suggests higher image quality with less distortion and noise.<sup>25</sup>

**Subjective Evaluation.** Fifty images from each of the internal and external test sets were randomly assigned to 2 experienced ophthalmologists (F.S. and S.H.) for visual quality assessment. The ophthalmologists subjectively rated the quality of the translated images, including the overall authenticity and the validity of anatomical structures and pathological lesions, on a scale of 1 to 5 (1 = excellent, 2 = good, 3 = normal, 4 = poor, and 5 = very poor), with score 1 referring to the image quality of the real FFA image. Fig S1 demonstrates the grading criteria. Interrater agreement was evaluated using Cohen’s linearly weighted kappa

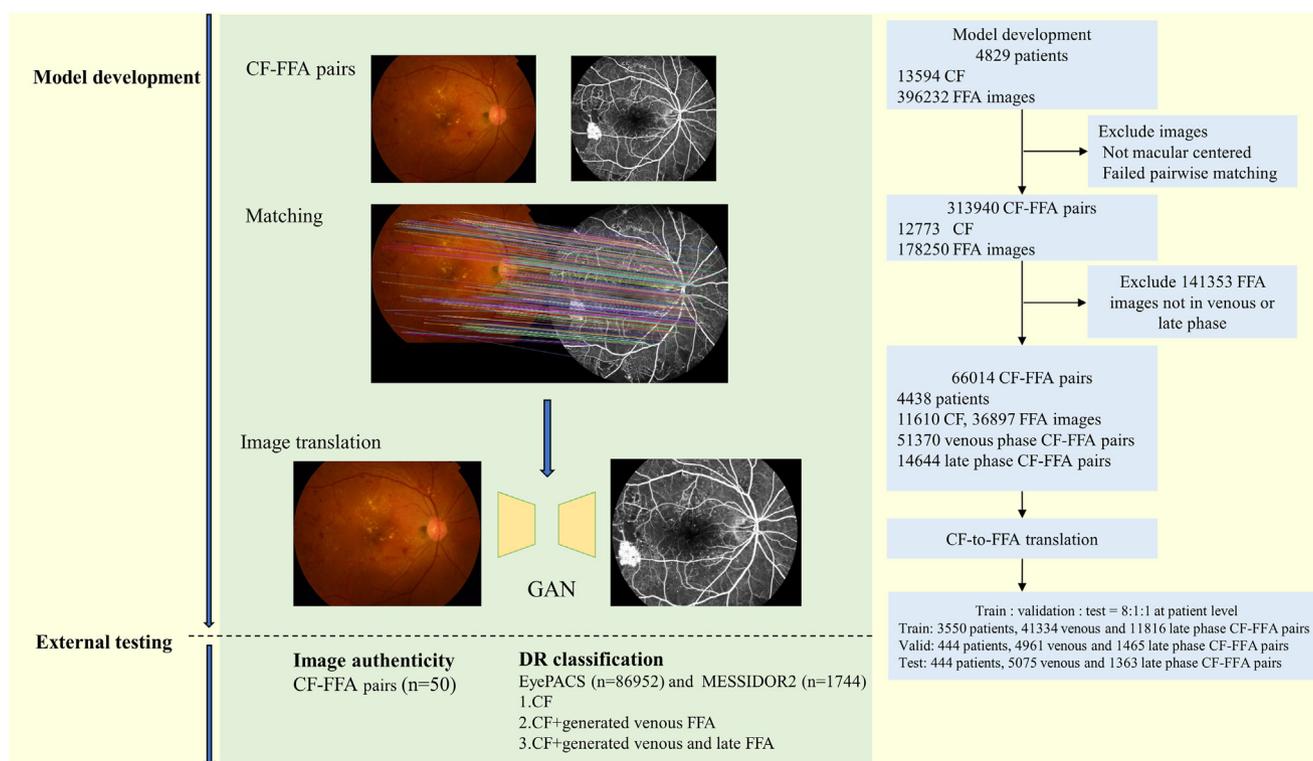
score,<sup>26</sup> which ranges from  $-1$  to  $1$ . It can be interpreted as follows: 0.40 to 0.60 represents a moderate agreement, 0.60 to 0.80 represents a substantial agreement, and 0.80 to 1.00 represents an almost perfect agreement.

**DR Classification.** A comparative study was conducted to assess how adopting generated FFA images may improve the accuracy of DR classification. Experiments were done using the Swin-transformer<sup>27</sup> model under the same hyperparameter setting to classify DR based on CF, CF plus the generated venous FFA, and CF plus the generated venous plus late-phase FFA images. The experiment was conducted on 2 datasets, with DR classes of 0 to 4 on both EyePACs and MESSIDOR2 datasets, and with or without DME on the MESSIDOR2 dataset. In experiments where the generated FFA images were added, the CF and generated FFA images were fed into the Swin-transformer and encoded into embeddings of length 512, and the embeddings were concatenated and passed through fully-connected layers and a softmax layer to obtain the classification output. We used the Swin Transformer model because its capacity to capture both local and global relationships makes it ideal for DR classification. Its patch-based attention mechanism accommodates the processing of high-resolution retinal images and enables effective multiscale analysis, which is suitable for identifying microaneurysms and other lesions. Furthermore, the Swin Transformer excels at capturing long-range dependencies crucial for detecting lesions spanning extensive areas. The training, validation, and test sets were divided in a ratio of 6:2:2. Images were resized to 512\*512, and augmented with random horizontal flips and random rotations of  $-30$  to  $30$  degrees during training. We used Adam optimizer with a learning rate of  $1e-5$  and a batch size of 4. Each experiment was trained for 30 epochs, and the models with the highest area under the receiver operating characteristic curve value on the validation set were used for testing. On the test set, we calculated the F1-score, sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve for assessing DR classification performance. To obtain fine-grained class-level performance, we also provided confusion matrixes.

We used Pytorch to develop the deep learning algorithm and trained the models on an NVIDIA GeForce RTX 3090 card.

## Results

A total of 1984 CF and 359 335 FFA images were excluded from the analysis, including those that were not macular centered, failed CF-FFA pairwise matching, or did not belong to the venous or late phase. The final dataset included 51 370 venous CF-FFA pairs and 14 644 late-phase



**Figure 1.** Flow chart of the study. CF = color fundus photography; FFA = fundus fluorescein angiography; GAN = generative adversarial networks.

pairs from 4438 participants for model development. The medium (interquartile range) age of the participants was 51.47 (25.39) years, and 2430 (54.8%) were male. Of these participants, the majority were diagnosed with eye diseases, including DR, retinal vein occlusion, and age-related macular degeneration. Among them, 2669 (60.1%) had DR, with 1865 at stage 1, 637 at stage 2, 366 at stage 3, and 801 at stage 4. The study flow chart is shown in [Figure 1](#). Characteristics of DR in the dataset are presented in [Table 1](#), characteristics of other eye diseases are presented in [Table S1](#), the diagnoses were automatically extracted from FFA reports.

## Objective Evaluation

Pixel-wise comparison between the real and CF-translated FFA was performed on the internal test set; the MAE, PSNR, SSIM, and FID were 111.46, 21.07, 0.61, and 46.28 respectively for venous-phase FFA and 123.07, 22.11, 0.65, and 32.72 respectively for late-phase FFA. The higher the SSIM, PSNR, and the lower the FID, the better the generated images. For the FFA generation task, the reported SSIM in the literature were generally around 0.4 to 0.6.<sup>7-12</sup>

## Subjective Evaluation

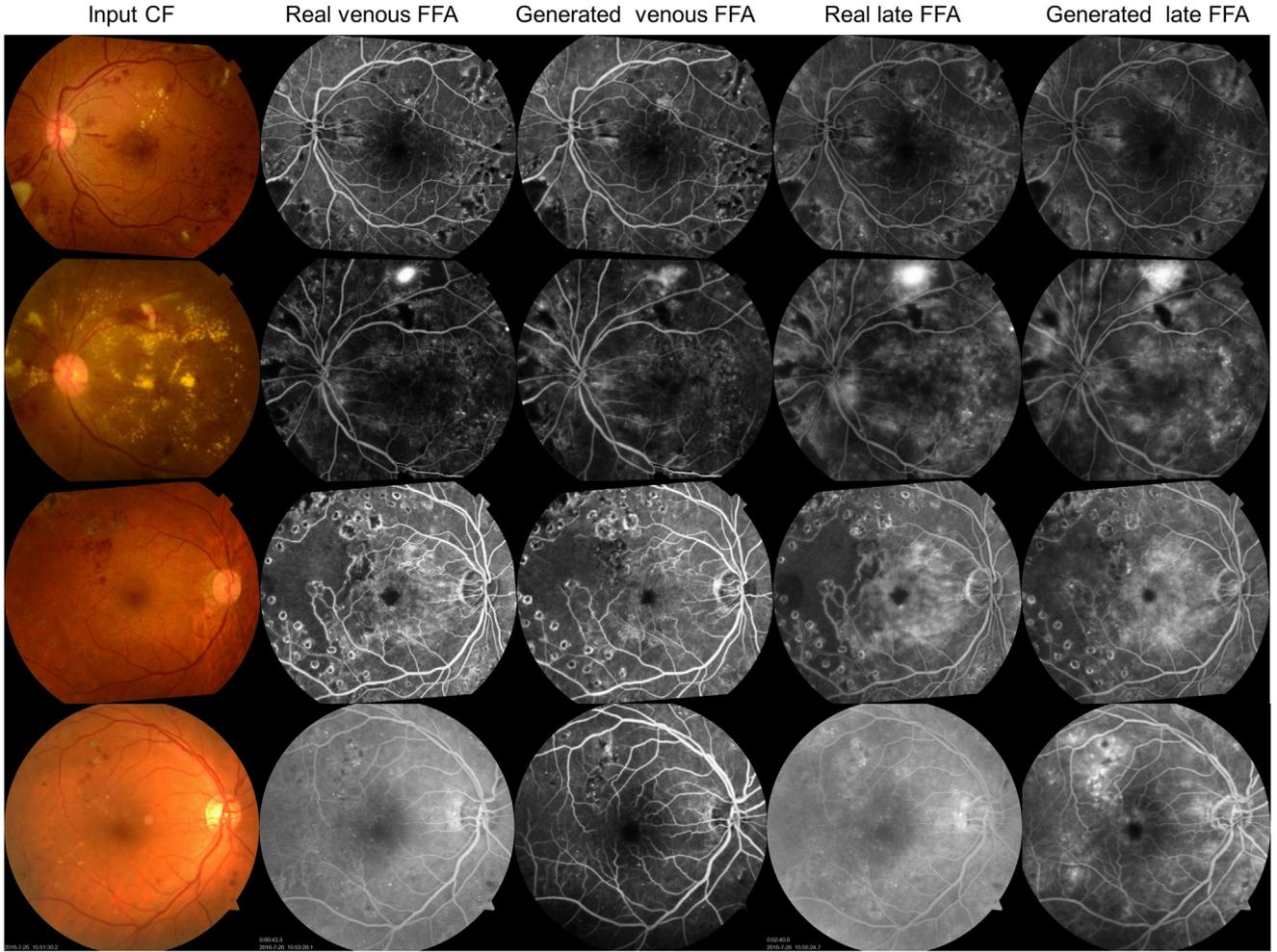
Example-generated images in the internal and external test sets are shown in [Figure 2](#). The model efficiently mapped the features from CF into FFA, resulting in globally, anatomically, and lesion-consistent FFA images. Notably, the black and white background noise sometimes present in the real FFA was learned as unrelated and was effectively

excluded in the translated images. The synthesized output images are visually very close to real ones.

Image quality assessment, which considers retinal structure and lesions, was based on a 5-point scale ([Table 2](#)). The mean (standard deviation) of the scores for venous-phase FFA was 2.12 (0.77) and 2.26 (0.60) for the internal and external test set, respectively, assessed by the first grader, and 2.10 (0.75), 2.28 (0.61) by the second grader. The mean (standard deviation) of the scores for late-phase FFA were 1.37 (0.53) and 2.50 (0.65) in the internal and external test sets, respectively, assessed by the first grader, and 1.47 (0.54), 2.60 (0.64) by the second grader. Cohen's kappa values indicate an excellent agreement between the 2 graders for assessing image quality, with a kappa value of 0.84, 0.81 for venous-phase FFA, and 0.80, 0.83 for late-phase FFA in the internal and external test sets, respectively. This reflects the high quality of synthesized images for anatomical features (vessel, optic disc, and macula) and lesions (nonperfusion, neovascularization, and macular edema).

In the internal and external test sets, 1.8% and 4% of the generated FFA images were of poor quality ( $\geq 4$  points) due to the following reasons: lesions that were not prominent on CF (such as nonperfusion and microaneurysms) could be missed, and a blurry CF could result in a blurry generated FFA. Additionally, a blurry CF may also produce a false positive generation of DR lesions, as demonstrated in [Figs S1](#) and [S2](#).

While this study primarily focused on DR, [Fig S3](#) provides generation examples for other diseases, such as polypoidal choroidal vasculopathy, central serous chorioretinopathy, normal CF, wet age-related macular degeneration, retinitis



**Figure 2.** Examples of real and translated fundus fluorescein angiography (FFA) images. First row, severe nonproliferative diabetic retinopathy (DR), second row, proliferative DR, third row, DR post-laser treatment with severe nonperfusion and diabetic macula edema. First to third rows: internal test set, color fundus photographs (CF) were registered with FFA, rotation occurs during this process. Fourth row: external test set.

pigmentosa, and retinal vein occlusion, to further demonstrate the model’s capabilities.

### DR Classification

Table 3 and Figures 3–5 present the quantitative results of DR classification on the EyePACs (Fig 3) and MESSIDOR2 datasets (Figs 4 and 5). The addition of generated FFA on top of CF significantly improved the overall DR classification accuracy, as evidenced by the improved area under the

receiver operating characteristic curve for 5-class DR and DME classification in Table 3. Incorporating generated FFA images led to decreased error rates for each specific DR category, as demonstrated in Figures 3–5.

### Discussion

We developed a CF-to-FFA translation framework and further demonstrated for the first time that this framework

Table 2. Subjective Evaluation of Real and Translated FFA Image Quality

	Internal Test Set (N = 50)			External Test Set (N = 50)		
	Rater 1 Mean (SD)	Rater 2 Mean (SD)	Kappa	Rater 1 Mean (SD)	Rater 2 Mean (SD)	Kappa
Venous-phase FFA	2.12 (0.77)	2.10 (0.75)	0.84, $P < 0.001$	2.26 (0.60)	2.28 (0.61)	0.81, $P < 0.001$
Late-phase FFA	1.37 (0.53)	1.47 (0.54)	0.80, $P < 0.001$	2.50 (0.65)	2.60 (0.64)	0.83, $P < 0.001$

FFA = fundus fluorescein angiography; SD = standard deviation.

Table 3. DR Classification Based on CF and CF + Translated FFA Images on the EyePACs (n = 86 952) and MESSIDOR2 Dataset (n = 1749), train:validation:test = 6:2:2

Dataset	F1-score	Sensitivity	Specificity	Accuracy	AUC
EyePACs					
Five-class DR					
CF	0.715	0.694	0.740	0.694	0.912
CF + venous FFA	0.760	0.736	0.795	0.736	0.934
CF + venous + late FFA	<b>0.775</b>	<b>0.757</b>	<b>0.795</b>	<b>0.757</b>	<b>0.939</b>
MESSIDOR2					
Five-class DR					
CF	0.793	0.791	0.896	0.791	0.952
CF + venous FFA	0.845	0.846	<b>0.918</b>	0.846	0.966
CF + venous + late FFA	<b>0.866</b>	<b>0.870</b>	0.916	<b>0.870</b>	<b>0.972</b>
DME					
CF	0.698	0.763	0.941	0.922	0.927
CF + venous FFA	0.706	0.763	<b>0.945</b>	0.925	0.967
CF + venous + late FFA	<b>0.713</b>	<b>0.789</b>	0.941	<b>0.925</b>	<b>0.974</b>

AUC = area under the receiver operating characteristic curve; CF = color fundus photography; DME = diabetic macular edema; DR = diabetic retinopathy; FFA = fundus fluorescein angiography. Highest metric values for each dataset are marked in bold.

can reliably generate realistic FFA images that are sufficiently close to the true FFA images in terms of both anatomical structures and pathological lesions and also proved that the generated FFA images can further improve the accuracy of DR classification in the EyePACs and MESSIDOR2 dataset.

Given that noninvasive imaging has become an emerging trend in clinical practice,<sup>28</sup> CF to FFA image translation holds several advantages. First, FFA has been extensively used in clinical ophthalmic practice, resulting in a large

number of images that can be used as training labels at a low cost, providing a good opportunity to satisfy data-hungry deep learning algorithms in this field.<sup>29</sup> Second, by mimicking the FFA images at 2 critical periods during the dynamic circulation of the fluorescein injection, we can leverage the advantages of FFA to obtain dynamic information for the diagnosis and monitoring of DR. Third, although OCT angiography is becoming an increasingly popular option for detecting retinal microvascular abnormalities in a noninvasive way, it may

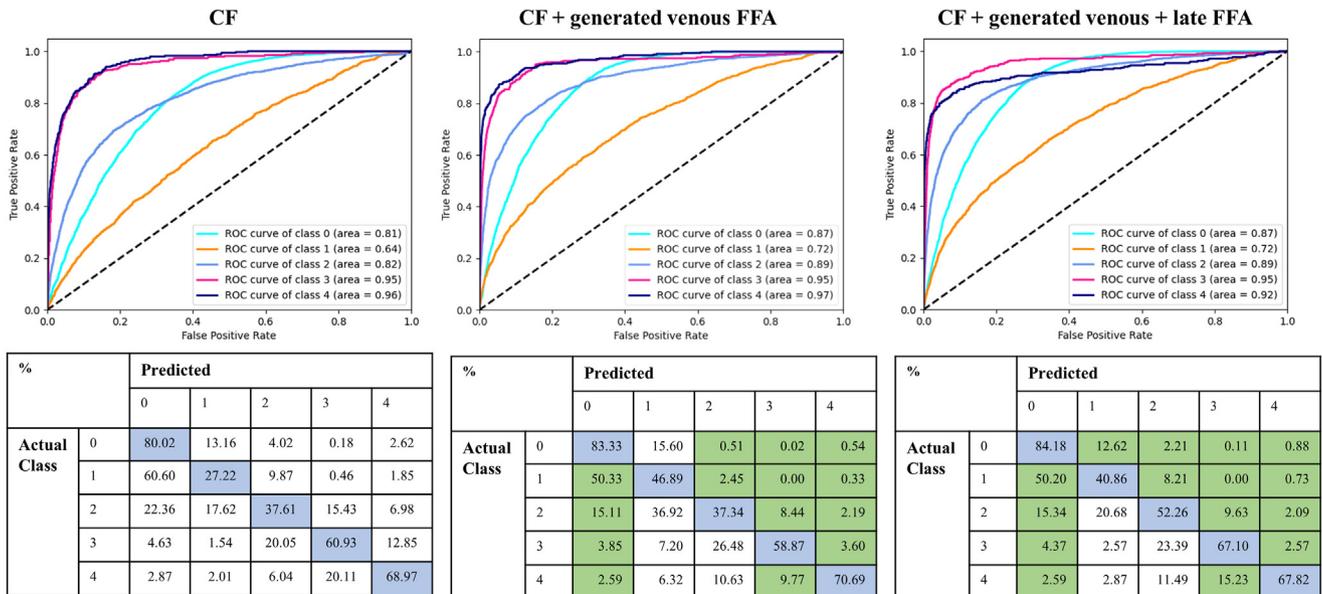
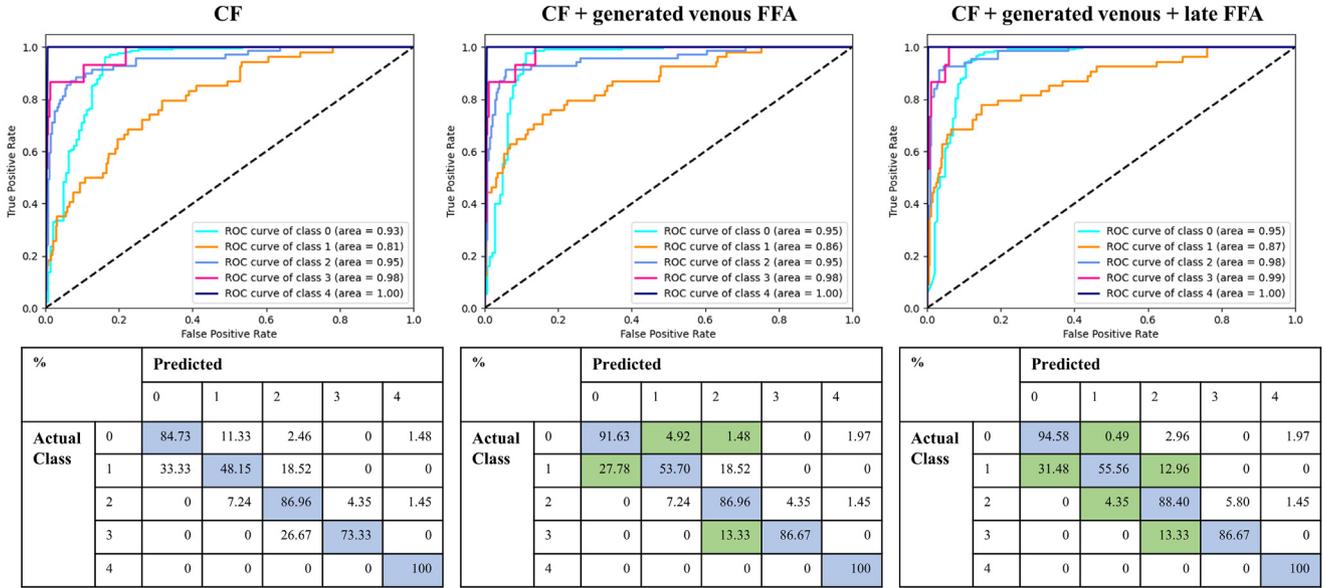


Figure 3. Comparison of diabetic retinopathy (DR) classification results on the EyePACs dataset (n = 86 952) with and without the addition of translated fundus fluorescein angiography (FFA) on top of color fundus photography (CF). The classification is based on 5 categories: 0 = no DR, 1 = mild DR, 2 = moderate DR, 3 = severe nonproliferative DR, and 4 = proliferative DR. The first row shows the receiver operating characteristic (ROC) curves for each model, and the second row displays the corresponding confusion matrices. The blue color in the confusion matrix represents correct predictions, while the green color highlights a reduced error rate with the addition of the generated FFA.



**Figure 4.** Comparison of diabetic retinopathy (DR) classification results on the MESSIDOR2 dataset ( $n = 1749$ ) with and without the addition of generated fundus fluorescein angiography (FFA) on top of color fundus photography (CF). The classification is based on 5 categories: 0 = no DR, 1 = mild DR, 2 = moderate DR, 3 = severe nonproliferative DR, and 4 = proliferative DR. The first row shows the receiver operating characteristic (ROC) curves for each model, and the second row displays the corresponding confusion matrices. The blue color in the confusion matrix represents correct predictions, while the green color highlights a reduced error rate with the addition of the generated FFA.

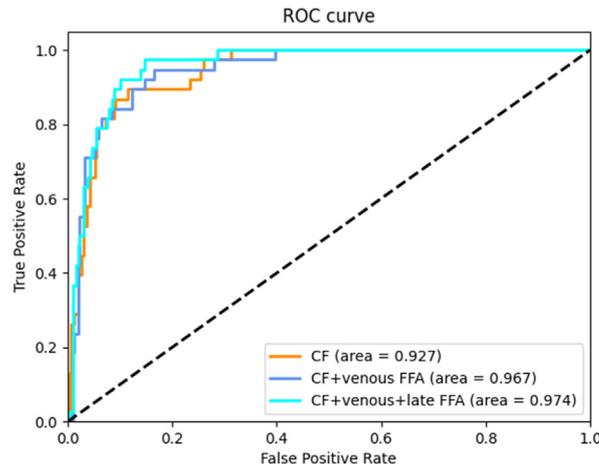
not be applied to large-scale screening because of its expensive examination fees, long scanning time, limited scanning scope, and unstable image quality.<sup>28,30,31</sup> Generating CF-based FFA images offers a wider field of view, takes less than a second, and is more cost-effective than OCT angiography. In addition, FFA remains the most discriminative imaging that can identify more lesions than CF or OCT angiography.<sup>17,28,32</sup> Therefore, CF-to-FFA translation may be a preferred approach under certain scenarios and warrants future investigation.

To ensure the algorithm’s generalizability, we tested it on external datasets. When subjectively evaluated by 2 ophthalmologists, the translated FFA in both the internal and external test sets achieved good image authenticity using real FFA as references, indicating that the algorithm learned critical feature mapping instead of just memorizing the training set samples. While our dataset is primarily composed of patients with DR, the model’s potential for generalization to other common retinal conditions such as age-related macular degeneration, retinitis pigmentosa, retinal vein occlusion, and central serous chorioretinopathy also seems plausible, as indicated in Fig S3. This suggests that the synthesized FFA images may serve as an important surrogate for routine ophthalmic examinations when traditional FFA is not available.

Under the same experimental setting for DR screening, we observed significant improvement in DR classification accuracy on the EyePACS and MESSIDOR2 datasets after adding generated FFA images. The addition of venous-phase FFA resulted in a prominent improvement in 5-scale DR classification, with a minor increase observed after further adding late-phase FFA. This suggests that the

generated venous-phase FFA contains substantial information for discriminating the 0 to 4 DR. For detecting DME, the addition of late-phase FFA further increased sensitivity, which is intuitive as late-phase FFA is characterized by leakage and edema. Compared to previous studies,<sup>33,34</sup> the algorithm showed similar accuracy in DR classification, which was further improved by incorporating generated FFA images. As most DR screening programs are based on conventional CF,<sup>35–37</sup> the CF-to-FFA framework as a new technology serves as a simple add-on to improve DR screening accuracy with no additional cost.

This study has limitations. First, we utilized a static version of FFA images with a field of view limited to  $55^\circ$ , which may have missed some dynamic and peripheral manifestations of retinal diseases. Future investigations should explore the generation of dynamic CF-based FFA images with extended views to improve the comprehensiveness of this method. Second, blurry CF images could result in blurry generated FFA, or even worse, result in false positive DR generation due to the wrong association of blurry images with proliferative DR, which was incorrectly learned from biased data. Therefore, patients with suboptimal CF quality due to refractive media opacity should be excluded and are warranted to take real-FFA examinations. Third, indistinctive lesions on the CF such as microaneurysms and nonperfusion may be missed; therefore, caution needs to be taken using this technology. Meanwhile, the translated FFA images should always be used in conjunction with CF. Finally, though the subjective quality assessment of generated FFA images in both internal and external datasets ranged from excellent to normal, the data size of our internal and external test datasets is limited. Further explorations should include more data to verify and improve the algorithm’s performance



% Actual Class		Predicted	
		No	Yes
No	94.14	5.86	
Yes	23.68	76.32	

% Actual Class		Predicted	
		No	Yes
No	94.46	5.54	
Yes	23.68	76.32	

% Actual Class		Predicted	
		No	Yes
No	94.14	5.86	
Yes	21.05	78.95	

**Figure 5.** Comparison of diabetic macular edema classification results on the MESSIDOR2 dataset (n = 1749) with and without the addition of translated fundus fluorescein angiography (FFA) on top of color fundus photography (CF). The classification is based on 2 categories. The first row shows the receiver operating characteristic (ROC) curves for 3 models, and the second row displays the corresponding confusion matrices. The blue color in the confusion matrix represents correct predictions, while the green color highlights a reduced error rate with the addition of the generated FFA.

and generalizability on a full spectrum of retinal diseases, including rare conditions.

### Conclusion

In this study, we developed and validated a framework for translating CF images into realistic venous and late-phase

FFA images. The model showed high authenticity in generating anatomical structures and pathological lesions, and was proven to improve DR classification accuracy. These results imply that this technology is a promising add-on for large-scale DR screening and assisting in clinical decision-making for the diagnosis and management of DR. Future prospective validations are needed to prove its utility in clinical practice.

### Footnotes and Disclosures

Originally received: June 9, 2023.  
 Final revision: August 23, 2023.  
 Accepted: September 8, 2023.  
 Available online: September 15, 2023. Manuscript no. XOPS-D-23-00128.  
<sup>1</sup> School of Optometry, The Hong Kong Polytechnic University, Kowloon, Hong Kong.  
<sup>2</sup> Research Centre for SHARP Vision (RCSV), The Hong Kong Polytechnic University, Kowloon, Hong Kong.  
<sup>3</sup> State Key Laboratory of Ophthalmology, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Zhongshan Ophthalmic Center, Guangdong Provincial Clinical Research Center for Ocular Diseases, Sun Yat-sen University, Guangzhou, China.  
<sup>4</sup> Department of Ophthalmology, Guangdong Academy of Medical Sciences, Guangdong Provincial People’s Hospital, Guangzhou, China.  
<sup>5</sup> Department of Ophthalmology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China.

\*D.S., W.Z. and S.H. contributed equally to this work. Y.Z. and M.H. are co-senior authors.  
 Disclosures:  
 All authors have completed and submitted the ICMJE disclosures form. The authors made the following disclosures:  
 M.H.: All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.), Global STEM Professorship Scheme (P0046113) by the Government of HKSAR, and the National Natural Science Foundation of China (grant no.: 81420108008).  
 M.H. and D.S. are inventors of the technology mentioned in the study patented as “A method to translate fundus photography to realistic angiography” (CN115272255A).  
 HUMAN SUBJECTS: No human subjects were included in this study. The study adheres to the tenets of the Declaration of Helsinki. The institutional

review board approved the study (No.2021KYPJ164-3) and individual consent for this retrospective analysis was waived.

Author Contributions:

Conception and design: Shi, M. He, Liu

Analysis and interpretation: Shi, Zhang, S. He, Song, Liu

Data collection: Shi, Zhang, S. He, Zheng, M. He, Wang

Obtained funding: The study was performed as part of regular employment duties at the Hong Kong Polytechnic University, the study was supported by the Global STEM Professorship Scheme, P0046113

Overall responsibility: Shi, Zhang, S. He, M. He, Chen

Abbreviations and Acronyms:

**CF** = color fundus; **DME** = diabetic macular edema; **DR** = diabetic retinopathy; **FFA** = fundus fluorescein angiography; **FID** = Fréchet inception distance; **GAN** = generative adversarial network; **MAE** = mean absolute error; **SSIM** = structural similarity measures; **PSNR** = peak signal-to-noise ratio.

Keywords:

Image-to-image translation, FFA generation, Generative Adversarial Networks, Diabetic retinopathy.

Correspondence:

Mingguang He, MD, PhD, Experimental Ophthalmology, School of Optometry, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China. E-mail: mingguang.he@polyu.edu.hk.

## References

- Sabanayagam C, Banu R, Chee ML, et al. Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol.* 2019;7:140–149.
- Kornblau IS, El-Annan JF. Adverse reactions to fluorescein angiography: a comprehensive review of the literature. *Surv Ophthalmol.* 2019;64:679–693.
- Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology.* 2021;128:1580–1591.
- Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: an overview. *IEEE Signal Process Mag.* 2018;35:53–65.
- You A, Kim JK, Ryu IH, Yoo TK. Application of generative adversarial networks (GAN) for ophthalmology image domains: a survey. *Eye Vis.* 2022;9:6.
- Chen JS, Coyner AS, Chan RVP, et al. Deepfakes in ophthalmology. *Ophthalmol Sci.* 2021;1:100079.
- Tavakkoli A, Kamran SA, Hossain KF, Zuckerbrod SL. A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs. *Sci Rep.* 2020;10:21580.
- Kamran SA, Fariha Hossain K, Tavakkoli A, et al. *Fundus2Angio: a conditional GAN architecture for generating fluorescein angiography images from retinal fundus photography.* International Symposium on Visual Computing. San Diego, CA, USA: Springer International Publishing; 2020:125–138.
- Li W, Kong W, Chen Y, et al. *Generating fundus fluorescence angiography images from structure fundus images using generative adversarial networks.* In: International Conference on Medical Imaging with Deep Learning. 121. Montreal, QC, Canada: PMLR; 2020:424–439.
- Chen Y, He Y, Li W, et al. Series-parallel generative adversarial network architecture for translating from fundus structure image to fluorescence angiography. *Appl Sci.* 2022;12:10673.
- Kamran SA, Hossain KF, Tavakkoli A, Zuckerbrod SL. *Attention2AngioGAN: synthesizing fluorescein angiography from retinal fundus images using generative adversarial networks.* 2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE; 2021:9122–9129.
- Huang K, Li M, Yu J, et al. Lesion-aware generative adversarial networks for color fundus image to fundus fluorescein angiography translation. *Comput Methods Programs Biomed.* 2023;229:107306.
- Abrahamyan L, Truong AM, Philips W, Deligiannis N. Gradient variance loss for structure-enhanced image super-resolution. In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Singapore: IEEE; 2022:3219–3223.
- Decencière E, Zhang X, Cazuguel G, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Anal Stereol.* 2014;33:231–234.
- Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology.* 1991;98(5 Suppl):786–806.
- Shi D, Lin Z, Wang W, et al. A deep learning system for fully automated retinal vessel measurement in high throughput image analysis. *Front Cardiovasc Med.* 2022;9:823436.
- Mohamed A, Omar A, Siamak S, et al. Evaluation of diabetic retinopathy severity on ultrawide field colour images compared with ultrawide fluorescein angiograms. *Br J Ophthalmol.* 2023;107:534.
- P. Alcantarilla, J. Nuevo and A. Bartoli, Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans, Patt. Anal. Mach. Intell.* 34(7):1281–1298.
- Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM.* 1981;24:381–395.
- Wang T-C, Liu M-Y, Zhu J-Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): IEEE.* 2018:8798–8807.
- Heusel M, Ramsauer H, Unterthiner T, et al. *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.* Advances in neural information processing systems; 2017:30.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004;13:600–612.
- Wang Z, Pan W, Cuppens-Boulahia N, et al. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans Image Process.* 2013;13:600–12.
- Chourak H, Barateau A, Tahri S, et al. Quality assurance for MRI-only radiation therapy: a voxel-wise population-based methodology for image and dose assessment of synthetic CT generation methods. *Front Oncol.* 2022;12:968689.
- Tanabe Y, Ishida T. Quantification of the accuracy limits of image registration using peak signal-to-noise ratio. *Radiol Phys Technol.* 2017;10:91–94.

26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
27. Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *2021 IEEE/CVF, International Conference on Computer Vision (ICCV)*. 2021: 9992–10002.
28. Kashani AH, Chen CL, Gahm JK, et al. Optical coherence tomography angiography: a comprehensive review of current methods and clinical applications. *Prog Retin Eye Res*. 2017;60: 66–100.
29. Shi D, He S, Yang J, et al. One-shot retinal artery and vein segmentation via cross-modality pretraining. *Ophthalmol Sci*. 2023;100363.
30. Schreur V, Larsen MB, Sobrin L, et al. Imaging diabetic retinal disease: clinical imaging requirements. *Acta Ophthalmol*. 2022;100:752–762.
31. Koutsiaris AG, Batis V, Liakopoulou G, et al. Optical coherence tomography angiography (OCTA) of the eye: a review on basic principles, advantages, disadvantages and device specifications. *Clin Hemorheol Microcirc*. 2023;83: 247–271.
32. La Mantia A, Kurt RA, Mejer S, et al. Comparing fundus fluorescein angiography and swept-source optical coherence tomography angiography in the evaluation of diabetic macular perfusion. *Retina*. 2019;39:926–937.
33. Zhou Y, Wang B, Huang L, et al. A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. *IEEE Trans Med Imag*. 2021;40:818–828.
34. Saxena G, Verma DK, Paraye A, et al. Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets. *Intell Based Med*. 2020;3-4:100022.
35. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care*. 2018;41:2509–2516.
36. Ruamviboonsuk P, Tiwari R, Sayres R, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. *Lancet Digit Health*. 2022;4:e235–e244.
37. Dai L, Wu L, Li H, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat Commun*. 2021;12:3242.