Supplementary Information for "Interpretable Fundus Image Classification via Ring-Based Retinal Vasculature Features"

Xiaoyan Li[1], Shixin Xu[3], Arvind Gupta[1], and Huaxiong Huang[1,2,3,4] *

[1]Department of Computer Science, University of Toronto, Toronto, ON, Canada
[2]Department of Mathematics and Statistics, York University, Toronto, ON, Canada
[3]Zu Chongzhi Center, Duke Kunshan University, Kunshan, Jiangsu, China
[4]Mathematics, Analytics, and Data Science Lab, Fields Institute for Research in Mathematical Sciences, Toronto, ON, Canada

S1 Sensitivity to Annular Ring Width

The proposed ring-structured representation partitions the fundus into concentric annuli centered on the optic disc. In the main HRF and SYSU experiments, we used a ring width of 0.5× optic disc diameter (DD). This interval was chosen as a disc-size-normalized spatial sampling unit that balances regional specificity and feature reliability. Narrower rings provide finer spatial localization but may contain too few vessel pixels for stable feature estimation, whereas wider rings improve measurement stability but may average out localized vascular changes.

To evaluate the sensitivity of the method to this design choice, we repeated the HRF classification experiment using ground-truth vessel masks with alternative annular widths. Specifically, we tested widths of 0.25×DD, 1.0×DD, and 1.5×DD. The results are summarized in Supplementary Table S1. Across all tested settings, classification performance remained high. The 0.25×DD setting achieved an accuracy of 0.911, while the 1.0×DD and 1.5×DD settings both achieved an accuracy of 0.933. In the original HRF ground-truth-mask experiment using 0.5×DD rings, the proposed method achieved an accuracy of 1.000. These results suggest that, at least on HRF with ground-truth vessel masks, the performance is not driven by a fragile choice of annular interval and remains high across the tested ring-width settings.

We retained 0.5×DD for HRF and SYSU because it balances regional specificity and feature reliability. Finer 0.25×DD rings may suffer from sparse vessel support, whereas coarser 1.0×DD and 1.5×DD rings may average out localized vascular variation.

For FIVES, we used 1.0×DD rings to preserve dataset coverage. Although 0.5×DD with eight rings improved accuracy to 0.792, it reduced the valid sample size from 421 to 231, likely due to insufficient vessel measurements in narrow annuli. Therefore, the FIVES setting was chosen to balance spatial resolution, feature stability, and retained sample size.

Supplementary Table S1 Sensitivity analysis of annular ring width on the HRF dataset using ground-truth vessel masks. Performance remained high across different spatial granularities, suggesting that the proposed ring-structured representation is not highly sensitive to the exact ring-width setting.

Ring width	Rings used	Valid images	Accuracy	95% CI
0.25×DD	18	45	0.911	[0.793, 0.965]
0.5×DD	9	45	1.000	[0.921, 1.000]
1.0×DD	5	45	0.933	[0.821, 0.977]
1.5×DD	3	45	0.933	[0.821, 0.977]

S2 Derivation of the RGB Optical-Density Approximation

Under the modified Beer–Lambert model commonly used in retinal vessel oximetry [1, 2], we model the vessel signal as the corresponding vessel-free reference multiplied by an exponential attenuation term due to blood:

$$I_{\mathrm{in}}(\lambda, p) \approx I_{\mathrm{out}}(\lambda, p)\, \exp\big(-\varepsilon(\lambda)\, h(p)\, l(p)\big), \qquad \lambda \in \Lambda_c, \tag{S1}$$

where $\varepsilon(\lambda)$ is the effective haemoglobin extinction coefficient at wavelength λ, $h(p)$ is an effective haemoglobin concentration term, and $l(p)$ is the effective optical path length.

As described in the oxygenation-related feature extraction section of the main manuscript, the band-integrated vessel-free and vessel intensities for channel c are denoted by $\bar{I}_{\mathrm{out},c}(p)$ and $\bar{I}_{\mathrm{in},c}(p)$, respectively. Using these band-integrated quantities together with (S1), we obtain the practical OD computed from the channel measurements:

$$\begin{aligned} OD_c(p) &\approx -\ln\left(\frac{\bar{I}_{\mathrm{in},c}(p)}{\bar{I}_{\mathrm{out},c}(p)}\right) \\ &= -\ln\left(\frac{\sum_{\lambda\in\Lambda_c} \Phi_c(\lambda)\, I_{\mathrm{out}}(\lambda, p)\, \exp\big(-\varepsilon(\lambda)\, h(p)\, l(p)\big)}{\sum_{\lambda\in\Lambda_c} \Phi_c(\lambda)\, I_{\mathrm{out}}(\lambda, p)}\right). \end{aligned} \tag{S2}$$

The effective weighting in (S2) is proportional to $\Phi_c(\lambda)\, I_{\mathrm{out}}(\lambda, p)$ and therefore depends, in principle, on the vessel location p through the local background spectrum.

Defining the corresponding local normalized weighting as

$$W_c(\lambda, p) \triangleq \frac{\Phi_c(\lambda)\, I_{\text{out}}(\lambda, p)}{\sum_{\lambda' \in \Lambda_c} \Phi_c(\lambda')\, I_{\text{out}}(\lambda', p)},$$

(S2) can be rewritten as

$$OD_c(p) \approx -\ln\left(\sum_{\lambda \in \Lambda_c} W_c(\lambda, p)\, \exp\big(-\varepsilon(\lambda)\, h(p)\, l(p) \big) \right). \tag{S3}$$

Publicly available RGB fundus datasets rarely provide measured illumination spectra, sensor spectral sensitivities, or CFA transmission curves; consequently, the true wavelength-dependent channel weighting is not directly observable. In our SO_2-proxy formulation, we therefore adopt a dataset-level approximation in which the local normalized weighting $W_c(\lambda, p)$ is replaced by a channel-specific effective weighting $W_c(\lambda)$, assuming that after local referencing the normalized spectral shape of $I_{\text{out}}(\lambda, p)$ varies only weakly across vessel locations within each channel band. Under this approximation,

$$OD_c(p) \approx -\ln\left(\sum_{\lambda \in \Lambda_c} W_c(\lambda)\, \exp\big(-\varepsilon(\lambda)\, h(p)\, l(p) \big) \right). \tag{S4}$$

In this formulation, illumination and background reflectance are absorbed into the spectral radiance terms $I_{\text{in}}(\lambda, p)$ and $I_{\text{out}}(\lambda, p)$, while $W_c(\lambda)$ represents a normalized effective channel weighting that summarizes the wavelength selectivity of the imaging system, including sensor quantum efficiency, CFA transmission, and optical transmittance. Although this assumption cannot be verified directly without camera calibration data, its plausibility is indirectly supported by sensitivity analyses showing that moderate perturbations of the assumed channel centers and bandwidths preserve the overall SO_2-proxy descriptors and class-separation trends (Supplementary Section S3.1).

We further adopt a first-order approximation to obtain an interpretable effective band-averaged form. Specifically, for valid vessel pixels the within-band attenuation is typically small to moderate, and we use $\exp(-x) \approx 1 - x$ together with $-\ln(1-u) \approx u$ to obtain

$$OD_c(p) \approx h(p)\, l(p) \sum_{\lambda \in \Lambda_c} W_c(\lambda)\, \varepsilon(\lambda) \;\triangleq\; \bar{\varepsilon}_c\, h(p)\, l(p), \tag{S5}$$

where the effective band-averaged extinction for channel c is defined as

$$\bar{\varepsilon}_c \triangleq \sum_{\lambda \in \Lambda_c} W_c(\lambda)\, \varepsilon(\lambda).$$

This approximation is a modeling simplification rather than a calibrated physical identity, and its accuracy may degrade in low-signal or strongly attenuated regions; in

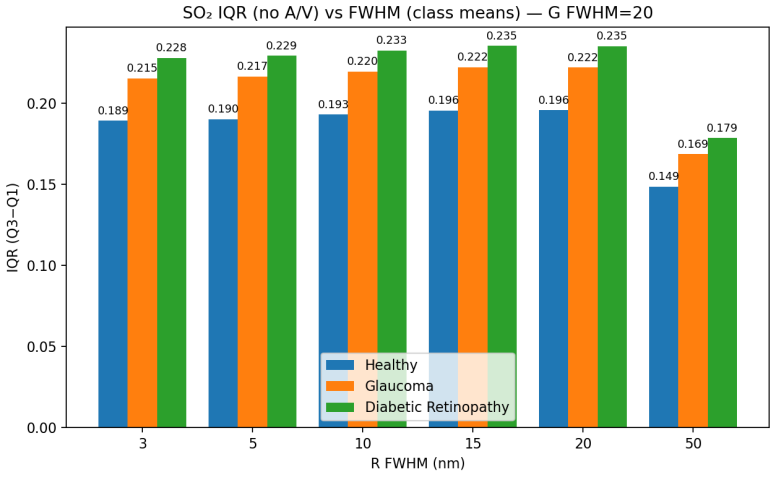
practice, such cases are mitigated by masking and thresholding in our OD/SO_2-proxy pipeline.

S3 RGB-derived appearance feature analysis

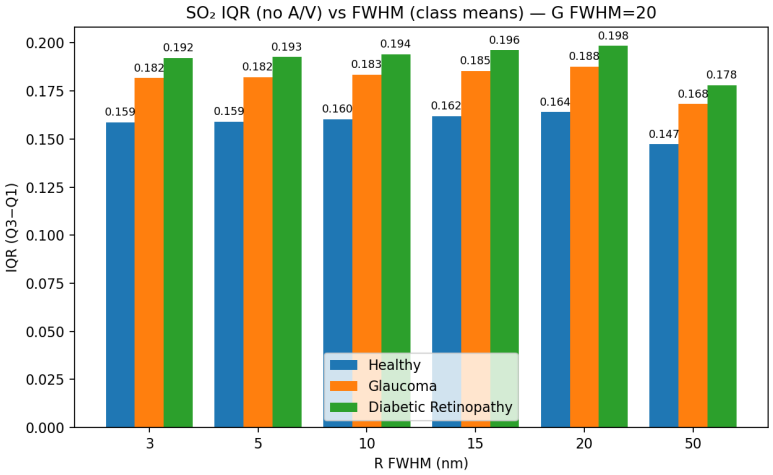
S3.1 Robustness of OD-Derived SO_2-Proxy Features to Assumed Sensor Bandwidth

To approximate RGB sensor spectral responsivities, we model the blue, green, and red channels as Gaussian band-pass profiles with nominal center wavelengths $\lambda_B = 465\,\text{nm}$, $\lambda_G = 545\,\text{nm}$, and $\lambda_R = 610\,\text{nm}$, respectively. Channel bandwidth is parameterized by the FWHM (full width at half maximum). We use $\text{FWHM}_B = 45\,\text{nm}$, $\text{FWHM}_G = 20\,\text{nm}$, and $\text{FWHM}_R = 20\,\text{nm}$. The broader blue band provides additional spectral support at shorter wavelengths, where haemoglobin absorption is typically stronger.

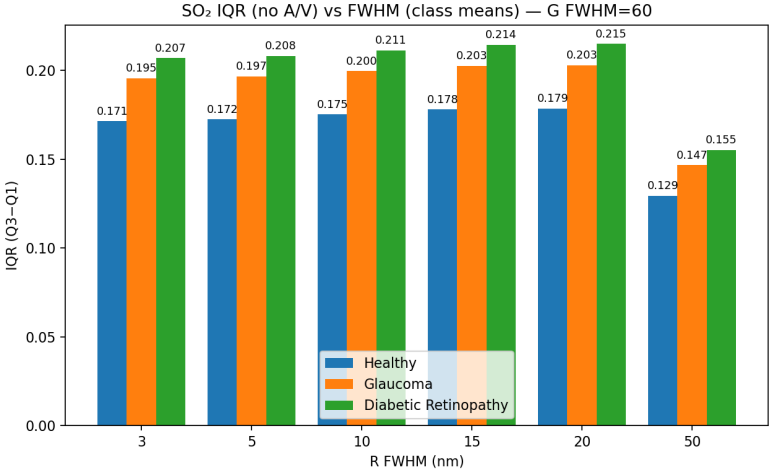
For the HRF dataset, Supplementary Fig. S1–S3 report the class-wise mean interquartile range (IQR) of the resulting SO_2-proxy index as the red-channel FWHM is varied, under different choices of the red-band center wavelength and green-band bandwidth. Across the tested settings, the mean IQR values for the three disease classes remain generally separated, and the overall trend is largely preserved, suggesting that the SO_2-proxy variability metric is not highly sensitive to moderate changes in the assumed red-channel bandwidth. Since artery–vein labeling is not consistently available in public fundus datasets, these HRF results further support the use of vessel-wide IQR, without artery/vein separation, as a practical group-level descriptor for retinal disease discrimination.



Supplementary Fig. S1 HRF dataset: Class-wise mean SO_2-proxy IQR versus red-channel FWHM with $\text{FWHM}_G = 20\,\text{nm}$ and $\lambda_R = 600\,\text{nm}$.



Supplementary Fig. S2 HRF dataset: Class-wise mean SO_2-proxy IQR versus red-channel FWHM with $\mathrm{FWHM}_G = 20\,\mathrm{nm}$ and $\lambda_R = 610\,\mathrm{nm}$.



Supplementary Fig. S3 HRF dataset: Class-wise mean SO_2-proxy IQR versus red-channel FWHM with $\mathrm{FWHM}_G = 60\,\mathrm{nm}$ and $\lambda_R = 600\,\mathrm{nm}$.

S3.2 Dataset-Dependent Distribution of RGB Color and SO_2-Proxy Features

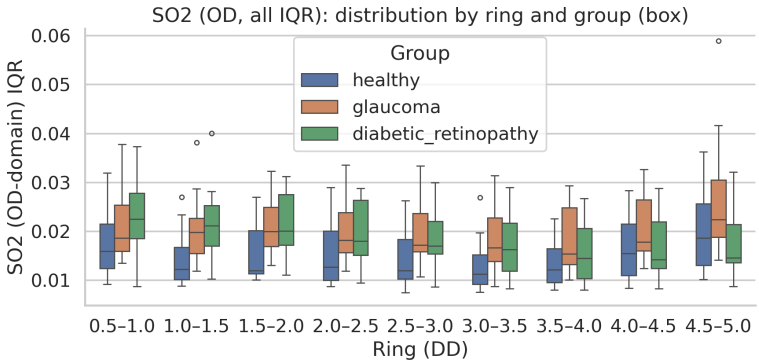
We further examined whether RGB-derived appearance descriptors, including raw RGB color features and OD-derived SO_2-proxy features, were stable across datasets acquired under different imaging conditions. This analysis was motivated by the fact that the proposed SO_2-proxy descriptors are estimated from RGB fundus images using approximate channel passbands and haemoglobin extinction coefficients, rather than

from calibrated multispectral retinal oximetry. Therefore, their absolute values may be influenced by camera spectral response, illumination, white balance, pigmentation, image compression, and dataset-specific post-processing.

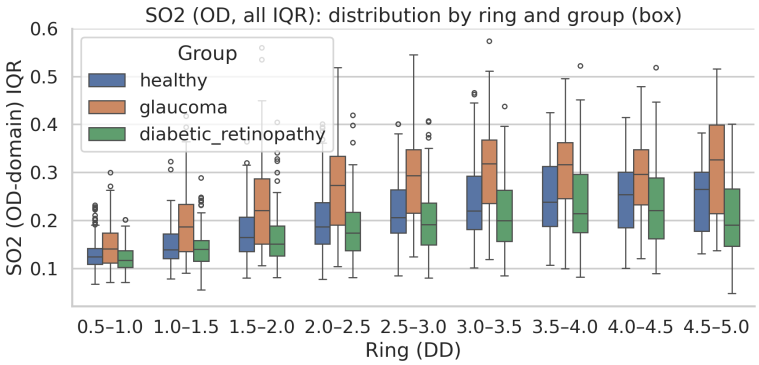
Supplementary Fig. S4 and Fig. S5 show the ring-wise distributions of the OD-derived SO_2-proxy IQR feature in HRF and FIVES, respectively. The two datasets exhibit substantially different absolute scales. In HRF, SO_2-proxy IQR values are relatively compressed across annuli, and the class ordering is ring-dependent rather than globally monotonic. In contrast, FIVES shows larger SO_2-proxy IQR values and a more pronounced increase from the inner to outer annuli. This cross-dataset difference indicates that the absolute scale of the RGB-derived SO_2-proxy descriptor is dataset-dependent and should not be interpreted as calibrated retinal oxygen saturation.

Supplementary Fig. S6 further compares RGB color and SO_2-proxy features between FIVES and HRF. Representative images from the two datasets show visible differences in color appearance. The ring-wise R/G ratio also shows a clear dataset-level shift, with HRF generally occupying a higher value range than FIVES. Similarly, the absolute SO_2-proxy variation shows a large dataset-level shift between the two datasets. These results indicate that both raw RGB color descriptors and absolute RGB-derived SO_2-proxy descriptors are influenced by dataset-specific color domains.

Within each dataset, however, SO_2-proxy features still show disease-associated variation. In HRF, diabetic retinopathy tends to show relatively higher values in several inner rings, but overlaps more with healthy and glaucoma in some outer rings. In FIVES, glaucoma generally shows higher SO_2-proxy variability than healthy and diabetic retinopathy across multiple rings. These patterns suggest that the descriptor can capture relative disease-associated color attenuation variation within a dataset, even though its absolute scale is not directly transferable across datasets.



Supplementary Fig. S4 HRF dataset: Ring-wise distribution of the OD-derived SO_2-proxy IQR feature across healthy, glaucoma, and diabetic retinopathy groups. The values are relatively compressed across annuli, and the class ordering is ring-dependent rather than globally monotonic.



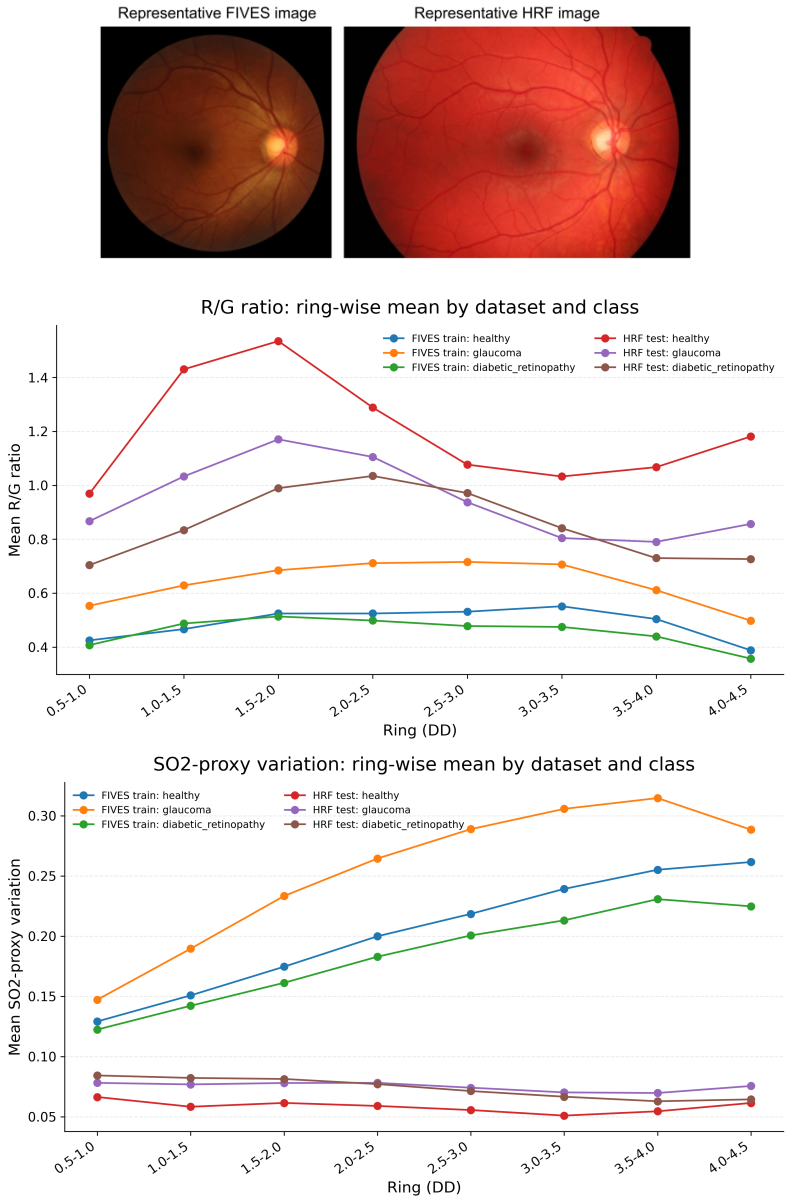
Supplementary Fig. S5 FIVES dataset: Ring-wise distribution of the OD-derived SO_2-proxy IQR feature across healthy, glaucoma, and diabetic retinopathy groups. Compared with HRF, the values show a larger absolute scale and a more pronounced increase from the inner to outer annuli, indicating dataset-dependent behavior of the RGB-derived proxy.

S3.3 Robustness to Illumination Perturbation

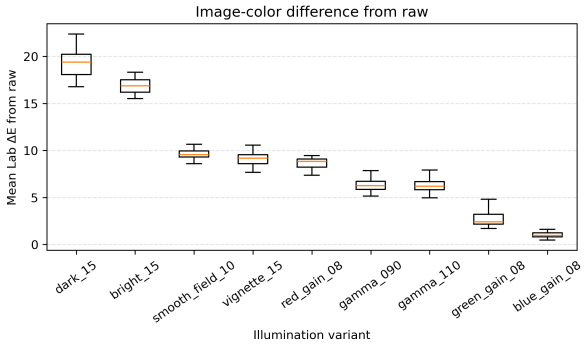
To assess the robustness of the OD-derived SO_2-proxy features to photometric variation, we applied controlled illumination and color perturbations to the HRF fundus images. Vessel masks, optic-disc annotations, and annular regions were kept fixed so that the analysis isolated the effect of image appearance changes on the extracted features and downstream classification. The perturbations included global brightness changes, gamma adjustment, channel-gain changes, vignetting, and smooth illumination-field variation. Channel-gain perturbations were implemented by multiplying one RGB channel by a fixed factor while leaving the other channels unchanged, thereby simulating color-balance or sensor-response shifts. These perturbations should be interpreted as controlled photometric stress tests rather than exact simulations of a specific camera or acquisition protocol.

Supplementary Fig. S7 summarizes the feature-level effect of these perturbations. The Lab ΔE analysis confirms that the simulated variants introduced measurable image-level color and illumination differences relative to the raw images. The two OD-derived SO_2-proxy features used in classification, namely the SO_2-proxy IQR and the SO_2-proxy artery–vein median ratio, also showed measurable changes under several perturbations. Among the tested variants, global brightness and gamma changes produced relatively larger SO_2-proxy feature changes, whereas the blue-channel gain perturbation had little effect. Spatially varying perturbations, including vignetting and smooth illumination-field variation, produced smaller but still measurable changes. These results indicate that OD-derived SO_2-proxy measurements can be affected by photometric variation at the feature level, even when vessel masks, optic-disc annotations, and annular regions are unchanged.

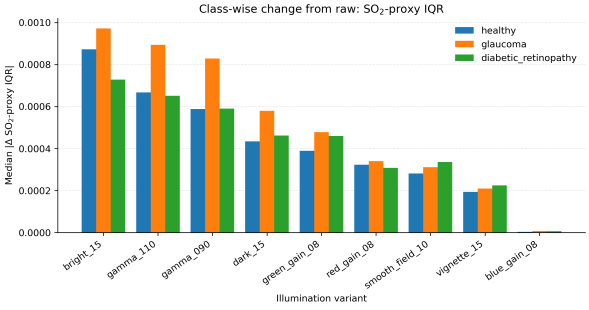
We also examined the RGB-derived vessel color feature used in the classifier. Unlike the OD-derived SO_2-proxy features, this feature directly summarizes the R/G ratio within the vessel region and therefore provides a complementary color-based descriptor.



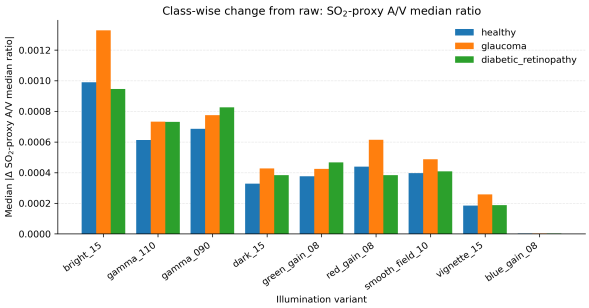
Supplementary Fig. S6 Distribution of RGB color and SO_2-proxy features across FIVES and HRF. Top: Representative FIVES and HRF images show visible dataset-level appearance differences. Middle: Ring-wise R/G ratio distributions show clear dataset-level shifts. Bottom: Ring-wise absolute SO_2-proxy variation also shows a large dataset-level shift.



(a) Image-color difference

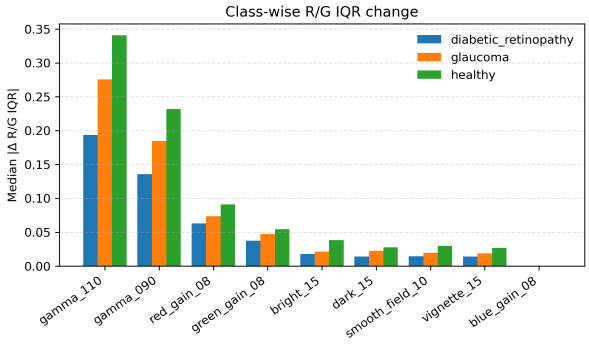


(b) SO_2-proxy IQR change



(c) SO_2-proxy A/V ratio change

Supplementary Fig. S7 Illumination sensitivity analysis for OD-derived SO_2-proxy features. Controlled photometric perturbations were applied to the fundus images while keeping vessel masks, optic-disc annotations, and annular regions fixed. (a) Image-level Lab ΔE shows that the perturbations introduced measurable color and illumination differences relative to the raw images. (b) Class-wise absolute changes in the OD-derived SO_2-proxy IQR feature were quantified across illumination variants. (c) Class-wise absolute changes in the OD-derived SO_2-proxy artery–vein median ratio were quantified. Both SO_2-proxy summaries showed measurable feature-level changes under several perturbations.



Supplementary Fig. S8 Illumination sensitivity analysis for the RGB-derived R/G IQR feature. Class-wise absolute changes in the R/G IQR feature were quantified across the same photometric perturbations used in the SO_2-proxy analysis. Although the magnitude of the R/G feature change varied across illumination variants, the class-wise patterns were broadly preserved. This suggests that RGB-derived color features may provide complementary information to OD-derived SO_2-proxy features under some perturbations.

As shown in Supplementary Fig. S8, the R/G IQR feature also changed under illumination perturbations, but its response pattern was not identical to that of the SO_2-proxy features. This supports the interpretation that the RGB-derived vessel color feature provides complementary information rather than duplicating the OD-derived SO_2-proxy descriptors.

Supplementary Table S2 Classification accuracy under illumination perturbations using different feature groups on HRF. SO_2-proxy denotes the two OD-derived SO_2-proxy features, vessel color denotes the RGB-derived R/G feature group, and all features denotes the full ring-wise representation.

Illumination variant	SO_2-proxy	Vessel color	All features
Raw	0.644	0.511	1.000
Bright +15%	0.578	0.600	0.978
Dark −15%	0.644	0.533	0.978
Gamma 0.90	0.622	0.556	1.000
Gamma 1.10	0.644	0.511	0.978
Red gain 0.80	0.600	0.556	1.000
Green gain 0.80	0.644	0.511	0.978
Blue gain 0.80	0.644	0.511	1.000
Smooth field	0.622	0.600	0.978
Vignette	0.644	0.556	1.000

Supplementary Table S2 summarizes the downstream classification results under the same perturbations. When only the OD-derived SO_2-proxy features were used, accuracy varied across variants, ranging from 0.578 to 0.644 compared with 0.644 for the raw images. The RGB-derived vessel color feature group also showed variable performance, ranging from 0.511 to 0.600 compared with 0.511 for the raw images.

Importantly, its response to the perturbations differed from that of the SO_2-proxy features. For example, under the brightening and smooth-field perturbations, vessel-color accuracy increased relative to the raw setting, whereas SO_2-proxy accuracy decreased.

In contrast, the full feature set remained substantially more stable, with accuracy ranging from 0.978 to 1.000 across all variants. These results suggest that the full ring-wise representation does not depend on a single photometrically affected feature group. Instead, complementary information from vascular geometry, vessel density, vessel color, OD-derived SO_2-proxy, vessel–background entropy, and connectivity descriptors helps maintain classification performance under controlled illumination perturbations.

S3.4 Ablation of OD-Derived SO_2-Proxy Features

The feature-family ablation experiments were designed to assess the contribution of the OD-derived SO_2-proxy features to the proposed classifier. In these experiments, the SO_2-proxy feature group was removed from the full ring-structured representation while all other feature groups were retained. The full-feature setting corresponds to the proposed method using predicted vessel masks generated by RIP-AV [3]. The ablated setting removes only the SO_2-proxy feature group while retaining the remaining vascular density, geometry, vessel color, vessel–background entropy, and graph-based connectivity descriptors. This analysis evaluates whether the SO_2-proxy descriptors provide complementary information beyond the other feature families.

Supplementary Table S3 summarizes the results. On HRF, removing the SO_2-proxy features reduced the accuracy from 0.911 to 0.778, corresponding to a drop of 0.133. This was the largest observed degradation among the HRF feature-family ablations, indicating that the SO_2-proxy descriptors provide substantial complementary information in this dataset. Importantly, this does not imply that the descriptor is a calibrated or globally stable oxygenation biomarker. Rather, it suggests that the ring-wise SO_2-proxy pattern contains information not fully captured by the remaining structural and appearance features.

On FIVES, removing the SO_2-proxy features reduced accuracy from 0.724 to 0.703, corresponding to a smaller drop of 0.021. This result is consistent with the feature-importance analysis shown in Fig. S23b, where SO_2-proxy features ranked highly but appeared alongside strong RGB color features. Thus, in FIVES, the SO_2-proxy descriptors may be discriminative but partially redundant with other appearance-based features. On SYSU, removing the SO_2-proxy features reduced accuracy only slightly, from 0.894 to 0.884, corresponding to a drop of 0.010. This suggests that the two-class classification task in this dataset is less dependent on this feature group. Overall, the ablation results indicate that OD-derived SO_2-proxy features can contribute useful complementary information, especially in HRF.

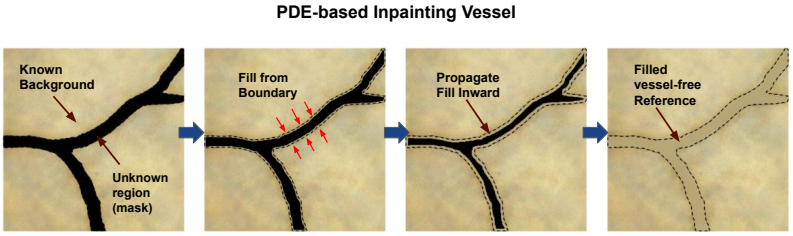
S4 Construction of the Vessel-Free Background Reference

This section describes the construction of a smooth vessel-free background reference map. First, the vessel region is dilated and removed to define the masked area. The resulting missing region is then inpainted using nearby non-vessel pixels

Supplementary Table S3 Feature-family ablation analysis of OD-derived SO_2-proxy features using predicted vessel masks generated by RIP-AV [3]. The full-feature setting corresponds to the proposed method with all feature groups included, whereas the ablated setting removes only the SO_2-proxy feature group.

Dataset	Feature set	Accuracy	F1	95% CI
HRF	All features	0.911	0.910	[0.793, 0.965]
	Without SO_2-proxy	0.778	0.778	[0.637, 0.875]
FIVES	All features	0.724	0.700	[0.677, 0.766]
	Without SO_2-proxy	0.703	0.697	[0.655, 0.746]
SYSU	All features	0.894	0.887	[0.873, 0.911]
	Without SO_2-proxy	0.884	0.883	[0.863, 0.902]

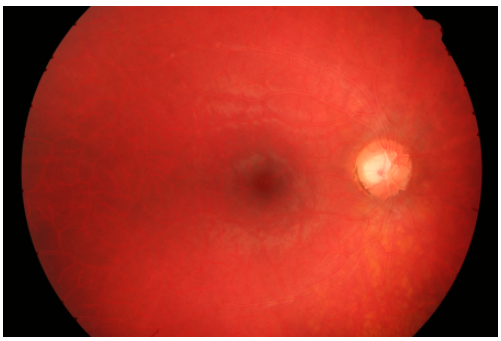
with OpenCV's Navier–Stokes partial differential equation (PDE)-based inpainting method. Finally, large-scale Gaussian smoothing is applied to obtain a smooth background reference. Supplementary Fig. S9 illustrates the overall procedure. Supplementary Fig. S10 illustrates an example.



Supplementary Fig. S9 Illustration of vessel-free background reference construction. The dilated vessel region is treated as the unknown masked area, while surrounding non-vessel pixels provide the known background. OpenCV's Navier–Stokes PDE-based inpainting propagates boundary information inward to fill the masked vessel region, producing a vessel-free background estimate. A subsequent large-scale Gaussian smoothing step is then applied to generate the final smooth background reference map.

S5 Photometric normalization and vessel-enhancing preprocessing

This section provides implementation details for the photometric normalization and vessel-enhancing preprocessing summarized in the main manuscript. The preprocessing is used to improve robustness to illumination variability and to generate the normalized green-channel image used for vessel–background texture/entropy measurements. In contrast, OD-derived SO_2-proxy features and ratio-based vessel color features are computed from the original RGB image to preserve raw vessel-to-background intensity relationships.



Supplementary Fig. S10 Example of the vessel-free background reference used for pixel-wise OD computation. The dilated vessel region is removed and filled by inpainting, followed by large-scale Gaussian smoothing to produce a smoothly varying background estimate within vessels.

To improve robustness to illumination variability and enhance vascular contrast, we apply a three-stage preprocessing pipeline to each RGB fundus image. Let $I \in \mathbb{R}^{H\times W\times 3}$ denote the input RGB image, with channel intensities $I_c(x)$ for $c \in \{R, G, B\}$ at pixel location x.

(1) Illumination field estimation and ratio correction.

Fundus images commonly exhibit low-frequency shading (e.g., vignetting) and spatially varying illumination. For each channel, we estimate a smooth illumination field by Gaussian blurring:

$$B_c(x) = \mathcal{G}_{\sigma_{\text{bg}}} * I_c(x), \tag{S6}$$

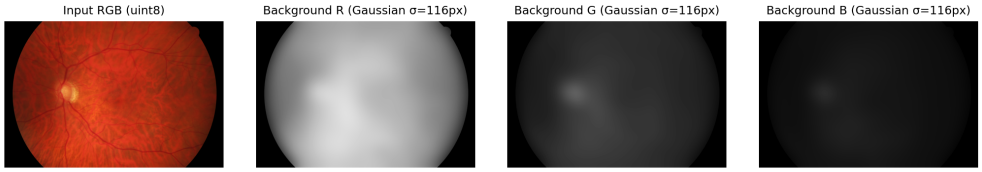
where $\mathcal{G}_\sigma$ denotes a Gaussian kernel with standard deviation σ (in pixels) and $*$ denotes convolution. Supplementary Fig. S11 visualizes the input image and the corresponding blurred background fields B_c for each color channel. We set σ_{bg} proportional to the image size,

$$\sigma_{\text{bg}} = \max\Big(31,\ \big\lfloor 0.05 \cdot \min(H, W)\big\rfloor\Big), \tag{S7}$$

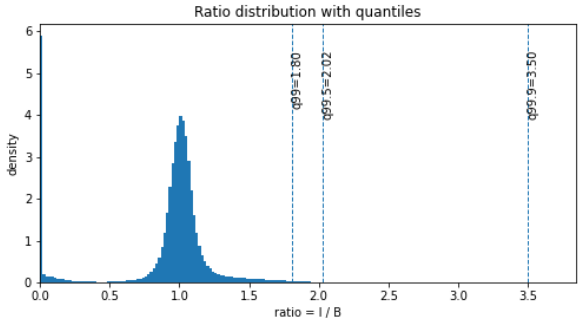
so that B_c captures only very low-frequency illumination trends while preserving anatomical structures such as vessels and the optic disc boundary; the minimum value prevents undersmoothing on smaller images. We then perform ratio-based correction,

$$I'_c(x) = \text{clip}\left(\frac{I_c(x)}{B_c(x)},\ 0,\ 4\right), \tag{S8}$$

where the lower and upper bounds truncate rare outliers in the ratio map that typically arise from saturated pixels or very dark regions with small values of $B_c(x)$. In our data, the ratio distribution $\{I_c(x)/B_c(x)\}$ is concentrated near 1, and even high-percentile ratio values remain below 4 (e.g., the 99.9-th percentile within the field of view is $q_{99.9} \approx 3.5$ in Supplementary Fig. S12). Thus, clipping at 4 preserves typical contrast while preventing extreme ratios from dominating the subsequent rescaling. Finally, the corrected image I' is linearly rescaled to $[0, 255]$ for subsequent color-space processing.



Supplementary Fig. S11 Illumination field estimation by large-σ Gaussian blurring. Shown are the input RGB image and the estimated smooth background fields B_c for the R/G/B channels (here $\sigma_{\mathrm{bg}} = 116$ px), which capture low-frequency illumination trends.



Supplementary Fig. S12 Histogram of the per-pixel ratio $I_c(x)/B_c(x)$ used for illumination correction, with vertical lines indicating high-percentile values (q_{99}, $q_{99.5}$, $q_{99.9}$). Most pixels yield ratios close to 1, while only a small fraction forms a right-tail of extreme values; this motivates clipping the ratio prior to rescaling to reduce sensitivity to outliers.

(2) Local contrast enhancement on lightness.

To enhance local vessel-to-background contrast while largely preserving chromatic relationships, we convert the illumination-corrected image I' to CIE Lab and apply contrast-limited adaptive histogram equalization (CLAHE) to the lightness channel L:

$$L' = \mathrm{CLAHE}\big(L;\ \texttt{clipLimit} = 2.0,\ \texttt{tileGridSize} = (8, 8)\big). \tag{S9}$$

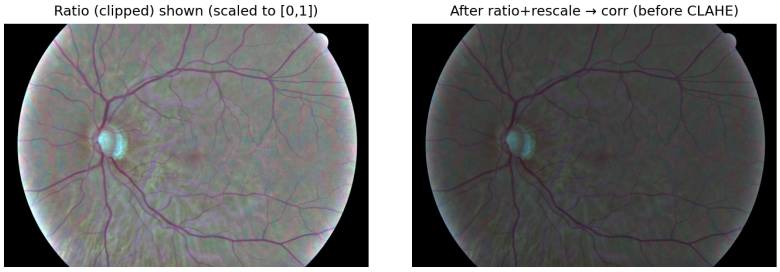
In our implementation, this corresponds to the OpenCV routine `createCLAHE` with `clipLimit`=2.0 and `tileGridSize`=$(8, 8)$, followed by `clahe.apply`(L) on the 8-bit lightness image. We then reconstruct the RGB image by combining (L', a, b) and converting back to RGB (Supplementary Fig. S14).

Principle. CLAHE is a local extension of histogram equalization. Instead of computing a single global intensity mapping, it partitions the image into contextual tiles, computes a histogram within each tile, and uses the tile cumulative distribution function (CDF) to form a local monotonic intensity remapping that spreads frequently occurring intensities over a wider output range. To mitigate noise over-amplification in near-uniform regions, which is a known limitation of adaptive histogram equalization, CLAHE introduces *contrast limiting*: histogram bins are clipped according to `clipLimit`, the clipped counts are redistributed across bins, and the resulting clipped

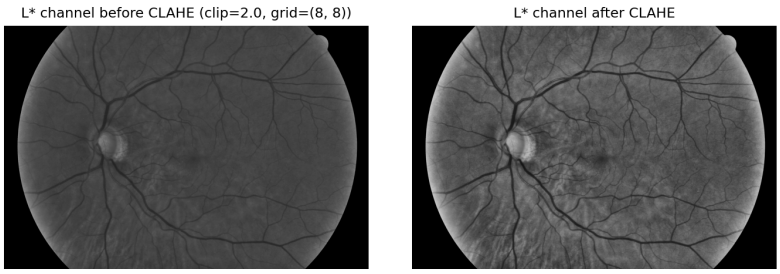
histogram is used to compute the CDF. Finally, to prevent discontinuities at tile boundaries, the per-tile mappings are blended via bilinear interpolation.

As illustrated in Fig. S13, applying CLAHE to the Lab lightness channel L^* enhances local vessel–background contrast (Supplementary Fig. S13b) while leaving the chromatic components (a, b) unchanged. Supplementary Fig. S13a additionally visualizes the illumination correction step: the clipped ratio map is shown for inspection after scaling to $[0, 1]$ for display, and the corresponding ratio-corrected image after linear rescaling to $[0, 255]$ provides the 8-bit input required for subsequent Lab conversion and CLAHE.

The `tileGridSize` parameter determines the locality of the operation by dividing the image into an 8×8 grid of contextual tiles and blending their mappings via interpolation; using too many tiles (very small regions) can lead to a patchy appearance, whereas too few tiles reduces locality and may weaken enhancement of fine vessels. The `clipLimit` parameter performs contrast limiting by clipping histogram peaks within each tile before computing the cumulative mapping, thereby reducing the risk of amplifying noise and small spurious variations; we use `clipLimit`=2.0 as a moderate setting that yields clear contrast enhancement without a noticeably grainy or over-enhanced appearance in the intermediate visualizations.

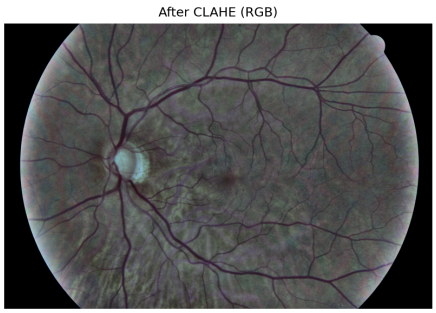


(a) Intermediate visualization of the ratio-corrected image prior to CLAHE (RGB domain).



(b) Lightness channel L^* before and after CLAHE, showing increased local contrast after equalization.

Supplementary Fig. S13 Step-wise visualization of the CLAHE stage. CLAHE is applied to the Lab lightness channel L^* and the image is reconstructed in RGB using (L', a, b).



Supplementary Fig. S14 RGB image reconstructed after applying CLAHE to the Lab lightness channel.

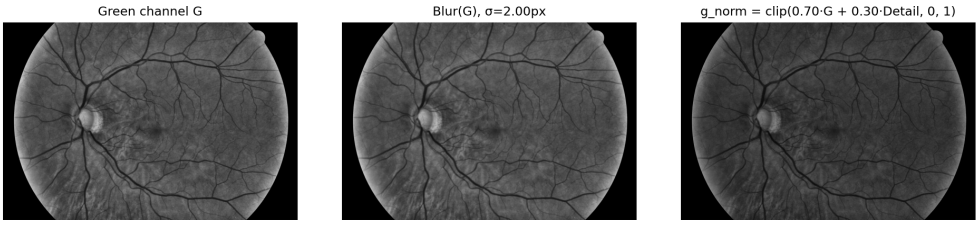
(3) Green-channel unsharp blending.

Because the green channel typically provides strong retinal vessel-to-background contrast, we extract the green channel from the CLAHE-corrected RGB image and normalize it to $G \in [0, 1]$. We compute a small-scale blurred version $\tilde{G} = \mathcal{G}_{\sigma_{\text{sharp}}} * G$ with $\sigma_{\text{sharp}} = 2$ pixels, and form an unsharp blend:

$$G_{\text{norm}} = \text{clip}\Big(w_{\text{base}} G + w_{\text{detail}}(G - \tilde{G}),\ 0,\ 1\Big), \tag{S10}$$

with $w_{\text{base}} = 0.7$ and $w_{\text{detail}} = 0.3$. where $*$ denotes two-dimensional convolution. This operation preserves the low-frequency brightness structure (via $w_{\text{base}} G$) while moderately enhancing fine-scale detail and vessel edges through the term $G - \tilde{G}$. The final clipping ensures a valid normalized range for downstream processing. Supplementary Fig. S15 visualizes the intermediate images.

The parameter σ_{sharp} controls the spatial scale of the high-pass component $G - \tilde{G}$, where $\tilde{G} = \mathcal{G}_{\sigma_{\text{sharp}}} * G$. Smaller values make $G - \tilde{G}$ dominated by very high-frequency content (including fine texture and noise), whereas larger values include broader edge neighborhoods and may introduce halo (edge-ringing) artifacts when added back. We use $\sigma_{\text{sharp}} = 2$ pixels to enhance vessel edges while avoiding visible halo artifacts.



Supplementary Fig. S15 Green-channel unsharp blending. Left: normalized green channel G. Middle: blurred version $\tilde{G} = \mathcal{G}_{\sigma_{\text{sharp}}} * G$ with $\sigma_{\text{sharp}} = 2$ pixels. Right: unsharp blend $G_{\text{norm}} = \text{clip}(0.70\, G + 0.30\, (G - \tilde{G}), 0, 1)$, which mildly enhances vessel edges while preserving overall intensity.

S6 Ring-Wise Feature Distributions

Supplementary Fig. S16 and Fig. S17 show the ring-wise distributions of geometry-based vessel features, vessel appearance features, and vessel–background entropy/texture features across different disease groups.

S7 Summary of feature statistics and rationale

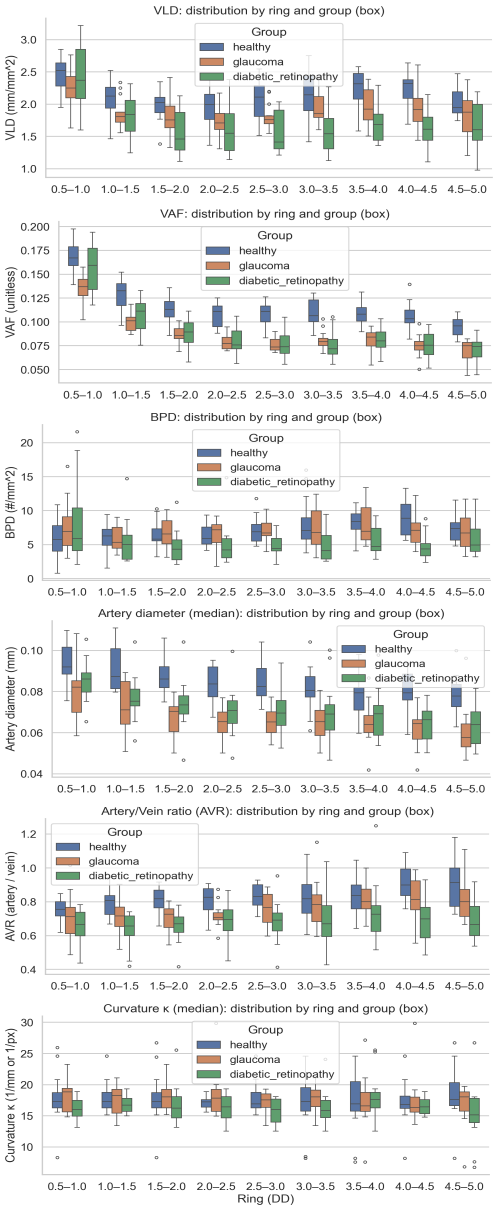
Supplementary Table S4 summarizes each ring-wise feature, the statistic used for summarization, and the corresponding rationale.

Supplementary Table S4 Summary of ring-wise features, the statistic used for summarization, and the corresponding rationale.

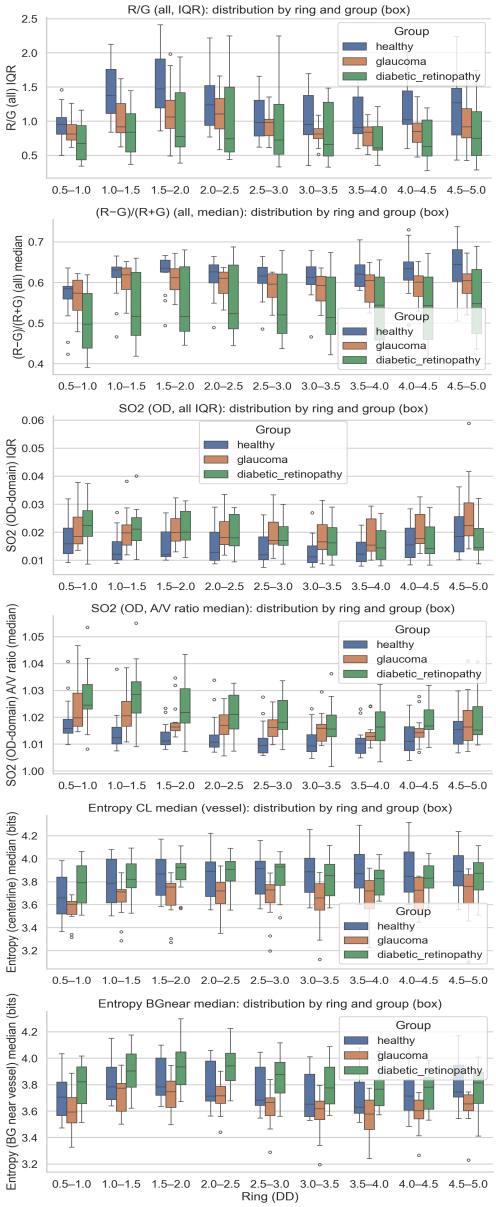
Feature	Statistic	Rationale
Vessel area fraction (VAF)	Raw value	Direct vessel area density within the ring.
Vessel length density (VLD)	Raw value	Direct skeleton length density within the ring.
Branch-point density (BPD)	Raw value	Direct junction density within the ring.
Vessel caliber	IQR	Robust to isolated unusually thick trunks or noisy thin branches.
Artery–vein ratio (AVR)	Ratio of medians	Stable summary of the typical relative artery–vein caliber.
Thin-vessel length	Raw value	Direct measure of small-caliber vessel content within the ring.
Diameter ratio	Ratio of medians	Robust contrast between thicker- and thinner-vessel subsets.
Vessel tortuosity	IQR	Captures within-ring heterogeneity of segment tortuosity.
Curvature	Median	Typical local bending magnitude; robust to isolated sharp turns and artifacts.
Vascular edge count	Raw value	Direct structural count of vascular connectivity within the ring.
Mean vascular connectivity	Mean	Standard graph-theoretic summary of local network connectivity.
SO_2-proxy feature	IQR	Captures within-ring heterogeneity of oxygenation-related appearance.
SO_2-proxy artery/vein contrast	Ratio of medians	Stable summary of typical artery–vein oxygenation-related contrast.
Red/green ratio	IQR	Captures within-ring variability in vessel color appearance.
Normalized red/green difference	Median	Typical per-pixel normalized red–green contrast within the ring.
Entropy on vessel centerlines	Median	Typical vessel-centerline texture complexity; robust to noisy pixels.
Entropy in near-vessel background	Median	Typical adjacent background texture for vessel–background contrast analysis.

S8 Controlled Non-Vessel Background Perturbations and Lesion-Aware Inpainting

This section describes the controlled perturbations used to examine the sensitivity of pretrained RETFound [4] to non-vessel information in the FOV-standardized FIVES



Supplementary Fig. S16 Ring-wise distributions of geometry-based vessel features across disease groups. Box plots show the distributions of representative vascular geometry features: VLD, VAF, BPD, median vessel caliber, AVR, and vessel tortuosity, computed within concentric annular rings centered at the OD. Ring radii are defined relative to the optic disc diameter. Results are shown for healthy, glaucoma, and DR fundus images, illustrating spatially varying and class-dependent differences in fundus vascular morphology.



Supplementary Fig. S17 Ring-wise distributions of vessel appearance and texture contrast features. Box plots summarize color-based vessel appearance features (red–green intensity ratios), SO_2-proxy appearance features, and vessel–background entropy features. Features are computed within concentric annular rings relative to the optic disc diameter and shown separately for healthy, glaucoma, and DR groups. These distributions highlight complementary appearance and texture patterns across rings that are not captured by geometric vessel descriptors alone.

images. We considered two complementary groups of controls: broad background-appearance perturbations that modified non-vessel color, brightness, and contrast, and lesion-aware inpainting that selectively suppressed automatically detected lesions in non-vessel regions. Together, these controls assess sensitivity to broad background appearance, localized lesion-related information, or a mixture of both.

S8.1 Editable Non-Vessel Background Region

Let $I : \Omega \to \mathbb{R}^3$ denote an RGB fundus image, $V : \Omega \to \{0, 1\}$ the vessel mask, and $F : \Omega \to \{0, 1\}$ the common FOV mask. The editable non-vessel background region is defined as

$$\Omega_{\text{bg}} = \{p \in \Omega : F(p) = 1,\ D(V)(p) = 0,\ \mathbf{1}_{\Omega_{\text{disc}}}(p) = 0\}\,, \tag{S11}$$

where $D(V)$ denotes a dilated vessel mask used to exclude pixels near the vasculature. A dilation radius of 5 pixels was used. The optic-disc region Ω_{disc} was defined from the annotated disc center and diameter as a circular region with radius equal to half of the annotated disc diameter. Perturbations were applied only within Ω_{bg}; vessel pixels, pixels near vessels, the optic-disc region, and pixels outside the common FOV were preserved.

From all valid pixels in Ω_{bg} across the dataset, we computed a shared background RGB median

$$\mathbf{m}^{\star}_{\text{bg}} = (m^{\star}_R, m^{\star}_G, m^{\star}_B) \in \mathbb{R}^3 \tag{S12}$$

and a shared grayscale background median $g^{\star}_{\text{bg}} \in \mathbb{R}$. These dataset-level quantities were used as common targets for the controlled background perturbations.

S8.2 Broad Background-Appearance Perturbations

S8.2.1 Background color matching.

Background RGB appearance was standardized using an additive per-channel shift. Let

$$\mathbf{m}_{\text{bg}}(I) = \big(\text{median}_{\Omega_{\text{bg}}}(I_R), \text{median}_{\Omega_{\text{bg}}}(I_G), \text{median}_{\Omega_{\text{bg}}}(I_B)\big) \tag{S13}$$

denote the channel-wise background medians of image I. The transformed background is defined as

$$\widetilde{I}_c(p) = I_c(p) + (m^{\star}_c - m_{\text{bg},c}(I))\,, \qquad p \in \Omega_{\text{bg}}, \quad c \in \{R, G, B\}. \tag{S14}$$

This perturbation aligns the channel-wise background medians with the shared dataset-level RGB reference while retaining local spatial structure. Because the transformation is performed in RGB space, it should be interpreted as standardizing background RGB appearance rather than isolating chromaticity in a perceptually disentangled color space.

S8.2.2 Background brightness matching.

Let $G(I)$ denote the grayscale version of image I. A scalar brightness factor is computed as

$$s_{\mathrm{bg}} = \frac{g^{\star}_{\mathrm{bg}}}{\mathrm{median}_{\Omega_{\mathrm{bg}}}(G(I))}, \tag{S15}$$

with the factor restricted to a bounded range in the implementation. The background pixels are then rescaled as

$$\widetilde{I}_c(p) = s_{\mathrm{bg}} I_c(p), \qquad p \in \Omega_{\mathrm{bg}}, \quad c \in \{R, G, B\}. \tag{S16}$$

This perturbation modifies overall background intensity while preserving its spatial structure.

S8.2.3 Background contrast reduction.

Background contrast was reduced by shrinking channel-wise deviations toward the regional mean. Let $\mu_{\mathrm{bg},c}(I)$ denote the mean of channel c within Ω_{bg}. The transformed background is

$$\widetilde{I}_c(p) = \mu_{\mathrm{bg},c}(I) + \alpha_{\mathrm{bg}} \left[I_c(p) - \mu_{\mathrm{bg},c}(I) \right], \qquad p \in \Omega_{\mathrm{bg}}, \tag{S17}$$

where $\alpha_{\mathrm{bg}} \in (0, 1)$ controls the retained contrast. Moderate contrast compression was used in the present implementation. This operation reduces background intensity variation without explicitly smoothing fine spatial detail.

S8.2.4 Joint color, contrast, and brightness standardization.

The composite variant sequentially applies background color matching, contrast reduction, and brightness matching. It should therefore be interpreted as a joint perturbation of multiple background-appearance factors, including RGB appearance, intensity spread, and overall brightness, rather than as an isolated single-factor manipulation.

S8.2.5 Constant-background control.

To suppress image-specific background variation more strongly, we also replaced the editable non-vessel background with the shared RGB reference:

$$\widetilde{I}(p) = \mathbf{m}^{\star}_{\mathrm{bg}}, \qquad p \in \Omega_{\mathrm{bg}}. \tag{S18}$$

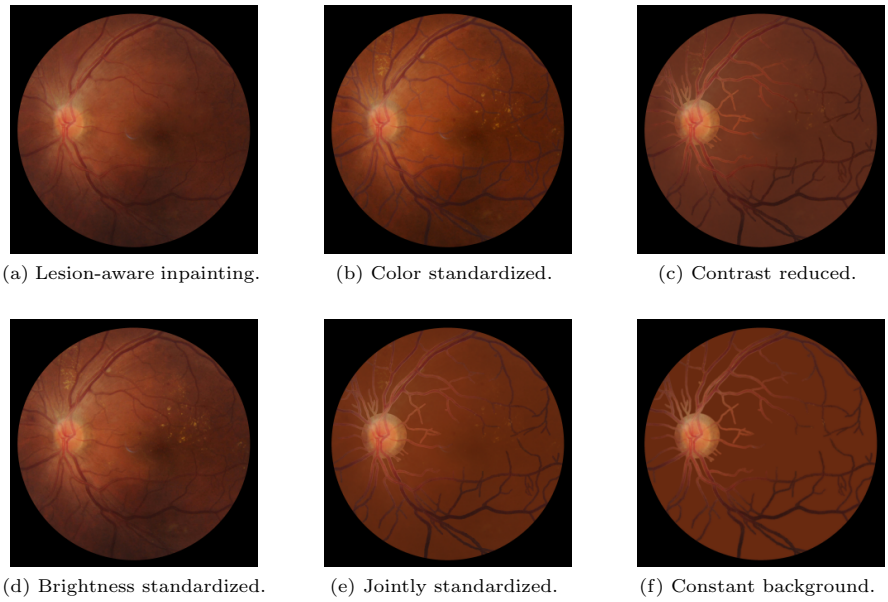
This variant removes most residual image-specific appearance variation within the editable background region and therefore serves as a strong control for evaluating the contribution of non-vessel background information.

S8.3 Lesion-Aware Background Inpainting

To examine the contribution of localized lesion-related information, common DR-related lesions, including exudates, hemorrhages, microaneurysms, and cotton-wool spots, were automatically segmented using a pretrained fundus-lesion segmentation method [5]. Only detected lesion pixels located within Ω_{bg} were inpainted using LaMa [6]. Vessel pixels, pixels near vessels, the optic-disc region, and other non-lesion pixels were preserved.

This variant should be interpreted as a lesion-suppression control rather than as complete removal of pathological information. Automatic lesion segmentation may not detect all abnormalities, and residual contextual or textural information may remain after inpainting.

Representative examples of all perturbation variants are shown in Supplementary Fig. S18.



(a) Lesion-aware inpainting. (b) Color standardized. (c) Contrast reduced.

(d) Brightness standardized. (e) Jointly standardized. (f) Constant background.

Supplementary Fig. S18 Representative non-vessel perturbations generated from a FIVES image. The lesion-aware variant selectively inpaints automatically detected lesions in non-vessel regions. The remaining variants modify broad non-vessel color, contrast, brightness, or their combination. For all variants, vessel pixels, pixels near vessels, the optic-disc region, and pixels outside the common FOV were preserved.

S8.4 Effects of Broad Background-Appearance Perturbations

Supplementary Table S5 summarizes RETFound performance for the broad background-appearance controls. Relative to the FOV-standardized baseline accuracy of 0.806, background color standardization produced the largest decrease among the

individual perturbations, reducing accuracy to 0.725. Contrast reduction and brightness standardization produced smaller decreases, to 0.763 and 0.760, respectively. Joint standardization of color, contrast, and brightness reduced accuracy to 0.682, similar to the constant-background result.

Supplementary Table S5 Classification performance of pretrained RETFound under controlled non-vessel perturbations in FIVES.

Input variant	Accuracy	Macro-F1
FOV-standardized baseline	**0.806**	**0.788**
Lesion-aware inpainting	0.803	0.787
Color standardized	0.725	0.718
Contrast reduced	0.763	0.757
Brightness standardized	0.760	0.758
Color + contrast + brightness standardized	0.682	0.678
Constant background	0.682	0.674

These results indicate that RETFound was sensitive to broad non-vessel appearance. However, the affected regions may contain both acquisition-related variation and clinically meaningful non-vessel pathology. The observed performance changes should therefore not be interpreted as isolating acquisition bias alone. In particular, broad background perturbations may alter lesion appearance, retinal texture, illumination, reflections, and pigmentation-related variation simultaneously.

S8.5 Effects of Lesion-Aware Inpainting

The lesion-aware analysis was conducted on the matched set of 381 images available for both the FOV-standardized and lesion-inpainted inputs. Compared with the FOV-standardized baseline, lesion-aware inpainting produced only minor changes in RETFound performance (Supplementary Table S6). Accuracy decreased from 0.806 to 0.803, macro-F1 decreased from 0.788 to 0.787, and macro ROC-AUC decreased from 0.931 to 0.930. Healthy and DR recall decreased slightly, whereas glaucoma recall increased slightly.

Supplementary Table S6 Effect of lesion-aware inpainting on RETFound performance after FOV standardization in FIVES.

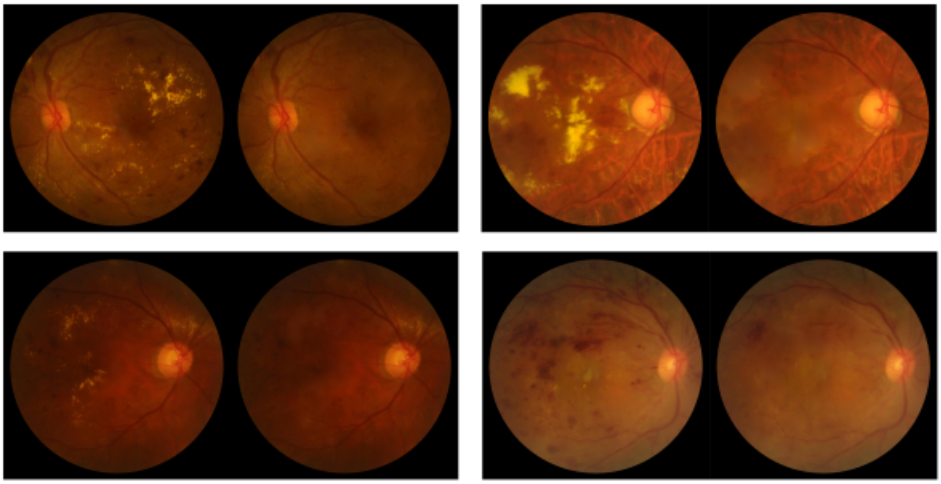
Input	Acc.	Macro-F1	Macro ROC-AUC	Healthy Rec.	Glaucoma Rec.	DR Rec.
FOV-std.	0.806 ± 0.020	0.788	0.931	0.857	0.651	0.838
Lesion-aware	0.803 ± 0.020	0.787	0.930	0.851	0.663	0.831
Change	-0.003	-0.001	-0.002	-0.006	$+0.012$	-0.008

Note: The lesion-aware variant was generated from the FOV-standardized images by segmenting visible non-vessel lesions and inpainting the detected regions. Metrics were computed on the matched set of 381 images available for both input conditions. FOV-std., FOV-standardized images; Rec., recall; DR, diabetic retinopathy. Accuracy is reported as the observed proportion correct ± binomial standard error.

Among the matched 381 images, seven predictions changed after lesion-aware inpainting. Four changed from correct to incorrect and three changed from incorrect to correct, reducing the number of correctly classified images from 307 to 306. For DR specifically, five of 130 predictions changed. Three correctly classified DR images were subsequently predicted as healthy, whereas two DR images previously predicted as healthy were corrected. Consequently, the number of correctly classified DR images decreased from 109 to 108, DR-to-healthy errors increased from 14 to 15, and DR-to-glaucoma errors remained unchanged at seven.

These mixed changes indicate that lesion-aware inpainting did not uniformly remove discriminative DR information. Suppressing the automatically detected lesion regions produced only a limited overall change, suggesting that RETFound predictions were not determined solely by these segmented lesions. Nevertheless, the limited effect does not establish independence from lesion-related information. Residual or undetected lesions, laser photocoagulation scars, local retinal texture, illumination artifacts, reflections, pigmentation-related variation, and other acquisition-dependent appearance characteristics may remain informative.

Because these factors may be clinically meaningful, acquisition-related, or both, the lesion-aware experiment should be interpreted as a targeted lesion-suppression control rather than a complete separation of pathology from acquisition bias. Representative before-and-after examples are shown in Supplementary Fig. S19.



Supplementary Fig. S19 Representative examples of lesion-aware background inpainting. For each example, the original FIVES image is shown on the left and the corresponding lesion-aware image is shown on the right. Visible non-vessel lesions are substantially suppressed, although residual local texture, diffuse background variation, and other non-vessel appearance information may remain after inpainting.

S8.6 Interpretation.

These controlled perturbations do not perfectly isolate individual image characteristics. The color, brightness, and contrast variants primarily target their corresponding background-appearance factors, whereas the composite and constant-background variants alter multiple aspects of non-vessel appearance. The lesion-aware variant suppresses automatically detected non-vessel pathology but depends on the completeness and accuracy of lesion segmentation and inpainting. Performance changes should therefore be interpreted as sensitivity to broader categories of non-vessel information, including acquisition-related background appearance and lesion-related pathology, rather than to perfectly isolated factors.

S9 Detailed Classification Results and Per-Class Metrics

To provide a more comprehensive evaluation beyond overall accuracy, we report additional classification results and disease-specific metrics for the proposed method and comparison baselines. Supplementary Table S7 reports per-class sensitivity, specificity, F1-score, and ROC-AUC. For multiclass datasets, sensitivity is computed as class-wise recall, specificity is computed by treating the target class as positive and all remaining classes as negative, and ROC-AUC is computed using a one-vs-rest formulation. For the binary SUSTech-SYSU dataset, the same definitions are applied to the healthy and diabetic retinopathy classes.

S10 Optic Disc Spatial Distribution Across Datasets

We further characterized acquisition-related spatial variability across the evaluated datasets by analyzing optic-disc center locations in HRF and SUSTech-SYSU. This analysis complements the FIVES acquisition-bias analysis and provides additional dataset-level characterization of spatial configuration differences, including left–right eye positioning and variation in optic-disc location within the fundus field of view.

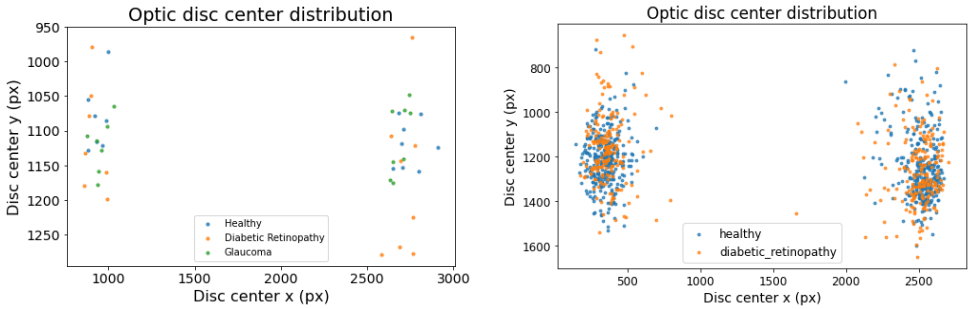
Supplementary Fig. S20 shows the optic disc center distributions for HRF and SUSTech-SYSU. HRF contains a smaller number of high-resolution fundus images, and the optic disc centers form two main spatial clusters corresponding to left- and right-eye images. In contrast, SUSTech-SYSU contains a much larger number of images and shows broader spatial variability in optic disc center locations. This broader distribution indicates greater acquisition heterogeneity and supports the use of SUSTech-SYSU as a larger-scale validation dataset beyond HRF.

Together with the FIVES spatial analysis, these results show that the proposed framework was evaluated across datasets with different image sizes, spatial positioning patterns, and acquisition characteristics. These plots characterize non-clinical spatial variation, such as left–right eye positioning and optic disc location, which may differ across datasets or disease groups and may influence model evaluation. Therefore, they provide context for the confounder-aware analyses performed in this study. However, these plots should be interpreted as dataset-characterization analyses rather than as direct evidence of clinical disease-related differences.

Supplementary Table S7 Detailed per-class classification metrics for the proposed method and comparison baselines.

Dataset	Method	Class	Sensitivity	Specificity	F1	ROC-AUC
HRF	Proposed (GT)	Healthy	1.000	1.000	1.000	1.000
		Glaucoma	1.000	1.000	1.000	1.000
		Diabetic retinopathy	1.000	1.000	1.000	1.000
	Proposed (Pred.)	Healthy	0.800	0.967	0.857	0.913
		Glaucoma	1.000	0.933	0.938	0.989
		Diabetic retinopathy	0.933	0.967	0.933	0.989
	RETFound	Healthy	0.933	0.900	0.875	0.956
		Glaucoma	0.933	0.967	0.933	0.976
		Diabetic retinopathy	0.867	1.000	0.929	0.969
	ViT-B/16	Healthy	0.733	0.833	0.710	0.873
		Glaucoma	0.933	0.933	0.903	0.969
		Diabetic retinopathy	0.733	0.933	0.786	0.918
	ConvNeXt-B	Healthy	0.933	0.833	0.824	0.940
		Glaucoma	0.933	0.967	0.933	0.976
		Diabetic retinopathy	0.733	1.000	0.846	0.982
	ResNet-50	Healthy	0.733	0.800	0.688	0.831
		Glaucoma	0.867	0.900	0.839	0.942
		Diabetic retinopathy	0.733	0.967	0.815	0.931
FIVES	Proposed (GT)	Healthy	0.899	0.779	0.825	0.902
		Glaucoma	0.590	0.923	0.632	0.868
		Diabetic retinopathy	0.708	0.924	0.763	0.881
	Proposed (Pred.)	Healthy	0.851	0.731	0.777	0.840
		Glaucoma	0.542	0.933	0.608	0.837
		Diabetic retinopathy	0.674	0.888	0.713	0.850
	RETFound	Healthy	0.857	0.821	0.823	0.920
		Glaucoma	0.651	0.923	0.675	0.919
		Diabetic retinopathy	0.838	0.948	0.865	0.954
	ViT-B/16	Healthy	0.756	0.808	0.756	0.851
		Glaucoma	0.675	0.869	0.629	0.855
		Diabetic retinopathy	0.800	0.944	0.839	0.930
	ConvNeXt-B	Healthy	0.786	0.808	0.774	0.876
		Glaucoma	0.651	0.903	0.651	0.882
		Diabetic retinopathy	0.823	0.928	0.839	0.917
	ResNet-50	Healthy	0.738	0.803	0.743	0.833
		Glaucoma	0.639	0.866	0.602	0.818
		Diabetic retinopathy	0.762	0.908	0.786	0.894
SUSTech-SYSU	Proposed (Pred.)	Healthy	0.933	0.834	0.917	0.947
		Diabetic retinopathy	0.834	0.933	0.858	0.947
	RETFound	Healthy	0.964	0.917	0.957	0.983
		Diabetic retinopathy	0.917	0.964	0.928	0.983
	ViT-B/16	Healthy	0.975	0.917	0.962	0.977
		Diabetic retinopathy	0.917	0.975	0.936	0.977
	ConvNeXt-B	Healthy	0.984	0.930	0.971	0.989
		Diabetic retinopathy	0.930	0.984	0.951	0.989
	ResNet-50	Healthy	0.960	0.917	0.955	0.977
		Diabetic retinopathy	0.917	0.960	0.925	0.977

Note: Sensitivity is computed as class-wise recall. Specificity is computed by treating the target class as positive and all remaining classes as negative. ROC-AUC is computed using a one-vs-rest formulation. All metrics are computed from out-of-fold predictions. GT, ground-truth vessel masks; Pred., predicted vessel masks. SUSTech-SYSU does not provide ground-truth vessel masks; therefore, only the predicted-mask result is reported for the proposed method.



(a) HRF. (b) SUSTech-SYSU.

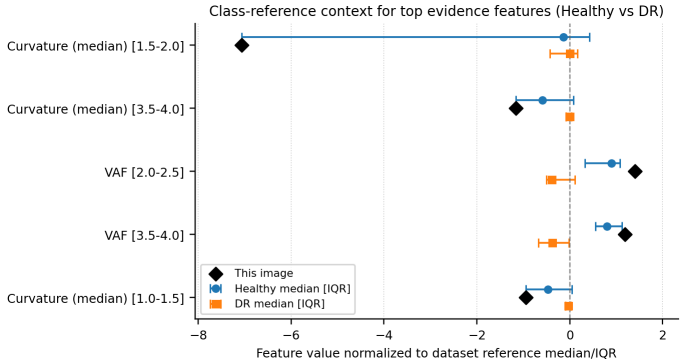
Supplementary Fig. S20 Optic disc center distributions for HRF and SUSTech-SYSU. Points indicate image-level optic disc center coordinates, colored by class. The vertical axis is inverted to follow image-coordinate convention. HRF shows two main clusters corresponding to left- and right-eye images, while SUSTech-SYSU shows broader spatial variability due to its larger sample size and more heterogeneous acquisition characteristics.

S11 Feature-Level Reference Context for Ring-Wise Evidence

To complement the qualitative and quantitative explanation comparison in the main manuscript, we further examined the feature-level context underlying the proposed ring-wise evidence heatmap. RETFound produces a spatial attribution map over image patches, whereas the proposed framework decomposes the predicted-vs-runner-up decision into explicit feature–ring contributions, as defined in the main manuscript. This allows the explanation to be linked not only to an annular retinal location, but also to measurable vascular descriptors such as vessel area fraction, curvature, caliber, color, entropy, and connectivity.

In the main manuscript, the top positive feature–ring contributions are shown quantitatively for the representative healthy example. To provide additional reference context for these top evidence features, we compared their raw feature values with class-specific distributions from the predicted and runner-up classes. As shown in Supplementary Fig. S21, the test-image feature values are plotted together with the median and interquartile range of the same features in the predicted healthy class and the runner-up diabetic retinopathy class. This visualization helps assess whether the feature values supporting the model decision are closer to the predicted-class distribution, closer to the runner-up-class distribution, or near the boundary of the class-reference ranges.

In this representative example, the vessel area fraction features are closer to the healthy reference distribution than to the diabetic retinopathy reference distribution, providing intuitive support for the predicted healthy class. The curvature-related features also generally align better with the healthy reference distribution than with the diabetic retinopathy reference distribution. In particular, Curvature (median) [3.5–4.0] and Curvature (median) [1.0–1.5] lie closer to the healthy reference distribution, while Curvature (median) [1.5–2.0] remains near the boundary of the healthy reference



Supplementary Fig. S21 Feature-level reference context for the representative healthy example used in the qualitative comparison. Black diamonds indicate the test-image feature values. Blue circles and error bars show the median and interquartile range (IQR) of the predicted healthy class, and orange squares and error bars show the median and IQR of the runner-up diabetic retinopathy class. Feature values are normalized for visualization. The plot provides dataset-level context for the top evidence features shown in the main manuscript, indicating whether the selected feature values are closer to the predicted-class distribution, the runner-up-class distribution, or near the boundary of the class-reference ranges.

range. Overall, the plot shows that several top evidence features are broadly consistent with the predicted healthy class.

S12 Feature Correlation, Redundancy, and Dimensionality Reduction

Because the proposed representation extracts multiple descriptors from several concentric annular regions, some dependence among the resulting feature columns is expected. For example, the same vascular descriptor measured in neighboring annuli may exhibit similar variation, while descriptors belonging to the same feature family may capture partially related aspects of retinal structure or appearance. We therefore analyzed both pairwise feature correlation and PCA-based dimensionality reduction.

S12.1 Pairwise feature correlation

Pairwise Spearman correlations were computed among all individual ring-wise feature columns within each dataset. Feature pairs were then summarized using four absolute-correlation intervals:

$$|\rho| < 0.50, \qquad 0.50 \le |\rho| < 0.70, \qquad 0.70 \le |\rho| < 0.90, \qquad |\rho| \ge 0.90.$$

As summarized in Table S8, most feature pairs exhibited weak absolute correlations in all three datasets. In HRF, 10,529 of 11,628 feature pairs (90.55%) had $|\rho| < 0.50$, while 653 pairs (5.62%), 342 pairs (2.94%), and 104 pairs (0.89%) fell within the 0.50–0.70, 0.70–0.90, and ≥ 0.90 intervals, respectively. In FIVES, the corresponding percentages were 88.24%, 6.65%, 4.61%, and 0.50%. In SUSTech-SYSU,

they were 88.68%, 6.38%, 3.38%, and 1.56%. Thus, approximately 88–91% of all feature pairs showed weak pairwise correlations, whereas only 0.5–1.6% exhibited very high correlations.

The strongest correlations were predominantly localized among related descriptors. Among feature pairs with $|\rho| \geq 0.90$, 95.19% in HRF, 80.00% in FIVES, and 78.87% in SUSTech-SYSU belonged to the same feature family. In addition, 86.54%, 70.00%, and 63.38% of these pairs, respectively, represented the same base descriptor measured in different annular regions. These findings indicate that the strongest correlations mainly reflect expected relationships among repeated or closely related measurements, particularly the same descriptor evaluated across neighboring annuli, rather than widespread duplication across unrelated feature categories.

S12.2 PCA-based dimensionality reduction

We further evaluated PCA to determine whether the multivariate ring-wise feature space could be compressed without reducing classification performance. Unlike pairwise correlation analysis, which evaluates associations between individual feature pairs, PCA assesses whether the joint feature variance can be represented by a smaller set of orthogonal components. PCA was fitted independently within each training fold to prevent information leakage, and components retaining 90%, 95%, and 99% of the training-set variance were evaluated.

As summarized in Supplementary Table S8, PCA substantially reduced dimensionality in all three datasets. At the 95% variance threshold, the original feature spaces were reduced from 153 to 26 components for HRF, from 64 to 28 components for FIVES, and from 96 to 40 components for SUSTech-SYSU, corresponding to reductions of 83.0%, 56.3%, and 58.3%, respectively. Similar trends were observed at the 90% and 99% thresholds.

Despite this compression, PCA did not improve classification performance. On HRF, the accuracies for PCA90, PCA95, and PCA99 were 0.822, 0.867, and 0.889, respectively, compared with 0.911 using the original features. The corresponding accuracies were 0.716 for all three PCA settings in FIVES, compared with 0.724, and 0.879, 0.883, and 0.882 in SUSTech-SYSU, compared with 0.895.

These results indicate that the feature space is compressible in a multivariate sense, but PCA provides no predictive advantage. Because principal components also combine descriptors from multiple feature categories and annular regions, we retained the original ring-wise representation to preserve descriptor-level and spatial interpretability.

Supplementary Fig. S22 shows the Spearman correlation heatmaps for the three datasets. Features are ordered by annular region from the innermost to the outermost retained ring, with the same set of descriptors repeated within each annulus. Under this ordering, diagonal blocks represent within-ring correlations, whereas off-diagonal blocks represent correlations between features measured in different rings.

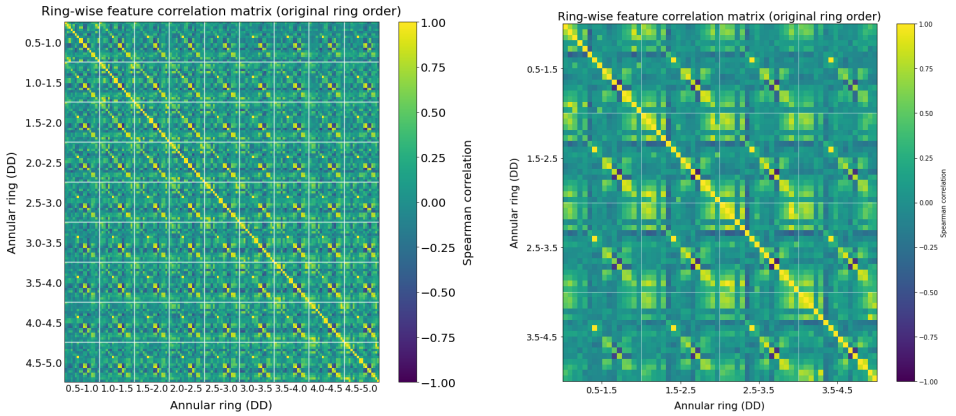
The heatmaps show that stronger correlations are confined to localized regions rather than being uniformly distributed across the complete feature space. Repeated patterns in the off-diagonal blocks arise because the same descriptors are evaluated across multiple annuli. Stronger local correlations are therefore primarily associated

Supplementary Table S8 Feature-correlation distribution and PCA dimensionality-reduction sensitivity analysis. Correlation entries report the percentage of feature pairs within each absolute Spearman-correlation interval. PCA columns report retained dimensionality and classification accuracy.

Dataset	N	$\lvert\rho\rvert$ < 0.50	$0.50 \leq \lvert\rho\rvert$ < 0.70	$0.70 \leq \lvert\rho\rvert$ < 0.90	$\lvert\rho\rvert$ ≥ 0.90	**Orig. Dim./Acc.**	**P90 Dim./Acc.**	**P95 Dim./Acc.**	**P99 Dim./Acc.**
HRF	45	90.55%	5.62%	2.94%	0.89%	153/**0.911**	19–20/0.822	26/0.867	36–37/0.889
FIVES	381	88.24%	6.65%	4.61%	0.50%	64/**0.724**	20–21/0.716	28/0.716	42/0.716
SUSTech-SYSU	1016	88.68%	6.38%	3.38%	1.56%	96/**0.895**	29/0.879	40/0.883	60/0.882

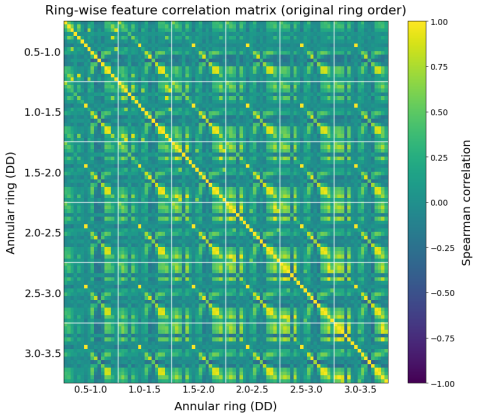
Note: N denotes the number of valid images. Orig. reports the dimensionality of the original per-image ring-wise feature vector and the corresponding classification accuracy. P90, P95, and P99 report the retained principal-component dimensionality, or its range across cross-validation folds, together with the corresponding classification accuracy after preserving 90%, 95%, and 99% of the training-set variance, respectively.

with related measurements across rings, whereas comparisons between heterogeneous feature families generally exhibit weak or mixed correlations. These families include vessel-density, caliber and diameter-ratio, tortuosity/curvature, color and SO_2-proxy, entropy, and graph-connectivity descriptors, which capture different aspects of retinal vascular structure and appearance. This visual pattern is consistent with the interval analysis in Table S8: approximately 88–91% of feature pairs have $|\rho| < 0.50$, while the relatively small number of stronger correlations is concentrated among related measurements.



(a) HRF.

(b) FIVES.



(c) SUSTech-SYSU.

Supplementary Fig. S22 Spearman correlation matrices of the ring-wise feature representations for HRF, FIVES, and SUSTech-SYSU. Features are shown in their original annular-ring order, with all selected descriptors grouped within each retained ring. White boundaries separate annular regions. Diagonal blocks describe within-ring correlations, whereas off-diagonal blocks describe correlations between different rings. Most feature pairs exhibit weak correlations, whereas stronger correlations are primarily localized among related descriptors and repeated measurements of the same descriptor across neighboring annuli.

S13 Feature-Family Importance Analysis

This section provides additional feature-family importance analyses for the proposed ring-based framework under different vessel-mask sources. RIP-AV is a pretrained retinal vessel segmentation model trained on the GAVE dataset [3]. We used the pretrained RIP-AV model to generate vessel masks for the HRF, FIVES, and SUSTech-SYSU datasets. For FR-UNet and SA-UNetv2, we trained the models using a collection of publicly available retinal vessel segmentation datasets, including AFIO [7], ARIA [8], CHASE_DB1 [9], DR HAGIS [10], DRIVE [11], LES-AV [12], STARE [13], and TREND [14]. The trained FR-UNet and SA-UNetv2 models were then applied to the same target datasets to generate vessel masks for downstream feature extraction and classification.

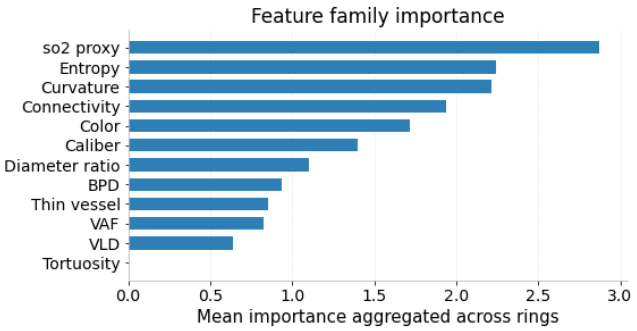
Supplementary Fig. S23, Fig. S24, and Fig. S25 show the feature-family importance rankings aggregated across all retained annular rings when RIP-AV, FR-UNet, and SA-UNetv2 vessel masks were used, respectively. The corresponding results obtained using ground-truth vessel masks are shown in Fig. S26.

Overall, these rankings are consistent with the trends discussed in the main text. Across the segmentation-based settings, appearance-related feature groups, including vessel color, SO_2 proxies, and vessel–background entropy, were generally more influential than geometric feature groups. This pattern was especially evident on FIVES and SUSTech-SYSU, where appearance-related descriptors consistently ranked among the most important feature families. In contrast, on the high-quality HRF dataset with ground-truth vessel masks, the top-ranked feature families included both geometric and appearance-related descriptors. This suggests that accurate vessel delineation can allow structurally meaningful vascular measurements to contribute more strongly to classification.

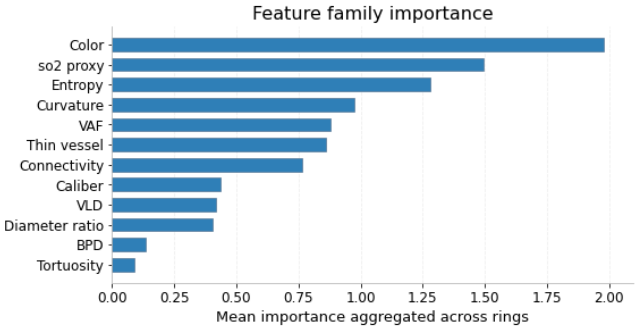
A further observation is that, on the lower-quality FIVES dataset, the dominant feature families remained predominantly appearance-related even when ground-truth vessel masks were used. This suggests that, under reduced image quality, appearance-related cues may remain more stable than fine-grained geometric measurements, even with expert vessel annotations. Taken together, these findings indicate that improved vessel masks alone do not fully remove the influence of image quality on the relative usefulness of different feature groups. They also support the importance of robust vessel segmentation, particularly for improving the reliability of geometry-based retinal vascular descriptors.

S14 Progressive Ring-Wise Evaluation

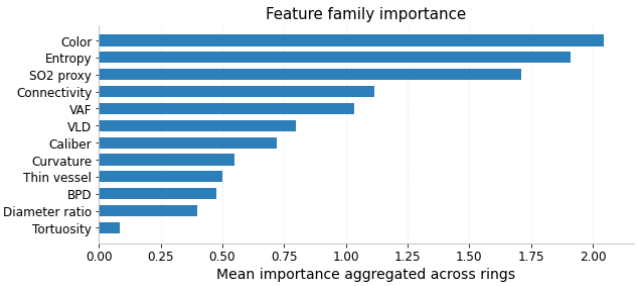
We performed a progressive ring-wise evaluation on the HRF dataset using ground truth vessel masks. Supplementary Fig. S27 shows the classification accuracy obtained by progressively including concentric annular regions from the optic disc outward. Overall, the combined feature set achieved the highest accuracy across most ring ranges, and its performance generally improved as more peripheral rings were included, with only minor fluctuations. Geometry-based features showed a similar trend, indicating that vascular structural information from more peripheral annuli contributed additional discriminatory value. In contrast, individual feature groups such as vessel



(a) HRF.



(b) FIVES.

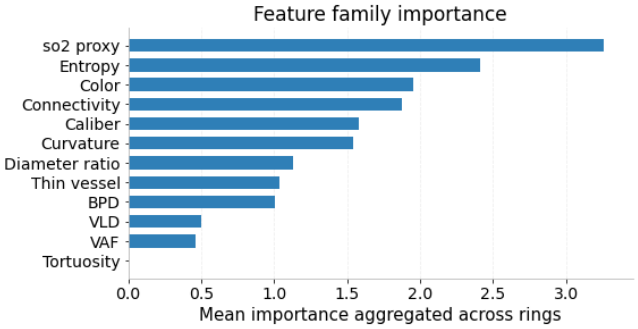


(c) SUSTech-SYSU.

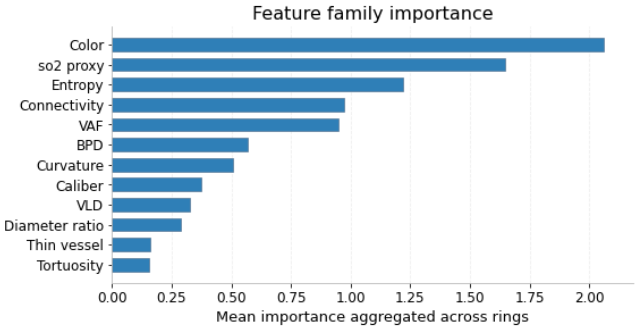
Supplementary Fig. S23 Feature-family importance aggregated across all retained annular rings using RIP-AV vessel segmentation on the (a) HRF, (b) FIVES, and (c) SUSTech-SYSU datasets. Larger values indicate greater overall contribution of each feature family to classification.

color, SO_2-proxy, and vessel–background entropy showed more modest changes across ring ranges.

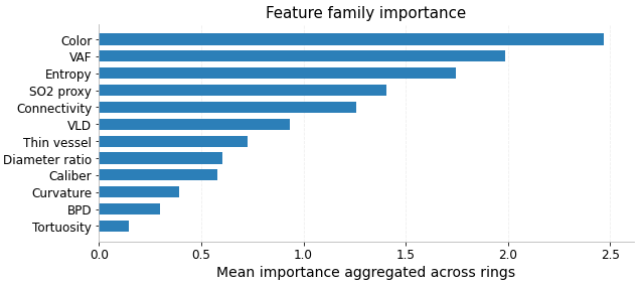
Although retinal vascular analysis often focuses on the peripapillary region for measurement consistency [15, 16], these results suggest that more peripheral rings also provide complementary discriminatory information. In particular, the best performance was achieved when a larger retinal extent was retained, especially for the combined feature set and geometry-based features. This indicates that peripheral



(a) HRF.



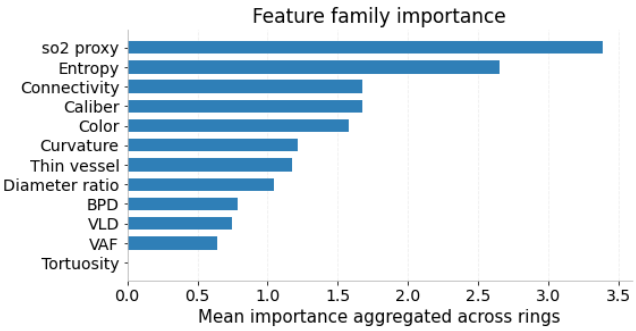
(b) FIVES.



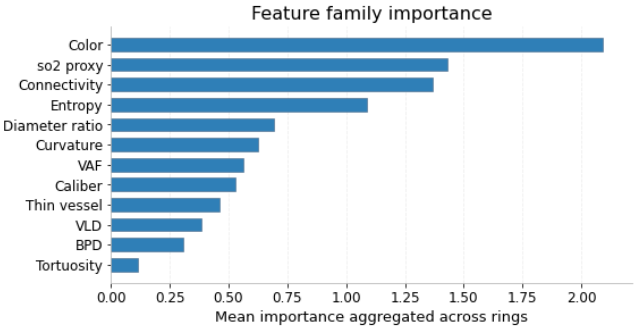
(c) SUSTech-SYSU.

Supplementary Fig. S24 Feature-family importance aggregated across all retained annular rings using FR-UNet vessel segmentation on the (a) HRF, (b) FIVES, and (c) SUSTech-SYSU datasets. Larger values indicate greater overall contribution of each feature family to classification.

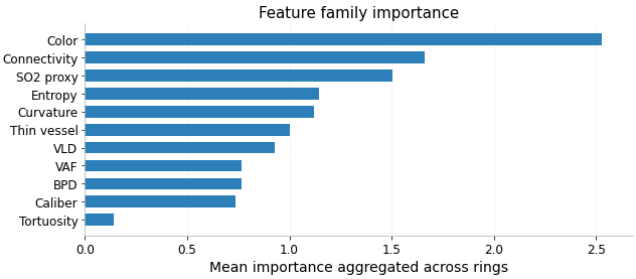
vascular structure provides complementary cues beyond the region immediately surrounding the optic disc. One possible explanation is that the HRF dataset provides high-resolution 45° fundus images with sufficient vessel detail across a relatively broad FOV, enabling meaningful feature extraction over multiple optic-disc diameters. In addition, DR often affects smaller vessels outside the immediate peripapillary region. Therefore, including outer rings may help capture disease-related abnormalities that are less apparent in central regions alone.



(a) HRF.



(b) FIVES.

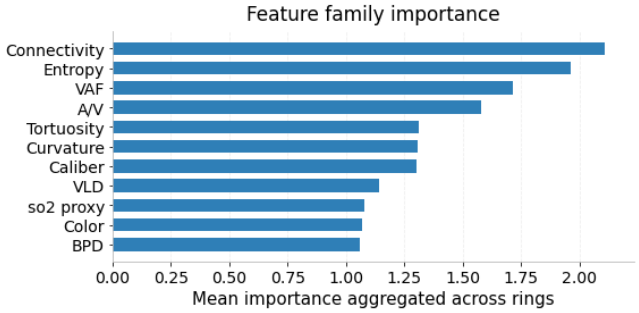


(c) SUSTech-SYSU.

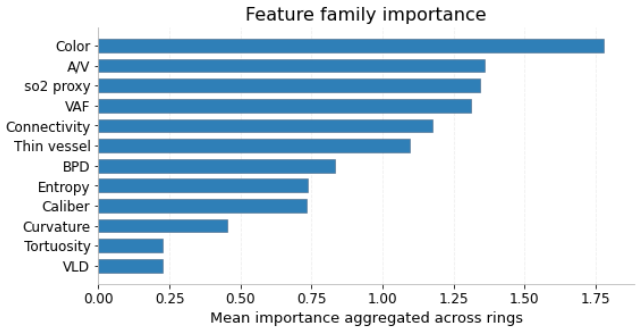
Supplementary Fig. S25 Feature-family importance aggregated across all retained annular rings using SA-UNetv2 vessel segmentation on the (a) HRF, (b) FIVES, and (c) SUSTech-SYSU datasets. Larger values indicate greater overall contribution of each feature family to classification.

S15 Effect of the Optic Disc on Classification Performance

Anatomically, the optic disc is a distinctive fundus region where retinal ganglion cell axons converge and exit the eye. It contains a bright central optic cup, has a relatively consistent size across individuals (average major axis $\approx 1.8\,\text{mm}$), and maintains stable spatial relationships with nearby landmarks such as the fovea [17]. These properties



(a) HRF.



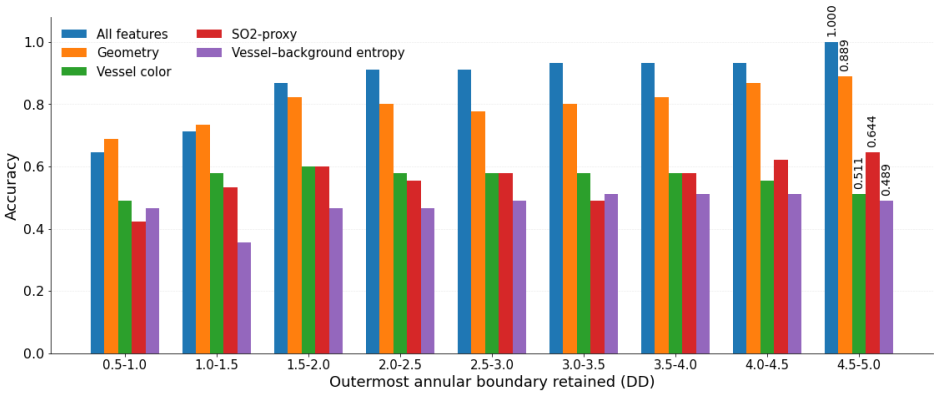
(b) FIVES.

Supplementary Fig. S26 Feature-family importance aggregated across all retained annular rings using ground-truth vessel masks on the (a) HRF and (b) FIVES datasets. Larger values indicate greater overall contribution of each feature family to classification.

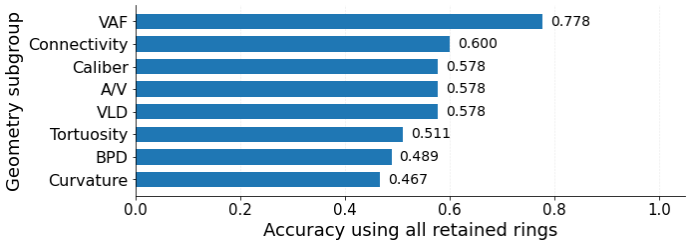
make it a clinically meaningful reference region for capturing disease-related structural and vascular changes.

To assess the contribution of optic disc information to classification performance, we conduct experiments on the FIVES dataset using both the proposed method and the pretrained RETFound model. For RETFound, we evaluate two image settings: (i) the full fundus image and (ii) the same image with the optic disc region masked (Supplementary Fig. S28). For the proposed method, we additionally extract optic disc-specific geometric and texture descriptors, including: (i) optic disc vessel area fraction (OD-VAF), defined as the ratio of vessel pixels within the disc region; (ii) disc vessel sparsity, computed as $1 - \text{OD-VAF}$; (iii) a disc-normalized caliber index, defined as the mean vessel diameter divided by the optic disc diameter; and (iv) optic disc entropy, quantifying local texture complexity within the disc region.

These descriptors are clinically motivated. In glaucoma, neuroretinal rim thinning and optic disc cupping can reduce the density and continuity of visible peripapillary vessels, potentially leading to lower OD-VAF and reduced local texture complexity. In diabetic retinopathy, widespread vascular dysfunction may also affect the disc and peripapillary region; in proliferative disease, neovascularization of the disc (NVD) may increase apparent vessel occupancy and local texture variability near the optic disc [18].

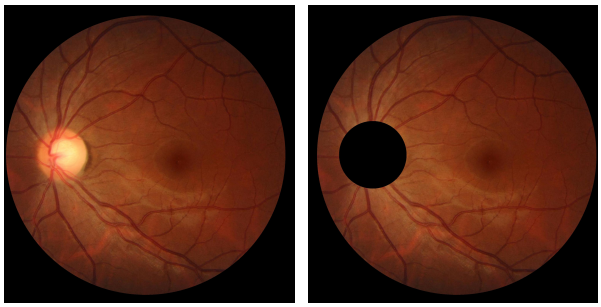


(a) Accuracy as progressively larger annular regions are included.



(b) All-ring accuracy using individual geometry subgroups.

Supplementary Fig. S27 Ring-wise classification accuracy on the HRF dataset using ground-truth vessel masks. (a) Accuracy is shown as progressively larger concentric annular regions are included from the optic disc outward. Each bar group corresponds to one feature category: all features, vessel geometry, vessel color, SO_2-proxy features, and vessel–background entropy. Accuracy values are annotated for the final retained annular extent. (b) Standalone all-ring classification accuracy using individual geometry subgroups. Overall, larger retinal extents generally improve performance, with the full feature set achieving the highest accuracy across most ring ranges.



Supplementary Fig. S28 Example DR fundus image. Left: original full image; right: same image with the optic disc region removed. After removal of optic disc information, the pretrained RETFound model misclassifies the image as healthy.

To capture vascular context beyond the disc interior, we expand the optic disc region to 1.3× DD, thereby including the immediate peripapillary zone while limiting the inclusion of more distant background tissue. Supplementary Fig. S28 shows an example DR case in which RETFound misclassifies the image as healthy after the optic disc region is masked, highlighting the contribution of disc-related cues. Supplementary Table S9 summarizes results with and without optic disc information for both methods on FIVES; incorporating optic disc information improves accuracy in both settings. Moreover, augmenting the proposed ring-based representation with optic disc-specific descriptors using ground-truth vessel masks yields performance comparable to that of RETFound. Supplementary Fig. S29 further shows representative DR cases that are correctly classified when optic disc and vessel features are combined but misclassified when only vessel features are used. Overall, these findings suggest that optic disc-centered vascular and appearance cues provide complementary information for fundus image classification. Because the optic disc is the convergence point of the major retinal vessels, disc-centered descriptors may reflect disease-related vascular remodeling.

Supplementary Table S9 Comparison of classification performance on FIVES dataset with and without optic disc information. Results for the proposed method are based on ground-truth vessel masks.

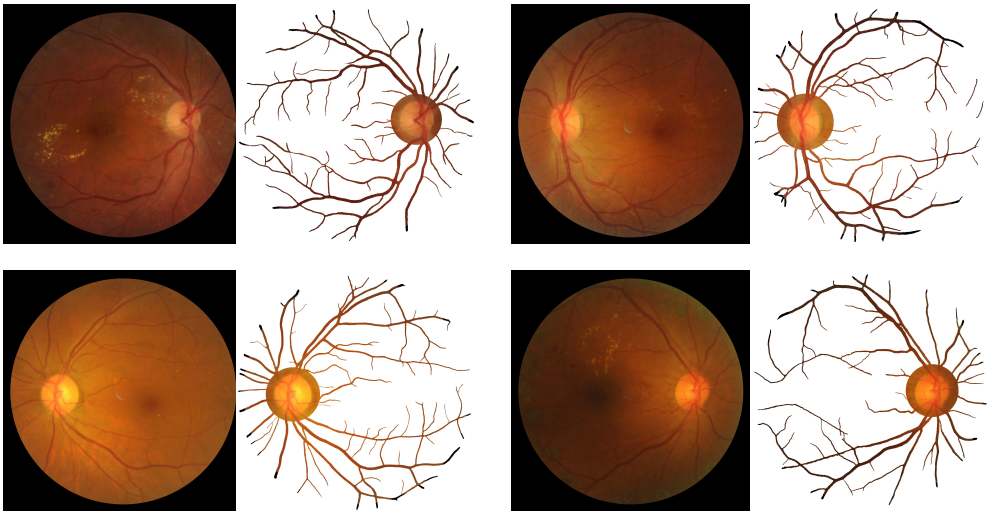
Method	Accuracy	95% CI
RETFound (full image)	0.806	[0.763, 0.842]
RETFound (image without disc)	0.761	[0.716, 0.801]
Proposed (GT vessel + disc)	0.790	[0.746, 0.828]
Proposed (GT vessel only)	0.764	[0.719, 0.804]

S16 Related Work

Retinal vascular analysis based on fundus images provides an effective approach to characterizing microvascular structure and function in ophthalmic diseases such as DR and glaucoma. A variety of quantitative vascular descriptors have been proposed to capture complementary aspects of retinal physiology, including SO_2-proxy oxygenation indices, vessel caliber, vascular density, and other geometric and textural descriptors. To contextualize the vascular descriptors employed in this work, this section first reviews representative studies across these feature categories, and then discusses a pretrained ViT-based foundation model for fundus image–based prediction.

S16.1 Oxygen Saturation and SO_2-Proxy Features

Retinal oximetry has been extensively investigated as a noninvasive method for characterizing retinal oxygen metabolism. Early systems employed fundus cameras equipped with dual-wavelength filters and color charge-coupled device (CCD) sensors to acquire oxygen-sensitive and oxygen-insensitive images, enabling estimation of retinal vessel



Supplementary Fig. S29 Representative DR fundus images correctly classified using vessel + optic disc features but misclassified using vessel-only features. Left: original fundus image; right: vessel structure with optic disc region, illustrating the complementary role of optic disc information.

oxygen saturation based on differences in optical density [1, 2, 19]. These approaches are grounded in the Beer-Lambert law, in which the optical density ratio (ODR) between selected wavelengths serves as a surrogate for the relative oxygenation of hemoglobin.

More recently, efforts have been made to extract oxygenation-related contrast from conventional RGB fundus photographs. For example, [20] reported that standard RGB images can approximate dual-wavelength oximetry by computing optical density along vessel centerlines, using the red channel as a relatively oxygen-sensitive band and the green channel as a comparatively oxygen-insensitive reference. This observation enables high-throughput extraction of relative, oxygenation-related appearance cues from standard RGB fundus images in widely available datasets and motivates the construction of oxygenation-related SO_2-proxy descriptors in this work.

Population-level retinal oximetry studies have reported systematic differences in vessel oxygen saturation across cohorts with DR, glaucoma, and other ocular diseases [21–25]. These findings identify oxygenation-related contrast as a relevant physiological signal in retinal imaging research and motivate the exploration of oxygenation-derived features beyond traditional oximetry settings. In this context, our work leverages RGB-derived SO_2-like descriptors as relative appearance cues for disease classification, acknowledging that such features are not calibrated physiological measurements but can still capture meaningful inter-cohort differences.

S16.2 Vessel Diameters and Diameter-Based Metrics.

Retinal diameters, known as retinal vessel caliber, and related diameter-based metrics, such as the arteriolar-to-venular diameter ratio, are well-established quantitative markers for retinal disease assessment [26–34]. Numerous population-based and clinical

studies have demonstrated disease-specific alterations in retinal vessel diameters. DR is commonly associated with venular dilation and heterogeneous arteriolar changes, reflecting metabolic and inflammatory microvascular damage, whereas glaucoma is more consistently characterized by progressive arteriolar narrowing related to chronic vascular dysregulation [26].

[35] reported enlarged retinal venules in individuals with DR in an Asian population, corroborating earlier findings in white populations and supporting the use of retinal caliber measurements as indicators of early diabetic microvascular injury. In the context of glaucoma, [36] demonstrated that narrower retinal arteriolar caliber is significantly associated with glaucomatous optic neuropathy. Differences in vessel diameters between eyes have also been linked to the risk of developing retinal vein occlusion [31]. Together, these findings highlight the clinical relevance of quantitative vessel diameter analysis for disease characterization.

S16.3 Vessel Density and Length Density.

Beyond vessel caliber, vessel area density and vessel length density are widely used descriptors of retinal microvascular integrity, particularly in studies based on optical coherence tomography angiography (OCTA). In DR, OCTA investigations have shown that retinal vessel density decreases with increasing disease severity and can reveal subtle microvascular abnormalities at early stages [32]. In glaucoma, OCTA-based studies have consistently reported reduced retinal vessel density relative to healthy controls, most prominently in the optic nerve head and peripapillary regions, with additional decreases also observed in the macular vasculature [37]. Moreover, [38] reported that faster OCTA vessel density loss is associated with faster concurrent and subsequent visual field deterioration, supporting OCTA metrics as complementary markers for disease monitoring.

In the context of color fundus imaging, analogous information can be approximated using vessel area fraction and vessel length density computed from binary vessel masks or skeletons, providing complementary structural cues without requiring OCTA acquisitions.

S16.4 Tortuosity.

Retinal vessel tortuosity and curvature are widely used quantitative descriptors of vascular morphology and branching organization [29, 30, 34, 39, 40]. Automated and semi-automated algorithms for estimating tortuosity and curvature from color fundus images have enabled large-scale analyses of vascular geometry in both population-based cohorts and disease-specific studies. Alterations in retinal vascular geometry have been linked to ocular diseases, supporting the hypothesis that microvascular structural remodeling contributes to disease pathogenesis.

In glaucoma, population-based investigations have reported significant associations between retinal vascular geometric parameters, including tortuosity and branching complexity, and glaucomatous optic neuropathy, independent of intraocular pressure [41]. In DR, longitudinal studies have also reported that retinal vascular tortuosity

is associated with disease progression and regression, suggesting that tortuosity-related alterations may reflect evolving microvascular change over time [42].

S16.5 Branch-Point Density.

Fundus vessel topology can be quantified using junction metrics such as branch-point (bifurcation) density, defined as the number of skeleton branch points per unit area (or normalized by total skeleton length). Junction/crossover density can be reported similarly when crossings are considered. Changes in these densities can reflect altered network complexity due to vascular rarefaction/pruning or pathological remodeling [29, 43].

S16.6 RETFound and Retinal Foundation Models.

Recent advances in machine learning for retinal imaging have shifted from task-specific supervised learning toward foundation models that learn transferable representations from large-scale unlabeled data. In this context, RETFound, introduced by [4], is a vision transformer–based retinal foundation model pretrained with a masked autoencoder (MAE) objective on 1.6 million unlabeled fundus images. This pretraining strategy enables RETFound to capture anatomical and vascular structures visible in fundus photographs and to adapt efficiently to diverse downstream tasks with limited labeled data. RETFound has demonstrated improved performance and label efficiency over ImageNet-pretrained and task-specific baselines across multiple ophthalmic diseases, including DR, glaucoma, and age-related macular degeneration.

S17 Training and Inference Runtime Comparison

To assess computational efficiency, Supplementary Table S10 summarizes the training and inference runtimes of the proposed ring-based pipelines and the frozen and fine-tuned RETFound [4] baselines on the high-resolution HRF dataset. Vessel segmentation, RETFound feature extraction, and RETFound fine-tuning were performed on an NVIDIA Quadro RTX 6000 GPU with 24 GB of memory. Ring-wise feature extraction and logistic-regression training and prediction were performed on an Intel Xeon W-2255 CPU with 10 physical cores and 20 logical threads. For frozen RETFound, inference included the GPU-based backbone forward pass, transfer of the extracted embedding to the CPU, feature scaling, and logistic-regression prediction. Training time is reported as the mean per cross-validation fold, whereas inference time is reported per image.

RETFound achieved substantially shorter per-image inference times, with its image encoder run on the GPU. In the fine-tuned setting, the complete encoder and integrated neural classification head were evaluated on the GPU as part of the same forward pass, while only the last two encoder blocks and associated trainable encoder parameters were updated during training. In the frozen setting, the extracted embeddings were transferred to the CPU for feature scaling and logistic-regression prediction. Fine-tuning the last two RETFound encoder blocks required approximately 169.36 s

Supplementary Table S10 Mean training and inference runtimes on the high-resolution HRF dataset. Training time is reported per cross-validation fold, whereas inference time is reported per image. The processing device used at each stage is indicated in parentheses.

Method	Seg. (s/img)	Feat. (s/img)	Infer./Pred. (s/img)	Total (s/img)	Train/fold (s)
RIP-AV + Ring	6.650 (G)	44.270 (C)	0.0003 (C)	50.920	0.060 (C)
SA-UNetV2 + Ring	0.027 (G)	44.270 (C)	0.0003 (C)	44.297	0.060 (C)
FR-UNet + Ring	0.223 (G)	44.270 (C)	0.0003 (C)	44.493	0.060 (C)
Frozen RETFound	–	–	0.021 (G+C)	0.021	2.240 (C)
Fine-tuned RETFound	–	–	0.017 (G)	0.017	169.356 (G)

Note: Seg.: vessel segmentation; Feat.: ring-wise feature extraction; Infer./Pred.: model inference or classification prediction; img: image; G: GPU; C: CPU.

per cross-validation fold, compared with 2.24 s for training the logistic-regression classifier in the frozen RETFound setting and 0.06 s for the logistic-regression classifier used with the proposed representation. The proposed pipeline was slower during inference, primarily because the current ring-wise feature-extraction implementation was executed on the CPU. Nevertheless, once a vessel mask was available, the ring-based representation and logistic-regression classifier required no additional GPU computation. This design provides a CPU-compatible option for downstream analysis in settings with limited GPU resources, although ring-wise feature extraction remains the principal computational bottleneck in the current implementation.

References

[1] Beach, J. M., Schwenzer, K. J., Srinivas, S., Kim, D. & Tiedeman, J. S. Oximetry of retinal vessels by dual-wavelength imaging: Calibration and influence of pigmentation. *J. Appl. Physiol.* **86**, 748–758 (1999).

[2] Hammer, M., Vilser, W., Riemer, T. & Schweitzer, D. Retinal vessel oximetry-calibration, compensation for vessel diameter and fundus pigmentation, and reproducibility. *J. Biomed. Opt.* **13**, 054015 (2008).

[3] GAVE. RIP-AV GAVE dataset and related resources (2026). URL https://drive.google.com/drive/folders/1pveLOncip6ioNsRx9O2rR7YfgztjyXyg. Accessed: Apr. 3, 2026.

[4] Zhou, Y. *et al.* A foundation model for generalizable disease detection from retinal images. *Nature* **622**, 156–163 (2023).

[5] Pla, C. Fundus lesions toolkit. https://github.com/ClementPla/fundus-lesions-toolkit (2023). Accessed: 2026-07-02.

[6] Suvorov, R. *et al.* Resolution-robust large mask inpainting with fourier convolutions. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2149–2159 (2022).

[7] Akram, M. U., Akbar, S., Hassan, T. & Basit, I. Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema. *Data Brief* **29**, 105282 (2020).

[8] El-Bendary, N., Hassanien, A. E., Corchado, E. & Berwick, R. C. ARIAS: Automated retinal image analysis system. *Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011*, 67–76 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011).

[9] Barman, S. *et al.* CHASE_DB1 retinal vessel reference dataset (2012).

[10] Holm, S., Russell, G., Nourrit, V. & McLoughlin, N. P. DR HAGIS—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. *J. Med. Imaging* **4**, 014503 (2017).

[11] Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A. & van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **23**, 501–509 (2004).

[12] Orlando, J. I. *et al.* Towards a glaucoma risk index based on simulated hemodynamics from fundus images. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 65–73 (2018).

[13] Hoover, A., Kouznetsova, V. & Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **19**, 203–210 (2000).

[14] Popovic, N., Vujosevic, S., Radunovic, M., Radunovic, M. & Popovic, T. TREND database: Retinal images of healthy young subjects visualized by a portable digital non-mydriatic fundus camera. *PLoS ONE* **16**, e0254918 (2021).

[15] Hubbard, L. D. *et al.* Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the Atherosclerosis Risk in Communities Study. *Ophthalmology* **106**, 2269–2280 (1999).

[16] Abràmoff, M. D., Garvin, M. K. & Sonka, M. Retinal imaging and image analysis. *IEEE Rev. Biomed. Eng.* **3**, 169–208 (2010).

[17] Quigley, H. A., Brown, A. E., Morrison, J. D. & Drance, S. M. The size and shape of the optic disc in normal human eyes. *Arch. Ophthalmol.* **108**, 51–57 (1990).

[18] Cheung, N., Mitchell, P. & Wong, T. Y. Diabetic retinopathy. *Lancet* **376**, 124–136 (2010).

[19] Hirsch, K., Cubbidge, R. P. & Heitmar, R. Dual wavelength retinal vessel oximetry – influence of fundus pigmentation. *Eye* **37**, 2246–2251 (2023).

[20] Zhang, J. *et al.* Oxygen-saturation-related functional parameter as a biomarker for diabetes mellitus—extraction method and clinical validation. *Front. Cell Dev. Biol.* **11**, 1195873 (2023).

[21] Blair, N. P. *et al.* Retinal oximetry and vessel diameter measurements with a commercially available scanning laser ophthalmoscope in diabetic retinopathy. *Invest. Ophthalmol. Vis. Sci.* **58**, 5556–5563 (2017).

[22] Olafsdottir, O. B., Hardarson, S. H., Gottfredsdottir, M. S., Harris, A. & Stefánsson, E. Retinal oximetry in primary open-angle glaucoma. *Invest. Ophthalmol. Vis. Sci.* **52**, 6409–6413 (2011).

[23] Shimazaki, T. *et al.* Relationship between oxygen saturation of the retinal vessels and visual field defect in glaucoma patients: comparison with each hemifield. *Acta Ophthalmol.* **94**, e683–e687 (2016). URL https://doi.org/10.1111/aos.13089.

[24] Mahmoudinezhad, G. *et al.* Relationship between retinal oxygen saturation and the severity of visual field damage in glaucoma. *J. Glaucoma* **33**, 728–734 (2024).

[25] Zhang, W. *et al.* Retinal oximetry: new insights into ocular and systemic diseases. *Graefes Arch. Clin. Exp. Ophthalmol.* **263**, 2101–2115 (2025). URL https://doi.org/10.1007/s00417-025-06831-8. Epub 2025 Apr 21.

[26] Sun, C., Wang, J. J., Mackey, D. A. & Wong, T. Y. Retinal vascular caliber: systemic, environmental, and genetic associations. *Surv. Ophthalmol.* **54**, 74–95 (2009).

[27] Niemeijer, M. *et al.* Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs. *IEEE Trans. Med. Imaging* **30**, 1941–1950 (2011).

[28] Habib, M. S., Al-Diri, B., Hunter, A. & Steel, D. H. W. The association between retinal vascular geometry changes and diabetic retinopathy and their role in prediction of progression – an exploratory study. *BMC Ophthalmol.* **14**, 89 (2014).

[29] Dashtbozorg, B. *et al.* Infrastructure for retinal image analysis. *Proceedings of the Ophthalmic Medical Image Analysis International Workshop (OMIA)*, Vol. 3, 105–112 (2016).

[30] Asl, M. E., Koohbanani, N. A., Frangi, A. F. & Gooya, A. Tracking and diameter estimation of retinal vessels using Gaussian process and Radon transform. *J. Med. Imaging* **4**, 034006 (2017).

[31] Kawasaki, R. & Grauslund, J. Clinical motivation and the needs for RIA in healthcare in *Computational Retinal Image Analysis: Tools, Applications and Perspectives* (Trucco, E., MacGillivray, T. & Xu, Y., eds) Ch. 2, 5–17 (Academic Press, 2019).

[32] Li, X. *et al.* Quantitative analysis of retinal vessel density and thickness changes in diabetes mellitus evaluated using optical coherence tomography angiography: a cross-sectional study. *BMC Ophthalmol.* **21**, 259 (2021).

[33] Ahmad, N., Lai, K.-T. & Tanveer, M. Retinal blood vessel tracking and diameter estimation via Gaussian process with rider optimization algorithm. *IEEE J. Biomed. Health Inform.* **28**, 1173–1184 (2024).

[34] Andreini, P. & Bonechi, S. Towards a comprehensive characterization of arteries and veins in retinal imaging. *Comput. Biol. Med.* **195**, 110516 (2025).

[35] Islam, F. M. A. *et al.* Quantitative retinal vascular caliber changes in diabetes and retinopathy: the Singapore Malay Eye Study. *Eye* **23**, 1719–1724 (2009).

[36] Amerasinghe, N. *et al.* Evidence of retinal vascular narrowing in glaucomatous eyes in an asian population. *Invest. Ophthalmol. Vis. Sci.* **49**, 5397–5402 (2008).

[37] Khayrallah, O. *et al.* Optical coherence tomography angiography vessel density parameters in primary open-angle glaucoma. *Ann. Med. Surg. (Lond.)* **69**, 102671 (2021).

[38] Nishida, T., Moghimi, S., Wu, J.-H. *et al.* Association of initial optical coherence tomography angiography vessel density loss with faster visual field loss in glaucoma. *JAMA Ophthalmol.* **140**, 319–326 (2022).

[39] Owen, C. G. *et al.* Measuring retinal vessel tortuosity in 10-year-old children: validation of the Computer-Assisted Image Analysis of the Retina (CAIAR) program. *Invest. Ophthalmol. Vis. Sci.* **50**, 2004–2010 (2009).

[40] Huang, F. *et al.* Vascular biomarkers for diabetes and diabetic retinopathy screening in *Computational Retinal Image Analysis: Tools, Applications and Perspectives* (Trucco, E., MacGillivray, T. & Xu, Y., eds) Ch. 16, 319–352 (Academic Press, 2019).

[41] Wu, R. *et al.* Retinal vascular geometry and glaucoma: The Singapore Malay Eye Study. *Ophthalmology* **120**, 77–83 (2013).

[42] Lim, L. S., Chee, M. L., Cheung, C. Y. & Wong, T. Y. Retinal vessel geometry and the incidence and progression of diabetic retinopathy. *Invest. Ophthalmol. Vis. Sci.* **58**, BIO200–BIO205 (2017).

[43] Abbasi-Sureshjani, S., Smit-Ockeloen, I. M., Bekkers, E. J., Dashtbozorg, B. & ter Haar Romeny, B. M. Automatic detection of vascular bifurcations and crossings in retinal images using orientation scores. *Proceedings of the 13th IEEE International Symposium on Biomedical Imaging (ISBI 2016)*, 189–192 (IEEE, 2016).