Synthetic Vasculature and Pathology Enhance Vision-Language Model Reasoning

Chenjun Li[1,2] CL2733@CORNELL.EDU
Cheng Wan[1,2] CW2222@CORNELL.EDU
Laurin Lux[2,3] LAURIN.LUX@TUM.DE
Alexander H. Berger[2,3] A.BERGER@TUM.DE
Richard B. Rosen[4] RROSEN@NYEE.EDU
Martin J. Menten[3] MARTIN.MENTEN@TUM.DE
Johannes C. Paetzold[2,5] JPAETZOLD@MED.CORNELL.EDU
[1] *School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA*
[2] *Weill Cornell Medicine, New York, NY 10021, USA*
[3] *School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany*
[4] *New York Eye and Ear Infirmary of Mount Sinai, New York, NY*
[5] *Cornell Tech, New York, NY 10044, USA*

Editors: Under Review for MIDL 2026

Abstract

Vision-Language Models (VLMs) offer a promising path toward interpretable medical diagnosis by allowing users to ask about clinical explanations alongside predictions and across different modalities. However, training VLMs for detailed reasoning requires large-scale image-text datasets. In many specialized domains, for example in reading Optical Coherence Tomography Angiography (OCTA) images, such precise text with grounded description of pathologies is scarce or even non-existent. To overcome this bottleneck, we introduce Synthetic Vasculature Reasoning (SVR), a framework that controllably synthesizes images and corresponding text, specifically: realistic retinal vasculature with Diabetic Retinopathy (DR) features: capillary dropout, microaneurysms, neovascularization, and tortuosity, while automatically generating granular reasoning texts. Based on this we curate OCTA-100K-SVR, an OCTA image-reasoning dataset with 100,000 pairs. Our experiments show that a general-purpose VLM (Qwen3-VL-8b) trained on the dataset achieves a zero-shot balanced classification accuracy of 89.67% on real OCTA images, outperforming supervised baselines. Through human expert evaluation we also demonstrate that it significantly enhances explanation quality and pathology localization on clinical data.

Keywords: VLM, CoT, OCTA, DR, Synthetic Pathology

1. Introduction

Vision-language models (VLMs) integrate visual processing with natural language reasoning. Many studies discuss how such models can be particularly valuable in medical image analysis, as they can move beyond simple classification to support interpretable diagnosis (Li et al., 2023; Zhang et al., 2024; Yang et al., 2025; Sellergren et al., 2025). Compared to traditional black-box classifiers, VLMs could explicitly describe *what* features are present, *where* pathologies are located, and *why* a specific diagnosis is suggested. This explainability is critical for clinical validation.

Chain-of-thought (CoT) reasoning enables models to generate step-by-step explanations that simulate human workflows (Wei et al., 2022). In medical imaging, this approach breaks down diagnosis into interpretable stages: identifying anatomical structures, detecting and localizing abnormalities, and synthesizing these findings into a conclusion. However, training VLMs to produce clinically accurate CoT requires extensive datasets with annotations that exceed standard image-level labels (Pan et al., 2025; Lai et al., 2025). Privacy regulations, significant acquisition costs, and the requirement for specialized expertise restrict data availability, particularly for modalities like Optical Coherence Tomography Angiography (OCTA), where even the largest public datasets that include OCTA typically involve only hundreds to a few thousand subjects (Li et al., 2024; AI-READI Consortium, 2024). While a recent approach has attempted to improve VLM interpretability on OCTA images via graph-based knowledge extraction (Li et al., 2025), it is limited to small sample sizes and lacks detailed localization of different pathological features. Without ground-truth annotations, it is difficult to train VLMs that provide high-quality, location-specific explanations.

Synthetic data generation offers a practical solution for data augmentation in VLM training (Ma et al., 2025; Wu et al., 2025). Synthetic generation provides control over visual features and automatically produces ground-truth annotations. Despite these benefits, synthetic data is underutilized in medical VLM training. Previous work by Kreitner et al. (2024) validated synthetic OCTA for vessel segmentation but was limited to healthy vasculature, without pathological features and text-based reasoning chains. A framework capable of generating both realistic pathology and corresponding grounded explanations is needed to enable VLMs to diagnose, understand and reason about OCTA images.

We investigate this approach for Diabetic Retinopathy (DR) staging using OCTA images. DR is a primary cause of vision loss (Lee et al., 2015), and accurate diagnosis of DR relies on the detection and localization of microvascular abnormalities (Alam et al., 2020; Sun et al., 2021). Current public datasets (Dai et al., 2021; Li et al., 2024; AI-READI Consortium, 2024) generally provide only image-level labels.

In this work, we introduce **Synthetic Vasculature Reasoning (SVR)**, a pathology-aware OCTA synthesis framework designed to improve VLM reasoning capabilities (Code and dataset available at: `https://github.com/d0ng231/OCTA-SVR`). Our work makes the following contributions:

- We developed the first module capable of simulating four distinct DR features on synthetic OCTA images: capillary dropout, microaneurysms, neovascularization, and vessel tortuosity.
- Based on the controllable features, we present **OCTA-100K-SVR**, a synthetic dataset with 100,000 pairs of high-quality OCTA images and texts for VLM training, and demonstrate that scaling synthetic training data improves VLM performance in both classification and reasoning.
- Through evaluations on proprietary and public datasets, we show that VLMs trained via SVR generate clinically accurate explanations and correct pathology localization, exceeding the performance of other models **without fine-tuning on real data**.

2. Method

2.1. Overview

In Fig. 1 we illustrate the core of our method, which is based on a simulation module that generates topologically accurate vascular graphs with controlled diabetic retinopathy features. A generative adversarial network (GAN) then converts vessel maps into realistic OCTA images, while a teacher VLM converts the corresponding pathology metadata into grounded Chain-of-Thought reasoning texts. By aligning synthetic OCTA images with granular reasoning across 100,000 samples, the framework enables the VLM to learn robust diagnostic representations prior to fine-tuning on limited real-world clinical data.

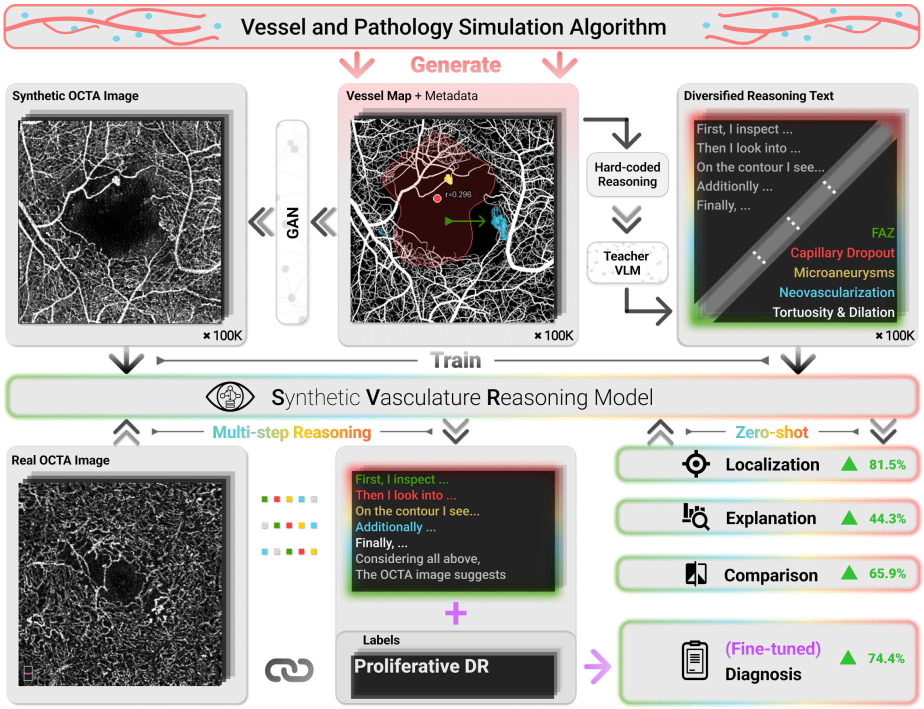


Figure 1: **Overview of the Synthetic Vasculature Reasoning Framework.** The pipeline begins with the simulation of vessels and pathology. In this way we generate ground-truth vessel maps with and without various DR hallmarks (detailed in Sec. 2.2 and Fig. 2). These maps are then used for: (1) visual synthesis, where a pretrained GAN from Kreitner et al. (2024) converts maps into realistic synthetic OCTA images, and (2) text generation, where structured pathology metadata is converted into diversified Chain-of-Thought reasoning text using GPT-5 (OpenAI, 2025) (see Pseudo code 1 for more details). On these image-text pairs we can fine-tune VLMs which are capable of multiple tasks including disease staging, locating pathological features, and explaining their reasoning. The marked improvements in localization, explanation, comparison and diagnosis are based on expert evaluation (see Sec. 4).

2.2. Synthetic Vessel Growth and Pathology

Synthetic OCTA generation pipeline. We build on the statistical angiogenesis simulator of Kreitner et al. (2024), which itself adapts the space–colonization model of Rauch and Harders (2021). In brief, the retinal vasculature is represented as a forest of rooted 3D binary trees growing in a normalized box $\Omega = [0,1] \times [0,1] \times [0,h_z]$, where the lateral coordinates $(x,y) \in [0,1]^2$ correspond, after global scaling, to a $3 \times 3\,\text{mm}^2$ macular crop and h_z denotes the normalized slab thickness in the axial (z) direction. Arterial and venous trees are grown in two successive layers that together form the superficial and deep vascular complexes, driven by randomly sampled oxygen sinks and CO_2 sources. Growth follows local attraction cones and Murray's law–based bifurcation rules, producing capillary beds and major vessels. After growth, the vessel graph is voxelized into a 3D volume, projected to an en-face map, and passed through a GAN-based contrast adaptation module to produce realistic synthetic OCTA images.

To increase variability in viewpoint and morphology, we add a random shift and jitter of the foveal avascular zone (FAZ) within the normalized image plane before projection. All variables and parameters of the growth model are described in Appendix A.

Pathology-aware graph augmentation. Going beyond previous works which have been focused on healthy vessel graph generation (Menten et al., 2022; Kreitner et al., 2024; Wittmann et al., 2024; Prabhakar et al., 2024), we introduce a pathology module that operates directly on the 3D tree representation. For each synthetic sample we first use the base simulator to complete the growth of arterial and venous forests, and then apply a sequence of pathology-specific graph remodeling operations that are clinically motivated by the four defining pathologies of DR that show in OCTA images of the retinal vasculature (Kaizu et al., 2017; Alam et al., 2020; Sun et al., 2021) (See Fig. 2). Although the vasculature is simulated in 3D with nodes $\mathbf{x} = (x,y,z)$, all pathology fields are parameterized in the normalized en-face plane using the lateral coordinates $\tilde{\mathbf{x}} = (x,y) \in [0,1]^2$.

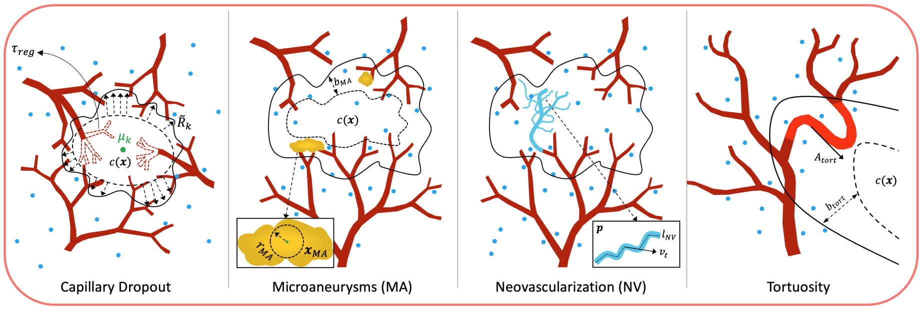


Figure 2: **Illustration of the synthesized DR pathologies.** DR cases show four main pathologies in OCTA images, which we implement in our vessel synthesis model: (1) capillary dropout, (2) microaneurysms (MA), (3) neovascularization (NV), and (4) increased tortuosity along dropout borders. Comprehensive mathematical description of the algorithms is provided in Appendix A. Afterwards, the generated (pathology-including) vessel maps are fed to the GAN.

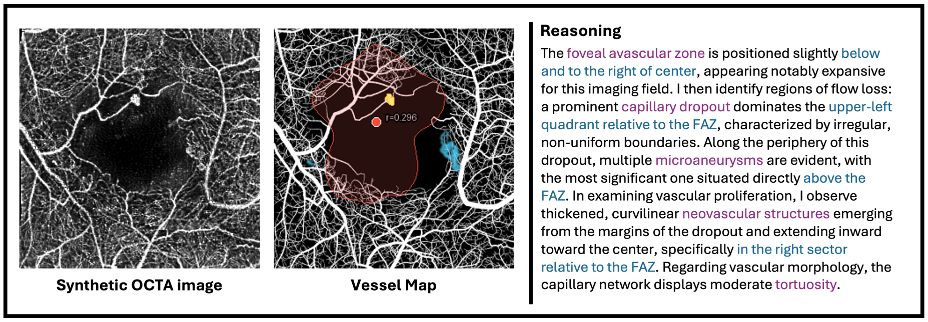
Capillary dropout. We first sample n non-perfusion regions around the FAZ. For a region centered at μ_k, its boundary $\tilde{R}_k$ is modeled as an irregular ellipse modulated with a series of angular harmonics and a smooth spatial noise field. We then define a continuous dropout field $c(\mathbf{x}) \in [0,1]$ that encodes how deep the regression should be at each location. Terminal vessel segments (capillaries) inside high-$c(\mathbf{x})$ areas (with threshold τ_{reg}) are stochastically pruned, with sampling weights that favor small-radius vessels and locations close to the lesion core. In addition, surviving vessels inside dropout regions undergo mild elongation and dilation to mimic remodeling at the borders of ischemic zones. The full definition of $c(\mathbf{x})$ and the pruning/remodeling rules is provided in Appendix A.2.

Microaneurysms (MA). Microaneurysms are modeled as short bulbous side branches that bud from existing vessels near dropout borders. For each arterial segment whose endpoint lies in a specified band b_{MA} of the dropout field, we sample a Bernoulli trial whose success probability is modulated by both the global dropout severity and the local value of $c(\mathbf{x})$, making microaneurysms more likely (but not guaranteed) near more severe dropout. Successful trials spawn a small perpendicular branch of length proportional to the simulator step size and with a radius r_{MA} drawn from a clinically plausible range (typically 20–80 μm). Additional child nodes around the MA center $\mathbf{x}_{\text{MA}}$ create irregular cluster-like shapes. All MA segments are stored explicitly in the graph and exported as part of the pathology metadata (see Appendix A.3).

Neovascularization (NV). Neovascularization is implemented as fine, tortuous sprouts that grow from arterial tips adjacent to dropout regions. We identify leaf tips in the 2D projection and, for a subset selected according to a global NV severity parameter, simulate short polylines $\mathbf{p}$ of effective length l_{NV} that extend away from the parent vessel, where l_{NV} corresponds to a small number of growth iterations per tuft. The step direction $\mathbf{v}_t$ at each growth step t combines the previous direction, a weak radial component relative to the nearest dropout center, and a small swirling field with random jitter, producing tuft-like shapes. We also spawn side branches along the main sprout to obtain clinically realistic NV shapes. A detailed description of the polyline construction and radius profiles is given in Appendix A.4.

Tortuosity along dropout borders. Finally, we increase vessel tortuosity in a narrow band b_{tort} around the dropout border. For arterial segments whose endpoints lie in this band, we jitter node positions in the direction perpendicular to the local vessel tangent by a small, zero-mean random offset whose amplitude A_{tort} is proportional to a tortuosity gain parameter. This preserves global connectivity and the FAZ geometry while increasing local curvature, producing the characteristic curling of vessels along non-perfusion borders. The exact definition of the tortuosity band and jitter distribution is provided in Appendix A.5.

Parameterization. All four pathology types share a small set of interpretable hyperparameters controlling count, size, strength, and probability. These parameters are sampled from ranges chosen to match clinical OCTA statistics. The full list of parameters and typical values is summarized in Tab. 5 in Appendix A.



Reasoning

The foveal avascular zone is positioned slightly below and to the right of center, appearing notably expansive for this imaging field. I then identify regions of flow loss: a prominent capillary dropout dominates the upper-left quadrant relative to the FAZ, characterized by irregular, non-uniform boundaries. Along the periphery of this dropout, multiple microaneurysms are evident, with the most significant one situated directly above the FAZ. In examining vascular proliferation, I observe thickened, curvilinear neovascular structures emerging from the margins of the dropout and extending inward toward the center, specifically in the right sector relative to the FAZ. Regarding vascular morphology, the capillary network displays moderate tortuosity.

Figure 3: **Synthetic OCTA image and generated reasoning text.** Purple phrases mark structures/pathologies that are directly controlled by the simulator, while the blue texts mark localizations.

2.3. VLM Pretraining and Fine-tuning

2.3.1. OCTA-100K-SVR Dataset

We refer to the full collection of our 100,000 synthetic image–reasoning pairs as the **OCTA-100K-SVR** dataset. For every simulated sample, the pathology module exports structured metadata describing FAZ geometry and the presence and location of capillary dropout, microaneurysms, neovascular tufts, and tortuosity. From this metadata we first construct a deterministic "template" reasoning paragraph that follows a fixed order (FAZ → dropout → microaneurysms → neovascularization → tortuosity) and states explicitly which patterns are present and where they lie relative to the FAZ and macular quadrants.

This template is then passed to a teacher model (GPT-5 (OpenAI, 2025)), which is prompted to rewrite the paragraph while strictly preserving all clinical facts and spatial relations. The teacher model diversifies the Chain-of-Thought descriptions that remain aligned with the metadata (see Fig. 3 and Appendix A.7). For each image we keep one such paragraph as the assistant answer a. The user question q is a short prompt asking the model to describe the OCTA image. Training with these image–reasoning pairs minimizes next-token cross-entropy. We use the same conversation format for the in-house OCTA data and add the explicit diagnosis sentence at the end of the reasoning paragraph for the fine-tuning stage.

2.3.2. SVR pretraining and fine-tuning

Our main reasoning model, denoted **SVR**, is obtained by pretraining the general-purpose Qwen3-VL-8b-Instruct (Yang et al., 2025) on the synthetic OCTA-100K-SVR dataset. We treat this as instruction-style supervised fine-tuning on the synthetic image–reasoning pairs while freezing the language module, so that only the vision encoder and multi-modal projection layers are updated. In this stage the model learns OCTA-specific visual features and their alignment to the existing language space without disturbing its general language capabilities. Models trained on smaller synthetic subsets (1k–75k samples) are used in the scaling experiments of Sec. 3.2.

To close the remaining domain gap, we then perform a second stage of supervised fine-tuning on clinical OCTA data, starting from an SVR-pretrained checkpoint. The resulting model is denoted **SVR-FT** and is trained on OCTA images from OCTA-500 and our in-house dataset together with reasoning and final DR labels. In this stage we train the model end-to-end, allowing it to adapt both its decision boundaries and linguistic style to real-world data while retaining the pathology-aware visual grounding acquired from synthetic vasculature. Training details and hyperparameters are summarized in Sec. 3.

3. Experiments

3.1. Datasets

We evaluate models on one synthetic dataset and two clinical datasets. Apart from OCTA-100K-SVR, the in-house dataset contains 1,286 OCTA deep capillary plexus scans labeled as Healthy, Non-proliferative DR (NPDR), or Proliferative DR (PDR) following the 5-fold setup of Lux et al. (2025). For OCTA-500 (Li et al., 2024), we follow the same test split of 189 scans as in Lux et al. (2025) and Li et al. (2025).

3.2. Tasks and metrics

We study two DR staging settings. On OCTA-500, models perform binary DR-vs.-Healthy classification. On the in-house dataset, models perform three-label staging (Healthy, NPDR, PDR). In both cases the VLM receives a single OCTA image and a prompt asking about DR staging. We report class-wise precision and recall and use *balanced accuracy* (mean per-class recall) as the primary metric.

We compare SVR-based models with supervised baselines (ResNet18 and a vessel-graph GNN (Lux et al., 2025)) and several VLM baselines: vanilla Qwen3-VL-8b/30b (Yang et al., 2025), LLaMA-3.2-11B-VL (Touvron et al., 2023), LLaVA-NEXT-8B (Liu et al., 2023), the graph-knowledge-based Qwen3-VL-8b-GFT (Li et al., 2025), and a baseline Qwen3-VL-8b-FT finetuned using only classification labels.

To quantify explanation quality we use GPT-5 as an automatic judge, similar to the ways in Liu et al. (2023) and Li et al. (2025). For each model, GPT-5 scores the candidate response on helpfulness, clinical and localization accuracy, and relevance, and we report the average score over these dimensions on both the in-house set and a held-out synthetic test set. Two ophthalmology experts then each evaluated 30 in-house cases, ranking the explanations under the same criteria. To verify if GPT-5 is a good judge and to study the correlation between the GPT-5 score and model performance (in terms of classification accuracy), we perform a scaling effects experiment in which Qwen3-VL-8b-SVR is trained on synthetic subsets from 1k to 100k samples and evaluated on OCTA-500. We track both balanced accuracy and GPT-5 Score as functions of synthetic dataset size, and observe a strong positive correlation (Fig. 6), indicating that GPT-5 scores are aligned with diagnostic accuracy and can therefore serve as a reliable automatic judge in our setting.

Ablation experiments isolate key components of SVR: (i) prompt diversification vs. static template reasoning, and (ii) the contribution of individual pathology types by removing dropout or microaneurysms from the reasoning. We additionally compare against a baseline (Qwen3-VL-8b CoT-FT) that fine-tunes on GPT-5-generated reasoning instead of SVR.

3.3. Training setup

All experiments are implemented with LLaMA-Factory (Zheng et al., 2024) on four NVIDIA H100 GPUs. OCTA images are resized to 512 × 512. In the SVR stage we fine-tune Qwen3-VL-8b-Instruct on OCTA-100K-SVR using a cosine learning-rate schedule, training the vision encoder and multimodal projector while freezing the language backbone. In the SVR-FT stage we start from the SVR checkpoint and continue supervised training end-to-end on clinical OCTA images with their reasoning and labels. Additional implementation details are provided in Appendix C.

4. Results and Discussion

4.1. DR Staging

4.1.1. Zero-shot and supervised fine-tuning

Tabs. 1 and 6 summarize DR staging performance on OCTA-500 and the in-house dataset. Directly fine-tuning a general-purpose VLM on limited classification labels (*Qwen3-VL-8b-FT*) is unstable and causes mode collapse, where the DR cases are neglected. In contrast, SVR pretraining (*Qwen3-VL-8b-SVR*) already produces competitive or better balanced accuracy in a zero-shot setting, despite never seeing real scans.

When subsequently fine-tuned on clinical data (*Qwen3-VL-8b-SVR-FT*), the model achieves consistently higher and more balanced recall across all stages, especially for advanced DR on both datasets. Compared with purely supervised CNN and GNN baselines, SVR-FT reaches similar or better balanced accuracy while additionally producing explanations. This suggests that SVR pretraining provides a strong reasoning capability regularizes downstream training, improves sensitivity to disease, and mitigates overfitting.

Table 1: DR staging on OCTA-500. H = Healthy, DR = Diabetic Retinopathy.

Model	**Prec(H)**	**Prec(DR)**	**Rec(H)**	**Rec(DR)**	**Bal.Acc**
ResNet18 (Lux et al., 2025)	0.8734	0.2903	0.8625	0.3103	0.5864
GNN (Lux et al., 2025)	0.9636	0.9583	0.9938	0.7931	0.8934
Qwen3-VL-8b (Yang et al., 2025)	0.8503	0.5000	0.9938	0.0345	0.5142
Qwen3-VL-30b (Yang et al., 2025)	0.8511	**1.0000**	**1.0000**	0.0345	0.5173
Qwen3-VL-8b-GFT (Li et al., 2025)	0.9718	0.5846	0.9347	0.7703	0.8525
	± 0.0073	± 0.0408	± 0.0091	± 0.0594	± 0.0275
Qwen3-VL-8b-FT	0.8551	0.2000	**1.0000**	0.0621	0.5310
	± 0.0189	± 0.4000	± 0.0000	± 0.1241	± 0.0620
Qwen3-VL-8b-SVR (ours)	0.9739	0.6944	0.9313	0.8621	0.8967
Qwen3-VL-8b-SVR-FT (ours)	**0.9762**	0.7944	0.9575	**0.8690**	**0.9133**
	± 0.0093	± 0.0567	± 0.0166	± 0.0581	± 0.0204

4.1.2. Scaling effects of synthetic data

The scaling experiment in Tab. 2 and Fig. 6 shows that increasing the size of OCTA-100K-SVR improves both zero-shot DR classification and GPT-5 scores. With very small synthetic subsets, performance is highly variable and remains close to chance level, indicating that limited coverage of vascular topologies and lesion patterns is insufficient for robust reasoning. Once the synthetic dataset reaches tens of thousands of samples, we observe a sharp improvement (See Fig. 6 in Appendix).

Interestingly, while classification performance begins to saturate at larger synthetic scales, explanation quality (GPT-5 score) continues to improve, and over the full range of dataset sizes the two metrics still exhibit a strong positive correlation (Fig. 6). This suggests that additional synthetic diversity could be utilized not only to refine decision boundaries but also to strengthen pathology localization and clinical explainability.

Table 2: Scaling Effects of Synthetic Data (zero-shot test on OCTA-500).

Size	Prec(H)	Prec(DR)	Rec(H)	Rec(DR)	Bal.Acc	GPT-5 Score
0	0.8503	0.5000	**0.9938**	0.0345	0.5142	50.3
1k	0.8281	0.1440	0.3312	0.6207	0.4759	57.7
5k	0.8508	0.2500	0.9625	0.0690	0.5158	62.7
10k	0.8492	0.1587	0.6687	0.3448	0.5067	60.9
25k	0.9292	0.1579	0.6207	0.6000	0.6103	63.9
50k	0.9329	0.6429	0.9375	0.6207	0.7791	69.0
75k	**0.9739**	**0.6944**	0.9313	**0.8621**	**0.8967**	68.5
100k	0.9533	0.5641	0.8938	0.7586	0.8262	**73.2**

4.1.3. ABLATION STUDIES

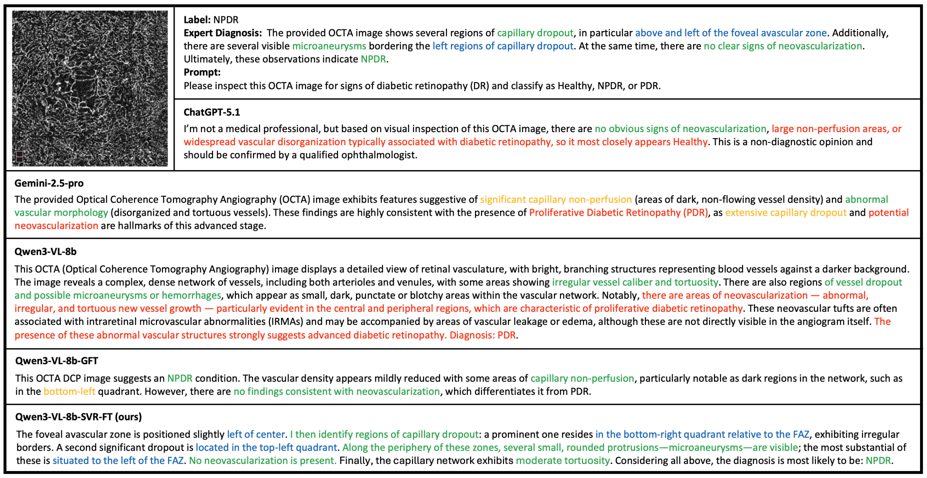
The ablation results in Tab. 3 highlight two central components of SVR. First, removing reasoning diversification (w/o diversifying) degrades downstream performance. This indicates that varied but fact-preserving CoT texts are critical for preventing the VLM from memorizing a fixed template and instead encouraging it to condition genuinely on the image content. Second, ablating individual pathology types (w/o dropout or microaneurysms) also decreases the performance, showing the importance of explicitly modeling all key DR hallmarks in the pipeline.

Table 3: Ablation Study.

Model	Prec(H)	Prec(DR)	Rec(H)	Rec(DR)	Bal.Acc
Qwen3-VL-8b (Yang et al., 2025)	0.8503	0.5000	0.9938	0.0345	0.5142
Qwen3-VL-8b CoT-FT	0.8951	0.4444	0.9063	0.4138	0.6601
Qwen3-VL-8b SVR (w/o diversifying)	0.8359	0.1311	0.6687	0.2759	0.4723
Qwen3-VL-8b SVR (w/o Dropout)	0.9444	0.7308	0.9563	0.6552	0.8058
Qwen3-VL-8b SVR (w/o MA)	0.8743	**1.0000**	**1.0000**	0.2069	0.6035
LLaMA-3.2-11B-VL-SVR (Touvron et al., 2023)	0.8889	**1.0000**	**1.0000**	0.3103	0.6552
LLaVA-NEXT-8b-SVR (Liu et al., 2023)	0.9433	0.2500	0.7870	0.6000	0.6935
Qwen3-VL-8b-SVR (ours)	0.9739	0.6944	0.9313	0.8621	0.8967
Qwen3-VL-8b-SVR-FT (ours)	**0.9762**	0.7944	0.9575	**0.8690**	**0.9133**

4.2. Quality of Explanation

We next examine explanation quality. Using GPT-5 as an automatic judge, we score each model response on both the in-house and synthetic test sets (Tab. 4), where SVR-FT obtains the highest GPT-5 scores in all settings. Expert ratings are also provided in Tab. 4, where SVR-FT again achieves the best result. Qualitative examples in Fig. 4 show that SVR-FT produces grounded descriptions that discuss FAZ, dropout, microaneurysms, neovascularization, and tortuosity, rather than generic or hallucinated findings.



Label: NPDR

Expert Diagnosis: The provided OCTA image shows several regions of capillary dropout, in particular above and left of the foveal avascular zone. Additionally, there are several visible microaneurysms bordering the left regions of capillary dropout. At the same time, there are no clear signs of neovascularization. Ultimately, these observations indicate NPDR.

Prompt:

Please inspect this OCTA image for signs of diabetic retinopathy (DR) and classify as Healthy, NPDR, or PDR.

ChatGPT-5.1

I'm not a medical professional, but based on visual inspection of this OCTA image, there are no obvious signs of neovascularization, large non-perfusion areas, or widespread vascular disorganization typically associated with diabetic retinopathy, so it most closely appears Healthy. This is a non-diagnostic opinion and should be confirmed by a qualified ophthalmologist.

Gemini-2.5-pro

The provided Optical Coherence Tomography Angiography (OCTA) image exhibits features suggestive of significant capillary non-perfusion (areas of dark, non-flowing vessel density) and abnormal vascular morphology (disorganized and tortuous vessels). These findings are highly consistent with the presence of Proliferative Diabetic Retinopathy (PDR), as extensive capillary dropout and potential neovascularization are hallmarks of this advanced stage.

Qwen3-VL-8b

This OCTA (Optical Coherence Tomography Angiography) image displays a detailed view of retinal vasculature, with bright, branching structures representing blood vessels against a darker background. The image reveals a complex, dense network of vessels, including both arterioles and venules, with some areas showing irregular vessel caliber and tortuosity. There are also regions of vessel dropout and possible microaneurysms or hemorrhages, which appear as small, dark, punctate or blotchy areas within the vascular network. Notably, there are areas of neovascularization — abnormal, irregular, and tortuous new vessel growth — particularly evident in the central and peripheral regions, which are characteristic of proliferative diabetic retinopathy. These neovascular tufts are often associated with intraretinal microvascular abnormalities (IRMAs) and may be accompanied by areas of vascular leakage or edema, although these are not directly visible in the angiogram itself. The presence of these abnormal vascular structures strongly suggests advanced diabetic retinopathy. Diagnosis: PDR.

Qwen3-VL-8b-GFT

This OCTA DCP image suggests an NPDR condition. The vascular density appears mildly reduced with some areas of capillary non-perfusion, particularly notable as dark regions in the network, such as in the bottom-left quadrant. However, there are no findings consistent with neovascularization, which differentiates it from PDR.

Qwen3-VL-8b-SVR-FT (ours)

The foveal avascular zone is positioned slightly left of center. I then identify regions of capillary dropout: a prominent one resides in the bottom-right quadrant relative to the FAZ, exhibiting irregular borders. A second significant dropout is located in the top-left quadrant. Along the periphery of these zones, several small, rounded protrusions—microaneurysms—are visible; the most substantial of these is situated to the left of the FAZ. No neovascularization is present. Finally, the capillary network exhibits moderate tortuosity. Considering all above, the diagnosis is most likely to be: NPDR.

Figure 4: **Comparison of model response.** True, partially true and false sentences are marked in green, yellow and red, respectively.

Table 4: Explanation quality evaluated by GPT-5 and human experts in ophthalmology (higher is better). Metrics: H = helpfulness, A = accuracy (localization & clinical), R = relevance, Avg = average score.

Model	**In-house Dataset**				**Synthetic Test Set**				**Expert Rating** (1–10)
	H	A	R	Avg	H	A	R	Avg	
GPT-5-mini (OpenAI, 2025)	76.2	68.7	82.9	75.9	73.4	65.1	79.2	72.6	4.915
Gemini-2.5-flash (Comanici et al., 2025)	69.7	52.1	78.3	66.7	46.8	37.9	53.7	46.1	5.860
Qwen3-VL-8b (Yang et al., 2025)	56.0	36.8	75.5	56.1	58.2	45.2	75.2	59.5	2.440
Qwen3-VL-8b-FT	40.1	37.1	59.0	45.4	43.4	47.8	57.2	49.5	2.178
Qwen3-VL-8b-GFT (Li et al., 2025)	71.3	67.9	74.2	71.1	72.5	69.4	75.1	72.3	6.445
Qwen3-VL-8b-SVR-FT (ours)	**80.8**	**70.6**	**91.6**	**81.0**	**87.6**	**83.9**	**95.3**	**89.0**	**6.985**

5. Conclusion

In this work, we presented a novel framework for training medical VLMs using synthesized images and pathological features. By simulating realistic retinal vasculature and specific DR pathologies we generated a large-scale synthetic dataset with precise ground-truth annotations and reasoning chains. Our experiments demonstrate that pretraining on this synthetic vasculature (SVR) substantially enhances VLM performance, allowing general-purpose models to outperform specialized supervised methods in diagnostic accuracy even without training on real clinical data, while providing interpretable, clinically meaningful explanations. Importantly, we showed that synthetic scaling could overcome the data scarcity bottleneck in medical imaging, transforming empirical knowledge into highly effective instruction-tuning data. Future work could extend this synthesis approach to 3D volumetric reasoning and multi-modal integration.

References

AI-READI Consortium. AI-READI:rethinking data collection, preparation, and sharing for propelling ai-based discoveries in diabetes research and beyond. *Nature Metabolism*, 6 (12):2210–2212, 2024. doi: 10.1038/s42255-024-01165-x.

Minhaj Alam, Yue Zhang, Jennifer I Lim, Robison VP Chan, Min Yang, and Xincheng Yao. Quantitative optical coherence tomography angiography features for objective classification and staging of diabetic retinopathy. *Retina*, 40(2):322–332, 2020.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025.

Ling Dai, Liang Wu, Huating Li, Chun Cai, Qiang Wu, Hongyu Kong, Ruhan Liu, Xiangning Wang, Xuhong Hou, Yuexing Liu, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature communications*, 12(1):3242, 2021.

Yoshihiro Kaizu, Shintaro Nakao, Shigeo Yoshida, Takehito Hayami, Mitsuru Arima, Muneo Yamaguchi, Iori Wada, Toshio Hisatomi, Yasuhiro Ikeda, Tatsuro Ishibashi, et al. Optical coherence tomography angiography reveals spatial bias of macular capillary dropout in diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, 58(11):4889–4897, 2017.

Linus Kreitner, Johannes C Paetzold, Nikolaus Rauch, Chen Chen, Ahmed M Hagag, Alaa E Fayed, Sobha Sivaprasad, Sebastian Rausch, Julian Weichsel, Bjoern H Menze, et al. Synthetic optical coherence tomography angiographs for detailed retinal vessel segmentation without human annotations. *IEEE Transactions on Medical Imaging*, 43(6):2061–2073, 2024.

Yuxiang Lai, Jike Zhong, Ming Li, Shitian Zhao, Yuheng Li, Konstantinos Psounis, and Xiaofeng Yang. Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. *arXiv preprint arXiv:2503.13939*, 2025.

Ryan Lee, Tien Y Wong, and Charumathi Sabanayagam. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye and vision*, 2(1):17, 2015.

Chenjun Li, Laurin Lux, Alexander H Berger, Martin J Menten, Mert R Sabuncu, and Johannes C Paetzold. Fine-tuning vision language models with graph-based knowledge for explainable medical image analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 198–207. Springer, 2025.

Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems*, 36:28541–28564, 2023.

Mingchao Li, Kun Huang, Qiuzhuo Xu, Jiadong Yang, Yuhan Zhang, Zexuan Ji, Keren Xie, Songtao Yuan, Qinghuai Liu, and Qiang Chen. Octa-500: a retinal dataset for optical coherence tomography angiography study. *Medical image analysis*, 93:103092, 2024.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning, 2023.

Laurin Lux, Alexander H Berger, Maria Romeo Tricas, Alaa E Fayed, Sobha Sivaprasada, Linus Kreitner, Jonas Weidner, Martin J Menten, Daniel Rueckert, and Johannes C Paetzold. Interpretable retinal disease prediction using biology-informed heterogeneous graph representations. *arXiv preprint arXiv:2502.16697*, 2025.

Weijian Ma, Ruoxin Chen, Keyue Zhang, Shuang Wu, and Shouhong Ding. Instruct where the model fails: Generative data augmentation via guided self-contrastive fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 5991–5999, 2025.

Martin J Menten, Johannes C Paetzold, Alina Dima, Bjoern H Menze, Benjamin Knier, and Daniel Rueckert. Physiology-based simulation of the retinal vasculature enables annotation-free segmentation of oct angiographs. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 330–340. Springer, 2022.

OpenAI. GPT-5. https://openai.com/gpt-5/, 2025. Accessed: 2025-11-13.

Jiazhen Pan, Che Liu, Junde Wu, Fenglin Liu, Jiayuan Zhu, Hongwei Bran Li, Chen Chen, Cheng Ouyang, and Daniel Rueckert. Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 337–347. Springer, 2025.

Chinmay Prabhakar, Suprosanna Shit, Johannes C Paetzold, Ivan Ezhov, Rajat Koner, Hongwei Li, Florian Sebastian Kofler, and Bjoern Menze. Vesselformer: Towards complete 3d vessel graph generation from images. In *Medical Imaging with Deep Learning*, pages 320–331. PMLR, 2024.

Nikolaus Rauch and Matthias Harders. Interactive synthesis of 3d geometries of blood vessels. In *Eurographics (Short Papers)*, pages 13–16, 2021.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, Madeleine Traverse, Timo Kohlberger, Shawn Xu, Fayaz Jamil, Cían Hughes, Charles Lau, et al. Medgemma technical report. *arXiv preprint arXiv:2507.05201*, 2025.

Zihan Sun, Dawei Yang, Ziqi Tang, Danny S Ng, and Carol Y Cheung. Optical coherence tomography angiography in diabetic retinopathy: an updated review. *Eye*, 35(1):149–161, 2021.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Bastian Wittmann, Lukas Glandorf, Johannes C Paetzold, Tamaz Amiranashvili, Thomas Wälchli, Daniel Razansky, and Bjoern Menze. Simulation-based segmentation of blood vessels in cerebral 3d octa images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 645–655. Springer, 2024.

Bin Wu, Wuxuan Shi, Jinqiao Wang, and Mang Ye. Synthetic data is an elegant gift for continual vision-language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2813–2823, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Kai Zhang, Rong Zhou, Eashan Adhikarla, Zhiling Yan, Yixin Liu, Jun Yu, Zhengliang Liu, Xun Chen, Brian D Davison, Hui Ren, et al. A generalist vision–language foundation model for diverse biomedical tasks. *Nature Medicine*, 30(11):3129–3141, 2024.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Bangkok, Thailand, 2024. Association for Computational Linguistics. URL `http://arxiv.org/abs/2403.13372`.

Appendix A. Details of Vessel and Pathology Simulation

This appendix provides the mathematical details of the vessel growth and pathology simulation summarized in Sec. 2.2. All coordinates are defined in a normalized 2D en-face domain $[0,1]^2$ unless otherwise stated, and radii in millimeters are mapped to pixel units via the same physical scale used by the baseline OCTA simulator (Kreitner et al., 2024).

A.1. Baseline healthy vessel simulation

We adopt the angiogenesis-based statistical growth model of Kreitner et al. (2024), which represents the vasculature as a forest of rooted binary trees growing in a 3D box

$$\Omega = [0,1] \times [0,1] \times [0,h_z]. \tag{1}$$

Each vessel segment is an edge

$$e = (\mathbf{x}_i, \mathbf{x}_j, r_{ij}), \tag{2}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \Omega$ are the 3D coordinates of the incident nodes and $r_{ij} > 0$ is the segment radius. Sibling segments satisfy Murray's law with bifurcation exponent κ,

$$r^{\kappa}_{\text{parent}} = r^{\kappa}_{\text{child},1} + r^{\kappa}_{\text{child},2}. \tag{3}$$

Growth proceeds in two phases (superficial and deep vascular complexes) by repeatedly sampling oxygen sinks (for arteries) and CO_2 sources (for veins) and letting leaf or inter-nodes sprout if the attraction points fall into a perception cone of distance δ and angle γ around the current segment. The growth direction is a weighted combination of the mean attraction vector and an optimal branching vector that minimizes deviation from the parent vessel.

FAZ center shift A circular exclusion zone around the FAZ with radius r_{FAZ} prevents sink placement in the foveal center. To increase variability, we randomly jitter the FAZ center by a vector $\Delta\mathbf{c}$ sampled from a disk of radius r_{jitter} and clamp the resulting center to a maximum normalized displacement $|\Delta\mathbf{c}| \leq r_{\text{max}}$. This effectively shifts the view while leaving the underlying vasculature unchanged.

After growth, the vessel graph is voxelized at high resolution and projected along the depth axis to obtain a grayscale vessel map X and a binarized segmentation label $\text{bin}(X)$. A pretrained GAN then converts X into a realistic OCTA image.

A.2. Capillary dropout field and vessel regression

Dropout regions and inside-score. Each dropout lesion k is specified by:

- center $\boldsymbol{\mu}_k = (c_{k,x}, c_{k,y}) \in [0,1]^2$,
- base radius $r_k > 0$ in normalized units,
- axis ratios $a_k, b_k > 0$ controlling ellipticity,
- harmonic amplitudes $A_{k,m}$ and phases $\phi_{k,m}$ for a small set of modes $\mathcal{H}_k$ (typically $m \in \{2,3,5\}$),

- shape exponent $\alpha_k > 0$,
- noise gain $g_k \geq 0$ and lesion strength $s_k \in [0, 1]$.

For a location $\mathbf{x} = (x, y)$ we define its coordinates relative to lesion k as

$$\boldsymbol{\delta}_k(\mathbf{x}) = \mathbf{x} - \boldsymbol{\mu}_k, \quad \rho_k(\mathbf{x}) = \|\boldsymbol{\delta}_k(\mathbf{x})\|_2, \quad \theta_k(\mathbf{x}) = \operatorname{atan2}(\delta_{k,y}, \delta_{k,x}). \tag{4}$$

The radius of the underlying ellipse in direction θ is

$$R_k(\theta) = \frac{r_k}{\sqrt{\left(\frac{\cos\theta}{a_k}\right)^2 + \left(\frac{\sin\theta}{b_k}\right)^2}}, \tag{5}$$

and we modulate it with angular harmonics to obtain an irregular boundary

$$\tilde{R}_k(\theta) = R_k(\theta)\left[1 + \sum_{m \in \mathcal{H}_k} A_{k,m}\cos\big(m\theta + \phi_{k,m}\big)\right]. \tag{6}$$

We use the notation $[z]_+ = \max(z, 0)$. The geometric “inside-score” of lesion k is

$$u_k(\mathbf{x}) = \left[1 - \frac{\rho_k(\mathbf{x})}{\tilde{R}_k(\theta_k(\mathbf{x}))}\right]_+^{\alpha_k}, \tag{7}$$

which smoothly decays from 1 at the center to 0 at the boundary.

To introduce additional boundary irregularity we construct a smooth noise field $n_k(\mathbf{x}) \in [0, 1]$ by summing a few sinusoidal components in x and y. With a noise gain g_k we define

$$c_k(\mathbf{x}) = u_k(\mathbf{x})\operatorname{clip}\big(0.75 + g_k(n_k(\mathbf{x}) - 0.5),\ 0,\ 1.2\big), \tag{8}$$

where $\operatorname{clip}(\cdot)$ clamps to the indicated range, and $c_k(\mathbf{x})$ is subsequently capped to $[0, 1]$.

The global dropout field is given by the maximum over lesions,

$$c(\mathbf{x}) = \max_k c_k(\mathbf{x}) \in [0, 1], \tag{9}$$

and the overall dropout severity by

$$s_{\max} = \max_k s_k \in [0, 1]. \tag{10}$$

Probabilistic pruning of capillaries. Let $\mathcal{L}$ denote the set of leaf nodes (terminal segments) in one vascular forest. Each leaf $i \in \mathcal{L}$ has position $\mathbf{x}_i$ and radius r_i, and we evaluate $c_i = c(\mathbf{x}_i)$. We only consider leaves with

$$c_i \geq \tau_{\text{reg}}, \tag{11}$$

where $\tau_{\text{reg}} \in (0, 1)$ is a regression threshold (we use $\tau_{\text{reg}} \approx 0.35$).

We target a global removal fraction

$$f_{\text{drop}} \approx s_{\max}, \tag{12}$$

with empirical lower/upper bounds to avoid trivial cases. Among eligible leaves we define sampling weights

$$w_i \propto (1 - c_i)^{\gamma}\, r_i^{-\alpha}, \quad i \in \mathcal{L}, \tag{13}$$

where $\gamma > 0$ biases removal towards the lesion core and $\alpha \in [0.3, 1]$ controls the preference for smaller vessels. We sample $\lfloor f_{\text{drop}}|\mathcal{L}|\rfloor$ leaves without replacement according to w_i and delete the corresponding segments from the graph.

Remodeling: elongation and dilation. For non-removed nodes we model subtle remodeling. Let node i with parent p have positions $\mathbf{x}_i$ and $\mathbf{x}_p$, local dropout $c_i = c(\mathbf{x}_i)$, direction

$$\mathbf{d}_{pi} = \mathbf{x}_i - \mathbf{x}_p, \tag{14}$$

and original radius r_i. We sample an elongation factor $e_i \in [e_{\min}, e_{\max}]$ and apply

$$\mathbf{x}_i^{(\text{new})} = \mathbf{x}_p + \big[1 + (e_i - 1)c_i\big]\, \mathbf{d}_{pi}, \tag{15}$$

which increases segment length more strongly near the lesion center.

Similarly, we dilate radii according to

$$r_i^{(\text{new})} = r_i \left[D_{\min} + (D_{\max} - D_{\min})c_i\right], \tag{16}$$

with $D_{\min}, D_{\max} \geq 1$ and optional global radius clamps.

A.3. Microaneurysm synthesis

Microaneurysms are modeled as short, roughly circular side branches emerging near dropout borders.

Spawn region. For each non-root arterial node i with parent p we compute $c_i = c(\mathbf{x}_i)$. If MAs are restricted to dropout borders, we require

$$c_{\min}^{\text{MA}} \leq c_i \leq c_{\max}^{\text{MA}}, \tag{17}$$

for some band $b_{MA} : (c_{\min}^{\text{MA}}, c_{\max}^{\text{MA}}) \subset (0, 1)$.

Spawn probability. For each eligible node we perform a Bernoulli trial with probability

$$p_{\text{MA}}(\mathbf{x}_i) = p_0\big(1 + \lambda_s s_{\max}\big)\big(1 + \lambda_c c_i\big), \tag{18}$$

where p_0 is a base MA density, λ_s couples MA counts to global dropout severity, and λ_c increases density near dropout borders. We also reweight p_{MA} by the area of the largest dropout region, so larger lesions tend to host more MAs.

MA geometry. We form a short side branch. Let

$$\mathbf{u}_{pi} = \frac{\mathbf{x}_i - \mathbf{x}_p}{\|\mathbf{x}_i - \mathbf{x}_p\|_2} \tag{19}$$

be the parent-to-child direction and

$$\mathbf{u}_{\perp} = (-u_{pi,y}, u_{pi,x}, 0) \tag{20}$$

a perpendicular unit vector in the en-face plane. With simulator step size d and MA length factor ℓ_{MA} we place the MA center at

$$\mathbf{x}_{\text{MA}} = \mathbf{x}_i + \ell_{\text{MA}} d\, \mathbf{u}_{\perp}. \tag{21}$$

The MA radius is sampled as

$$r_{\text{MA}} \sim \mathcal{U}\big(r_{\min}^{\text{MA}}, r_{\max}^{\text{MA}}\big), \tag{22}$$

where $r_{\min}^{\text{MA}}$ and $r_{\max}^{\text{MA}}$ are specified in millimeters. Additional child nodes can be sampled in a small disk around $\mathbf{x}_{\text{MA}}$ with slightly reduced radii to create irregular clusters.

A.4. Neovascular tufts

Neovascularization is represented as thin, tortuous sprouts grown from existing vessel tips.

Tip selection. We project arterial segments to 2D, identify leaf tips j with position $\mathbf{x}_j$, tangent direction $\mathbf{u}_j$, and local dropout value $c(\mathbf{x}_j)$. Tips too close to the FAZ or clearly outside dropout are down-weighted. A global NV severity parameter $s_{\mathrm{NV}} \in [0,1]$ determines the number of NV groups G and typical sprout length.

Main tuft growth. For each selected tip we initialize a polyline $\{\mathbf{p}_t^{(g)}\}_{t=0}^{T_{\mathrm{main}}}$ with

$$\mathbf{p}_0^{(g)} = \mathbf{x}_j. \tag{23}$$

At iteration t we update

$$\mathbf{p}_{t+1}^{(g)} = \mathbf{p}_t^{(g)} + \ell_{\mathrm{NV}} \mathbf{v}_t^{(g)}, \tag{24}$$

where ℓ_{NV} is a small step size and $\mathbf{v}_t^{(g)}$ is a unit direction obtained as a weighted combination of:

1. previous direction $\mathbf{v}_{t-1}^{(g)}$ (persistence),
2. a weak radial vector pointing away from the closest dropout center,
3. a low-frequency swirling field and isotropic jitter.

We clamp positions to remain within dropout regions and outside the FAZ.

Radius profile and side branches. Along each polyline we use a linearly tapering radius

$$r_{\mathrm{NV}}(t) = (1-\tau_t) r_{\mathrm{start}} + \tau_t r_{\mathrm{end}}, \quad \tau_t = \frac{t}{T_{\mathrm{main}}}, \tag{25}$$

with r_{start} proportional to the parent vessel radius and $r_{\mathrm{end}} < r_{\mathrm{start}}$. With probability increasing in s_{NV} we spawn side branches starting from intermediate points. These follow the same growth rule but with shorter maximum length T_{side}.

A.5. Tortuosity along dropout borders

To model increased tortuosity at dropout borders we jitter node positions perpendicular to the local vessel direction within a band of the dropout field.

For a non-root arterial node i with parent p, we define the tangent direction

$$\mathbf{u}_{pi} = \frac{\mathbf{x}_i - \mathbf{x}_p}{\|\mathbf{x}_i - \mathbf{x}_p\|_2}, \tag{26}$$

and choose a perpendicular direction $\mathbf{u}_\perp$ as above. Let $c_i = c(\mathbf{x}_i)$ and specify a tortuosity band b_{tort} where

$$c_{\mathrm{min}}^{\mathrm{tort}} \le c_i \le c_{\mathrm{max}}^{\mathrm{tort}}. \tag{27}$$

Nodes outside this band are left unchanged.

Given a gain parameter $g_{\mathrm{tort}} \in [0,1]$ and simulator step size d, we set the jitter amplitude

$$A_{\mathrm{tort}} = 0.35\, g_{\mathrm{tort}}\, d, \tag{28}$$

sample

$$\epsilon_i \sim \mathcal{U}(-A_{\text{tort}}, A_{\text{tort}}), \tag{29}$$

and update the node position as

$$\mathbf{x}_i^{(\text{new})} = \mathbf{x}_i + \epsilon_i \, \mathbf{u}_\perp, \tag{30}$$

clipping back to $[0, 1]^2$ if necessary. This leaves the topology and global FAZ shape intact but increases local curvature.

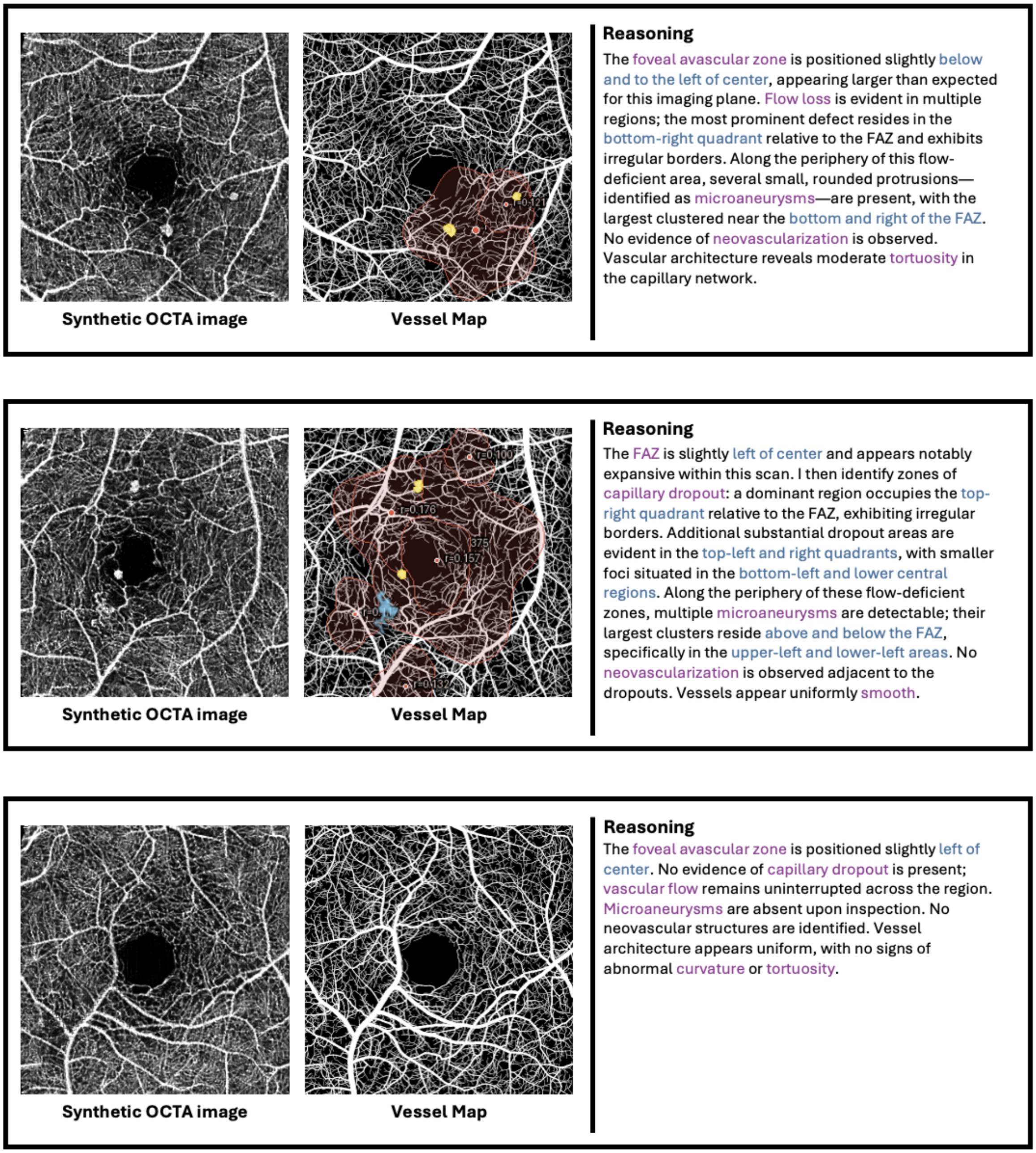
A.6. Parameter ranges and typical settings

Tab. 5 summarizes the key parameters that control the four DR lesion types and the typical ranges used in the OCTA-100K-SVR dataset. Spatial coordinates are normalized to $[0, 1]^2$; radii in millimeters refer to the physical scale of the baseline simulator.

Table 5: Key parameters of the simulator and typical settings used in this work.

Pathology	Parameter	Typical value / range
Dropout	# regions n	$[0, 6]$ per sample
	Radius r_{drop}	$[0.18, 0.32]$ (normalized)
	Lesion strength s_k	$[0.90, 0.99]$
	Gradient exponent α_k	$[2.0, 3.0]$
	Noise gain g_k	$[0.20, 0.40]$
MA	Base density p_0	≈ 0.03 (per segment)
	Radius r_{MA}	$[0.01, 0.08]$ mm
	Length factor ℓ_{MA}	$[0.3, 0.4]$ (in units of d)
	Strength coupling λ_s	≈ 15
NV	NV probability	≈ 0.4
	Severity s_{NV}	$[0.2, 0.7]$
	Footprint radius	$[0.015, 0.07]$ (normalized)
	Sprout length l_{NV}	$[3, 6]$ growth steps
Tortuosity	Gain g_{tort}	$[0.01, 0.5]$
	Band $[c^{\text{tort}}_{\text{min}}, c^{\text{tort}}_{\text{max}}]$	$[0.30, 0.75]$

A.7. Extra Examples of Reasoning Texts



Reasoning

The foveal avascular zone is positioned slightly below and to the left of center, appearing larger than expected for this imaging plane. Flow loss is evident in multiple regions; the most prominent defect resides in the bottom-right quadrant relative to the FAZ and exhibits irregular borders. Along the periphery of this flow-deficient area, several small, rounded protrusions—identified as microaneurysms—are present, with the largest clustered near the bottom and right of the FAZ. No evidence of neovascularization is observed. Vascular architecture reveals moderate tortuosity in the capillary network.

Reasoning

The FAZ is slightly left of center and appears notably expansive within this scan. I then identify zones of capillary dropout: a dominant region occupies the top-right quadrant relative to the FAZ, exhibiting irregular borders. Additional substantial dropout areas are evident in the top-left and right quadrants, with smaller foci situated in the bottom-left and lower central regions. Along the periphery of these flow-deficient zones, multiple microaneurysms are detectable; their largest clusters reside above and below the FAZ, specifically in the upper-left and lower-left areas. No neovascularization is observed adjacent to the dropouts. Vessels appear uniformly smooth.

Reasoning

The foveal avascular zone is positioned slightly left of center. No evidence of capillary dropout is present; vascular flow remains uninterrupted across the region. Microaneurysms are absent upon inspection. No neovascular structures are identified. Vessel architecture appears uniform, with no signs of abnormal curvature or tortuosity.

Figure 5: Extra Examples of Synthetic OCTA and Reasoning.

Appendix B. Pseudocode of the SVR Pipeline

```
Algorithm 1: Synthetic Vasculature Reasoning (SVR) pipeline
Input: number of synthetic samples $N$; vessel-growth config $\theta_{\text{vessel}}$; pathology profile
       $\theta_{\text{path}}$; optional GAN config $\theta_{\text{GAN}}$; teacher VLM $T$; base VLM $M$
Output: synthetic image–text dataset $\mathcal{D}$; pretrained model $M_{\text{SVR}}$; fine-tuned model
        $M_{\text{SVR-FT}}$
Initialize dataset $\mathcal{D} \leftarrow \varnothing$ for $i = 1$ to $N$ do
    Sample structural parameters (FAZ, seed, view shift, etc.)
    Simulate healthy vessel graph $G_i$
    Apply pathology profile $\theta_{\text{path}}$ to get augmented graph $G_i^*$
    Rasterize $G_i^*$ into a 2D vessel map $X_i$
    if GAN is enabled then
        Generate OCTA-like image $I_i$ from $X_i$ using $\theta_{\text{GAN}}$
    else
        Set $I_i \leftarrow X_i$
    end
    Extract structured metadata $m_i$ from the simulator (FAZ, dropout, MA, NV, tortuosity)
    Build template reasoning $r_i^{\text{tmp}}$ from $m_i$ (FAZ → dropout → MA → NV → tortuosity)
    Query teacher VLM $T$ with $(I_i, m_i, r_i^{\text{tmp}})$ to obtain a diversified, fact-preserving
      reasoning $r_i$
    Define question $q_i$ asking to describe $I_i$ and assess DR stage
    Define answer $a_i$ as $r_i$ (plus final diagnosis sentence for clinical data)  Add sample
      $(I_i, q_i, a_i, m_i)$ to dataset $\mathcal{D}$
end
Pretrain base VLM $M$ on $\mathcal{D}$ (freeze language backbone) to obtain $M_{\text{SVR}}$  Fine-tune
  $M_{\text{SVR}}$ end-to-end on clinical OCTA data with reasoning and DR labels to obtain
  $M_{\text{SVR-FT}}$
```

Appendix C. Training details

SVR training prompt (student VLM). During both SVR pretraining and SVR-FT fine-tuning, the VLM is trained in a single-turn instruction-following format. The human turn consists of the OCTA image token plus the following question (identical for synthetic and clinical data):

```
<image>
What features are visible in this OCTA image?
Please first describe the image features, then inspect it for signs
of diabetic retinopathy (DR) and classify it as Healthy, NPDR, or PDR.
```

The assistant turn contains the full Chain-of-Thought reasoning followed by the final diagnosis sentence.

CoT diversification prompts (teacher VLM). To obtain linguistically diverse and fact-preserving reasoning, we prompt GPT-5 with a system prompt and a user prompt that include the synthetic OCTA image, the original CoT and structured pathology metadata. The system prompt is:

```
You are an ophthalmology OCTA expert and skilled medical writer.
You will receive an OCTA image, concise metadata, and an original
chain-of-thought (CoT). Rewrite the CoT in a different language
style while preserving ALL medical facts, locations, and uncertainty.
Do not add new findings. Keep content consistent with the image and
metadata. Spatial terminology constraint: avoid eye-dependent terms
(e.g., temporal, nasal, superotemporal, inferonasal, superior/inferior
when tied to eye laterality). Use only absolute image directions such
as left, right, up, down, and center to describe locations.
Aim for similar length and clarity. Output only the rewritten CoT.
```

The corresponding user prompt is:

```
Here is an OCTA image <image>.
Metadata (JSON): <COMPACT_METADATA_JSON>

Original CoT describing the image:
<ORIGINAL_COT>

Task: Rewrite the CoT with a distinct language style (e.g., more
academic, more succinct, or slightly conversational) while preserving
all facts and spatial relations. Do not invent new content.
Use only absolute image directions such as left, right, up, down, and center.
Return only the rewritten CoT.
```

where `<COMPACT_METADATA_JSON>` is the compacted pathology metadata (FAZ, dropout, MA, NV, tortuosity) extracted from the simulator and `<ORIGINAL_COT>` is the deterministic template reasoning.

For robustness to natural-language variation in the input question, we also diversify the user prompt itself using a similar pair of prompts.

Appendix D. Supplementary Details in Evaluation

At inference we use temperature 0.1 and top-$p = 0.8$, and compute all metrics from unified JSONL outputs to ensure consistent parsing across models. To evaluate using GPT-5 scores, for each image, question, and model response, GPT-5 receives (i) the OCTA image, (ii) the dataset tag (synthetic vs. in-house), (iii) the question text, (iv) the ground-truth DR label when available, (v) an optional synthetic reference explanation (for synthetic data), and (vi) the candidate response. The system message is the fixed instruction:

```
You are a retina specialist. First write your response to the question with
all the information provided, and then score a single model response
```

```
by comparing it with yours on:
- helpfulness (0-10): clarity, usefulness, specificity to the question.
- accuracy (0-10): localization + clinical correctness (use the image;
  if GT label provided, consider it).
- relevance (0-10): how on-topic and focused the response is.
Return a compact JSON: {"helpfulness": int, "accuracy": int,
"relevance": int, "rationale": string}.
```

and the user message concatenates the textual context (dataset, question, GT label, reference, model response) with the OCTA image. We then parse the `helpfulness`, `accuracy`, and `relevance` integers and average them to obtain the per-sample "GPT-5 Score", which is then aggregated over all samples for each model and dataset.

Table 6: DR staging on the in-house OCTA dataset. H = Healthy, N = NPDR, P = PDR.

Model	**Prec(H)**	**Prec(N)**	**Prec(P)**	**Rec(H)**	**Rec(N)**	**Rec(P)**	**Bal.Acc**
ResNet18 (Lux et al., 2025)	0.943 ±0.008	0.335 ±0.044	0.426 ±0.070	0.793 ±0.021	0.544 ±0.094	0.563 ±0.131	0.633 ±0.030
GNN (Lux et al., 2025)	0.950 ±0.010	0.326 ±0.049	**0.456** ±0.107	0.720 ±0.035	0.594 ±0.091	**0.775** ±0.034	**0.697** ±0.034
Qwen3-VL-8b (Yang et al., 2025)	0.775	0.198	0.065	0.576	0.163	0.255	0.331
Qwen3-VL-30b (Yang et al., 2025)	0.787	0.228	0.444	**0.849**	0.241	0.041	0.377
Qwen3-VL-8b-GFT (Li et al., 2025)	0.889 ±0.007	**0.535** ±0.045	0.454 ±0.017	0.881 ±0.019	0.502 ±0.027	0.550 ±0.041	0.645 ±0.017
Qwen3-VL-8b-FT	0.908 ±0.014	0.364 ±0.084	0.245 ±0.322	0.828 ±0.071	**0.636** ±0.061	0.073 ±0.107	0.512 ±0.050
Qwen3-VL-8b-SVR (ours)	0.928	0.295	0.357	0.786	0.453	0.510	0.583
Qwen3-VL-8b-SVR-FT (ours)	**0.958** ±0.024	0.411 ±0.087	0.418 ±0.016	0.840 ±0.082	0.488 ±0.088	0.737 ±0.074	0.688 ±0.027

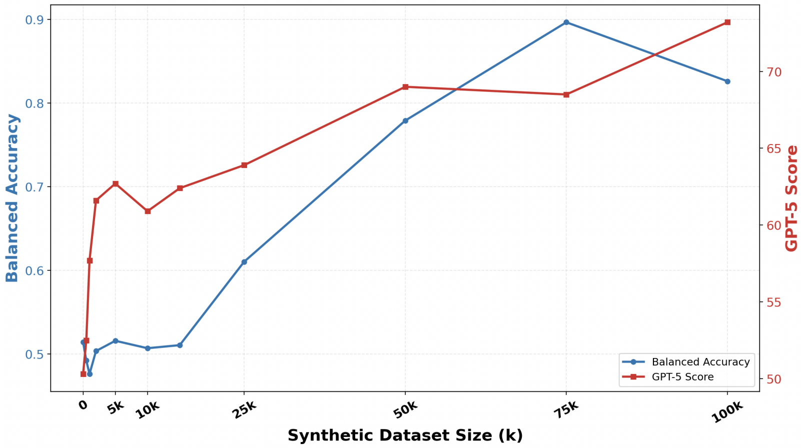


Figure 6: Correlation between Classification Performance and GPT-5 Score

Label: PDR
Expert Diagnosis: The main regions of non-perfusion are in the top-left and bottom-right areas neighboring the foveal avascular zone. There is one clearly visible microaneurysm at the edge of the bottom-right non-perfused area. There are indications of others in top-left area and top-right corner of the image.
Prompt:
Locate the capillary non-perfusion regions and microaneurysms on this OCTA image.

ChatGPT-5.1

The OCTA image shows a central capillary non-perfusion region, identifiable as the darker, vessel-depleted area in the middle where the normal capillary network is absent. Surrounding this zone, several small, focal hyper-reflective spots along dilated or irregular capillary segments suggest microaneurysms, particularly along the brighter, bulbous vessel outpouchings at the margins of the non-perfused region.

Gemini-2.5-pro

Based on the visual evidence in this OCTA scan, the capillary non-perfusion regions appear as irregular, dark flow voids where the capillary texture is noticeably absent; these are most prominent immediately surrounding the enlarged Foveal Avascular Zone (FAZ), particularly extending into the inferior (6 o'clock) and superior-temporal (10-11 o'clock) quadrants. Microaneurysms are identifiable as focal, hyper-reflective (bright) spots or tiny saccular dilations at the ends of capillary branches, with probable candidates visible in the inferotemporal region and clustered near the areas of vascular dropout.

Qwen3-VL-8b

Based on the provided OCTA image, here is the analysis for diabetic retinopathy (DR) features:
Capillary Non-Perfusion Regions:
The central dark area in the image represents a capillary non-perfusion region.
This is a hallmark sign of diabetic retinopathy, indicating areas where blood flow is absent due to vessel occlusion or damage.
The non-perfusion appears as a dark, well-defined zone surrounded by a network of perfused vessels.
Microaneurysms:
Microaneurysms are typically seen as small, round, bright spots on OCTA images.
In this image, there are no clearly identifiable microaneurysms.
The bright spots visible are likely to be retinal vessels or flow artifacts, not microaneurysms.

Qwen3-VL-8b-GFT

The capillary non-perfusion regions are located in the top-left, top-right, and bottom-right quadrants, while the microaneurysms are located in the bottom-left quadrant.

Qwen3-VL-8b-SVR (ours, trained on synthetic data only)

Flow voids are evident: the most prominent occupies the upper-left quadrant relative to the FAZ, exhibiting irregular margins. Additional substantial flow deficits are present in the upper and lower regions, while smaller ones cluster in the lower-left quadrant. Along the periphery of these flow-deficient zones, microaneurysms are visible, with the largest situated in the bottom-right quadrant relative to the FAZ.

Qwen3-VL-8b-SVR-FT (ours)

I first identify regions of capillary dropout: the most prominent is situated in the top-left quadrant relative to the FAZ, exhibiting irregular borders. Additional dropout zones are found in the bottom-left quadrant. Along the periphery of these dropout regions, several microaneurysms are visible, with the largest clustered in the bottom-right quadrant relative to the FAZ.

Figure 7: Extra Example for Qualitative Evaluation.

Label: PDR
Expert Diagnosis: The OCTA image shows large areas of capillary dropout, several microaneurysms, venous beading. Additionally, neovascularization has set in, resulting in an enlarged, irregularly shaped foveal avascular zone with a few tortuous vessels traversing it. The patient is most likely affected by proliferative diabetic retinopathy.
Prompt:
Please inspect this OCTA image for signs of diabetic retinopathy (DR) and classify as Healthy, NPDR, or PDR.

ChatGPT-5.1

Based on the OCTA image, the foveal avascular zone (FAZ) appears enlarged and irregular, with noticeable rarefaction of the surrounding capillary network. There is significant capillary dropout and reduced vascular density, but no clear signs of neovascularization such as abnormal, tuft-like proliferative vessels extending beyond the normal vascular boundaries. These findings are consistent with non-proliferative diabetic retinopathy (NPDR) rather than Healthy or PDR.

Gemini-2.5-pro

The provided Optical Coherence Tomography Angiography (OCTA) image shows a clear, central region lacking blood flow, known as Foveal Avascular Zone (FAZ) enlargement. The surrounding retinal capillary network also appears to be dilated, tortuous, and irregular, with areas of capillary non-perfusion (capillary dropout) extending into the periphery. Crucially, there is an absence of definitive neovascularization (new, abnormal vessel growth). These findings—FAZ enlargement and widespread capillary non-perfusion without neovascularization—are hallmark features of Non-Proliferative Diabetic Retinopathy (NPDR). Therefore, the image should be classified as NPDR.

Qwen3-VL-8b

This OCTA image displays a central avascular zone surrounded by a dense, irregular network of fine, branching vascular structures. The central area appears dark and devoid of flow, suggesting a significant loss of perfusion. The surrounding vasculature is tortuous and appears to be forming a “circular” or “ring-like” pattern around the central defect, with some vessels appearing dilated or irregular in caliber. There is no clear evidence of neovascularization (new, abnormal vessels) extending into the vitreous or subretinal space, nor are there obvious signs of hemorrhage or exudates. In the context of diabetic retinopathy, this pattern is highly suggestive of non-proliferative diabetic retinopathy (NPDR), specifically the “ischemic” or “severe NPDR” stage, where there is significant loss of capillary perfusion in the macula (central retinal ischemia).

Qwen3-VL-8b-GFT

This OCTA DCP image suggests a PDR (Proliferative Diabetic Retinopathy) condition. The extensive areas of capillary non-perfusion, characterized by large flow voids, are more severe and widespread compared to what would be expected in NPDR. Additionally, there are fine, tuft-like neovascularizations with brighter flow signals visible near the 7 o'clock position relative to the central FAZ, which is a hallmark of PDR but not typically seen in NPDR.

Qwen3-VL-8b-SVR (ours, trained on synthetic data only)

I first locate the foveal avascular zone: it is off center toward the bottom-right. The FAZ appears relatively large for this field of view. First, I check for areas of flow loss: There are multiple areas of marked capillary flow loss; the largest sits in the left area relative to the FAZ, with uneven edges. Additional sizeable regions lie in the top-right and bottom-right areas. Next, I look along the edges of flow loss: numerous small bulges (microaneurysms) are seen. The largest lies bottom-right of the FAZ. Finally, I look for new vessels: in multiple areas, thicker, curved branches sprout from dropout borders and extend inward, located in the bottom-left and bottom-right areas relative to the FAZ. Lastly, I look at vessel shape: capillaries show moderate extra twisting. Diagnosis: PDR.

Qwen3-VL-8b-SVR-FT (ours)

The foveal avascular zone is positioned slightly right of center and appears notably enlarged within this field. I then identify regions of capillary dropout: a a dominant, irregularly bordered area resides in the bottom-left quadrant relative to the FAZ. Along the periphery of these dropout regions, multiple small, rounded protrusions—microaneurysms—are evident, with the most prominent situated in the bottom-left. Further inspection reveals neovascular structures emerging from the borders of dropouts, extending inward toward the center, particularly in the bottom-left and left quadrants relative to the FAZ. Capillary architecture throughout shows higher tortuosity. Considering all above, the diagnosis is most likely to be PDR.

Figure 8: Extra Example for Qualitative Evaluation.