Evo-RAD: Navigating Rare Retinal Disease Diagnosis via Self-Evolving Agentic Retrieval

Wangding Xia[1]*, Ye Du[1]*, Jiashi Lin[2], Meng Wang[3,4], Danli Shi[5,6], and Shujun Wang[1]✉

[1] Department of Biomedical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China
[2] School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[3] Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore
[4] Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore
[5] School of Optometry, The Hong Kong Polytechnic University, Hong Kong SAR, China
[6] Research Centre for SHARP Vision (RCSV), The Hong Kong Polytechnic University, Hong Kong SAR, China
shu-jun.wang@polyu.edu.hk
✉ Corresponding author
*These authors contributed equally to this work.

Abstract. Large-scale pretrained foundation models have revolutionized general medical screening, but often falter on rare diseases because such conditions are underrepresented in real-world clinical datasets. While retrieval-augmented diagnosis attempts to mitigate this, conventional static methods frequently succumb to the *hubness problem*, retrieving visually similar but semantically incorrect common diseases. To address this, we propose **Evo-RAD**, a self-evolving agentic framework that transforms evidence acquisition into a dynamic decision-making task. We formulate retrieval as a *Markov Decision Process* (MDP) where a graph-based agent observes the reference set *state* and executes *actions* to purge discordant evidence (DELETE), acquire pathologically consistent samples (INSERT), or conclude the evolution (TERMINATE). Optimized via Group Relative Policy Optimization (GRPO) with a homogeneity-aware reward, the agent learns to maximize the diagnostic homogeneity of the support reference set. Experiments on retinal disease benchmarks show that Evo-RAD substantially improves rare-disease diagnosis, outperforming retinal foundation models by **+21.04%**, while also surpassing retrieval-based and parameter-efficient fine-tuning methods by **+3.56%**. Code is available at https://github.com/SDH-Lab/Evo-RAD.

Keywords: Retinal Disease · Rare Disease Diagnosis · Vision Language Model · Retrieval Augmented Diagnosis

1 Introduction

Foundation models are rapidly reshaping ophthalmic diagnosis by learning unified representations across retinal images and clinical language [21,19,15,16,18]. Recent Vision-Language Models (VLMs), such as FLAIR [16] and RetiZero [18], demonstrate strong performance on public benchmarks by aligning retinal imagery with clinical narratives, and have thus become attractive for automated screening. However, deploying these VLMs into rare retinal diseases diagnosis remains challenging. Rare diseases are infrequent and clinically heterogeneous [13], so their features are weakly constrained during pretraining and can be entangled with visually similar common diseases, leading to degraded real-world performance (lower sensitivity and higher confusion with prevalent diagnoses).

To mitigate this degradation, an intuitive strategy is to adapt foundation models via *Parameter-Efficient Fine-Tuning (PEFT)*, such as prompt tuning [22,2,9,4,10,12] and adapter-based methods [20,5]. PEFT freezes the backbone and updates only a small set of additional or selected parameters, thereby reducing computational costs. However, PEFT can be unreliable for rare classes. With only a handful of rare cases, even this limited trainable capacity can overfit to acquisition-specific noise and spurious cues. Conversely, when trained on the full imbalanced corpus, optimization is dominated by common classes, biasing the learned adaptation toward prevalent conditions and limiting gains in rare-disease sensitivity where robustness is most critical.

Besides PEFT, a more robust alternative is *Retrieval-Augmented Diagnosis (RAD)*. Given a query image, RAD retrieves a small set of similar reference cases from an external database and conditions diagnosis on the retrieved examples as supporting evidence. However, most existing RAD pipelines [15,11] still rely on one-shot nearest-neighbor retrieval with a fixed similarity metric, producing a static reference set. This rigidity design is fragile for rare diseases due to the *hubness problem* [14], where a small subset of common-disease samples repeatedly appears as nearest neighbors. As a result, the retrieved evidence can be visually similar but semantically incorrect, which biases prediction toward common classes and undermines the reliability of retrieval-based support.

To overcome these limitations, we argue that retrieval for rare-disease diagnosis should move beyond passive nearest-neighbor matching to an iterative, context-aware reasoning process, analogous to how clinicians refine hypotheses by accumulating consistent evidence and excluding mimics [3]. Motivated by this, we propose **Evo-RAD**, a self-evolving agentic retrieval framework that formulates retrieval as a *Markov Decision Process (MDP)* [17]: a graph-based agent observes the current reference-set *state* and sequentially evolves it via **DELETE** (purge discordant hub samples), **INSERT** (add pathologically consistent cases from external memory), or **TERMINATE**. We optimize the policy with Group Relative Policy Optimization (GRPO) [6] under a *homogeneity-aware reward* that promotes diagnostic coherence, so final prediction is conditioned on progressively evolved evidence, improving support for rare retinal disease diagnosis.

To summarize, our contributions are threefold. **First**, we propose **Evo-RAD**, an agentic retrieval framework moving beyond static similarity search to itera-

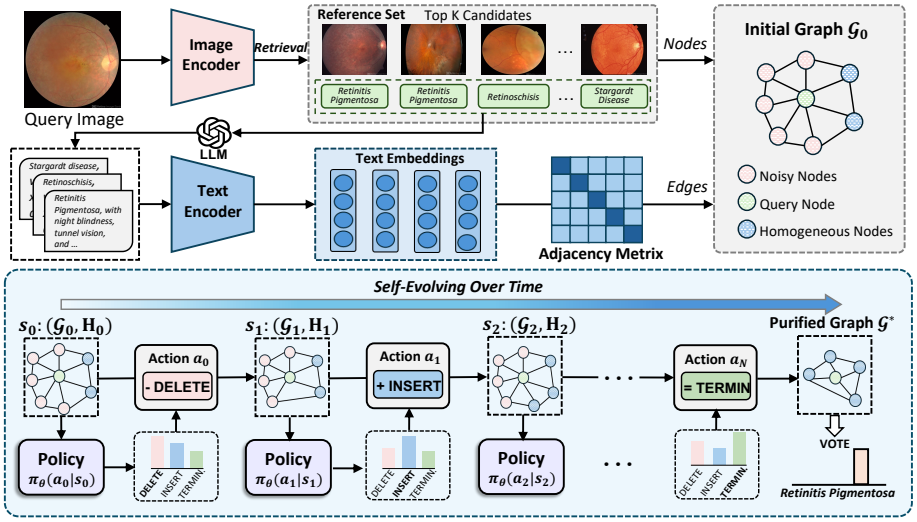


Fig. 1: **Overview of Evo-RAD.** Evo-RAD models retrieval as a Markov Decision Process where a graph policy evolves the evidence set through discrete DELETE, INSERT, and TERMINATE operations.

tive, dynamic reference set refinement. **Second**, we design a *homogeneity-aware* optimization objective and train the retrieval agent with GRPO, enabling discrete evidence evolving operations. **Third**, We conduct extensive experiments on retinal disease benchmarks, demonstrating that Evo-RAD yields a $+$**21.04%** improvement over retinal foundation models and a $+$**3.56%** gain over state-of-the-art static retrieval and PEFT methods.

2 Method

Evo-RAD (Fig. 1) is a self-evolving agentic retrieval framework for rare-disease diagnosis. It reformulates Top-K retrieval as a sequential decision process that iteratively evolves the reference set to remove hub-driven distractors and admit clinically concordant cases, terminating once the evidence is diagnostically coherent. Technically, Evo-RAD introduces: (i) a Markov Decision Process-based self-evolving agentic retrieval formulation; (ii) a GRPO-trained graph policy that jointly optimizes query affinity and intra-set agreement; and (iii) a final evidence-conditioned prediction head.

2.1 Markov Decision Process based Self-Evolving Agentic Retrieval

Given a query image q, we extract its visual embedding v_q using a frozen retinal foundation model and retrieve the Top-K most similar images with corresponding labels to form the initial reference set $\mathcal{X}_0 \subset \mathcal{D}$, where $\mathcal{D}$ is a

rare disease database (*e.g.*, the training subset). To enable reference set update, we model the self-evolving agentic retrieval as a Markov Decision Process $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R} \rangle$ [17], where the environment maintains an continual updating reference set $\mathcal{X}_t \subset \mathcal{D}$ and the agent learns to evolve from $\mathcal{X}_0$ to $\mathcal{X}_t$ to suppress hub-induced distractors and retain diagnostically consistent references.

1. State $\mathcal{S}$. At time step t, the state $s_t \in \mathcal{S}$ summarizes the current reference set $\mathcal{X}_t$ with respect to the query q and the intra-set semantic coherence. We represent s_t as a graph $s_t = (\mathcal{G}_t, \mathbf{H}_t)$, where $\mathcal{G}_t$ specifies the topology over the current references and $\mathbf{H}_t$ stores node features.

1) Graph topology. The topology $\mathcal{G}_t$ is instantiated by a semantic-consistency adjacency matrix defined within $\mathcal{X}_t$. Because raw disease labels are often coarse, for each disease label $y \in \mathcal{Y}$ we perform clinical concept expansion using a large language model (GPT-5.1), producing a set of fine-grained clinical tags $\mathrm{Tags}(y)$ (e.g., "bone-spicule pigmentation" and "cotton-wool spots"). For each reference case x, we attach its tag text $\mathrm{Tags}(y_x)$ and encode it with the frozen text encoder of the vision–language model to obtain a semantic embedding u_x. We then build the adjacency matrix $\mathbf{A}_t \in \mathbb{R}^{|\mathcal{X}_t| \times |\mathcal{X}_t|}$ by pairwise cosine similarity:

$$(\mathbf{A}_t)_{ij} := \frac{u_{x_i}^\top u_{x_j}}{\|u_{x_i}\|_2 \, \|u_{x_j}\|_2}, \quad x_i, x_j \in \mathcal{X}_t. \tag{1}$$

This construction links reference cases that share consistent expanded clinical semantics.

2) Node features. The matrix $\mathbf{H}_t \in \mathbb{R}^{|\mathcal{X}_t| \times D}$ stacks node features $h(x) \in \mathbf{H}_t$, defined as:

$$h(x) = [\, v_x \| \phi_{\mathrm{stat}}(x, \mathcal{X}_t, q) \,], \tag{2}$$

where $\|$ denotes concatenation. Descriptor $\phi_{\mathrm{stat}} \in \mathbb{R}^8$ stacks four base metrics and their mean deviations, capturing similarity between x and: (i) query q (visual), (ii) the initial rank, and set $\mathcal{X}_t$ regarding (iii) clinical text and (iv) disease categories.

2. Action $\mathcal{A}$. We use a discrete action space $\mathcal{A} = \mathcal{A}_{\mathrm{del}} \cup \mathcal{A}_{\mathrm{ins}} \cup \{a_{\mathrm{term}}\}$. *Deletion.* For each reference case $x \in \mathcal{X}_t$, a deletion action $a_{\mathrm{del}}(x) \in \mathcal{A}_{\mathrm{del}}$ removes x from $\mathcal{X}_t$, subject to a minimum-size constraint $|\mathcal{X}_t| \geq min\_size$. *Insertion.* An insertion action $a_{\mathrm{ins}} \in \mathcal{A}_{\mathrm{ins}}$ expands the set by adding one case from a prefetched candidate buffer $\mathcal{B} \subset \mathcal{D}$. To avoid searching the full datastore at each step, the environment deterministically inserts the most query-similar candidate not already selected: $x_{\mathrm{ins}} = \arg\max_{x \in \mathcal{B} \setminus \mathcal{X}_t} \mathrm{sim}(q, x)$. This design lets the agent learn *when* to expand the evidence set, while a fixed similarity prior determines *what* to insert, thereby reducing the effective action dimensionality. *Termination.* The agent stops the refinement process by selecting a_{term} when $\mathcal{X}_t$ is deemed sufficiently coherent.

3. Transition $\mathcal{T}$. Given action a_t, transition $\mathcal{T}$ updates the reference set as

$$\mathcal{X}_{t+1} = \begin{cases} \mathcal{X}_t \setminus \{x\}, & a_t \in \mathcal{A}_{\mathrm{del}}, \\ \mathcal{X}_t \cup \{x_{ins}\}, & a_t \in \mathcal{A}_{\mathrm{ins}}, \\ \mathcal{X}_t, & a_t = a_{\mathrm{term}}, \end{cases} \tag{3}$$

and reconstructs the graph $\mathcal{G}_{t+1}$ and node features $\mathbf{H}_{t+1}$ accordingly.

4. Homogeneity-aware Reward $\mathcal{R}$. We design a *homogeneity-aware reward* to encourage the agent to evolve a diagnostically coherent reference set. The total return of a trajectory $\tau = \{(s_t, a_t)\}_{t=0}^{T}$ is

$$R(\tau) = R_{\text{traj}}(\mathcal{X}_T) + \sum_{t=0}^{T-1} r_{\text{step}}(s_t, a_t), \tag{4}$$

where R_{traj} evaluates the quality of the final reference set and r_{step} provides immediate feedback for each evolving operation.

1) Trajectory-Level Reward. Let $\hat{y}(\mathcal{X})$ be the final diagnosis conditioned on reference set $\mathcal{X}$. To guide the policy toward label consistency and semantic homogeneity, we define the label purity $\text{Pur}(\mathcal{X})$ and the mean pairwise text-density $\text{Den}(\mathcal{X})$ as:

$$\text{Pur}(\mathcal{X}) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \mathbb{I}[y_x = y^\star], \quad \text{Den}(\mathcal{X}) = \frac{2}{|\mathcal{X}|(|\mathcal{X}|-1)} \sum_{i<j} \cos(u_{x_i}, u_{x_j}), \tag{5}$$

where $y^\star$ is the ground-truth label and u_x is the text embedding of clinical tags for x. We then construct the trajectory reward as:

$$R_{\text{traj}}(\mathcal{X}_T) = \underbrace{\alpha\, \mathbb{I}[\hat{y}(\mathcal{X}_T) = y^\star]}_{R_{\text{acc}}} + \underbrace{\beta\Big(\text{Pur}(\mathcal{X}_T) - \text{Pur}(\mathcal{X}_0)\Big)}_{R_{\text{purity}}} + \underbrace{\gamma\, \text{Den}(\mathcal{X}_T)}_{R_{\text{density}}}, \tag{6}$$

which rewards correct prediction, purity gain over the initial set $\mathcal{X}_0$, and intra-set semantic density.

2) Step-Level Reward. For each action, we provide positive-only feedback:

$$r_{\text{step}}(s_t, a_t) = \underbrace{\eta_{\text{ins}}\, \mathbb{I}[a_t \in \mathcal{A}_{\text{ins}}]\, \mathbb{I}[y_{x_{\text{ins}}} = y^\star]}_{r_{\text{ins}}} + \underbrace{\eta_{\text{del}}\, \mathbb{I}[a_t \in \mathcal{A}_{\text{del}}]\, \mathbb{I}[y_{x_{\text{del}}} \neq y^\star]}_{r_{\text{del}}}. \tag{7}$$

We adopt a zero-penalty principle by assigning 0 reward to incorrect insertions/deletions, avoiding negative shaping that can hinder early exploration.

2.2 Relational Policy Learning via Group-Relative Optimization

With the MDP defined above, we learn a retrieval evolving policy $\pi_\theta(a_t \mid s_t)$ that iteratively updates the reference set $\mathcal{X}_t$ toward a diagnostically consistent and semantically coherent subset. Instead of treating instances independently, we instantiate π_θ as a relational reasoner over the evidence graph and optimize it using a group-relative objective to stabilize learning without a value network.

Graph Policy Network. Given the state graph $s_t = (\mathcal{G}_t, \mathbf{H}_t)$ with adjacency matrix $\mathbf{A}_t$, we adopt a Graph Convolutional Network (GCN) [8] to propagate consistency signals among references. With self-loops $\hat{\mathbf{A}}_t = \mathbf{A}_t + \mathbf{I}$ and degree matrix $\hat{\mathbf{D}}_t$, the l-th layer updates node features as

$$\mathbf{H}_t^{(l+1)} = \sigma\Big(\hat{\mathbf{D}}_t^{-\frac{1}{2}} \hat{\mathbf{A}}_t \hat{\mathbf{D}}_t^{-\frac{1}{2}} \mathbf{H}_t^{(l)} \mathbf{W}^{(l)}\Big), \quad \mathbf{H}_t^{(0)} = \mathbf{H}_t, \tag{8}$$

where $\sigma(\cdot)$ denotes a nonlinear activation. After L layers ($L = 2$), we parameterize the action distribution by combining node-level deletion scores with graph-level operation scores:

$$\pi_\theta(a_t \mid s_t) = \mathrm{Softmax}\Big(\Big[\{\mathrm{MLP}_{\mathrm{local}}(h_x^{(L)})\}_{x\in\mathcal{X}_t},\ \mathrm{MLP}_{\mathrm{global}}(\mathrm{Pool}(\mathbf{H}_t^{(L)}))\Big]\Big)_{a_t}, \tag{9}$$

where $\mathrm{MLP}_{\mathrm{local}}$ and $\mathrm{MLP}_{\mathrm{global}}$ are two MLP heads and Pool is mean pooling.

GRPO Optimization. We optimize π_θ with Group Relative Policy Optimization (GRPO) [6], which eliminates the need for a learned critic by using group statistics as a dynamic baseline. For each query q, we sample a group of trajectories $\mathcal{T}_G = \{\tau_1, \ldots, \tau_G\}$ from $\pi_{\theta_{\mathrm{old}}}$, compute their returns $\{R(\tau_i)\}_{i=1}^G$, and form the normalized advantage

$$\hat{A}_i = \frac{R(\tau_i) - \mathrm{Mean}(\{R(\tau_j)\}_{j=1}^G)}{\mathrm{Std}(\{R(\tau_j)\}_{j=1}^G) + \epsilon}. \tag{10}$$

We then update the policy by maximizing

$$\mathcal{J}(\theta) = \mathbb{E}_q\left[\frac{1}{G}\sum_{i=1}^{G}\left(\frac{\pi_\theta(\tau_i)}{\pi_{\theta_{\mathrm{old}}}(\tau_i)}\hat{A}_i - \beta\,\mathbb{D}_{\mathrm{KL}}(\pi_\theta(\cdot \mid q)\,\|\,\pi_{\mathrm{ref}}(\cdot \mid q))\right)\right], \tag{11}$$

where π_{ref} is a reference policy and the KL term regularizes updates.

2.3 Evidence-Conditioned Prediction

After the agent terminates, we obtain the evolved reference set $\mathcal{X}_T$. We then use majority voting over the labels of evolved references for diagnosis:

$$\hat{y} = \arg\max_{c\in\mathcal{Y}} \sum_{x\in\mathcal{X}_T} \mathbb{I}(y_x = c). \tag{12}$$

3 Experiments

3.1 Experimental Setup

Datasets. Public fundus benchmarks are increasingly saturated for rare-disease diagnosis because recent retinal foundation models have been trained extensively on them [15,16], making generalization hard to assess. We therefore evaluate on the *Retina Image Bank* (RIB) [1], a public and continuously updated clinical repository that is less coupled to standard benchmarks. From RIB, we collect 30,662 images and curate a single-label fundus dataset to reduce comorbidity confounding. We construct **Rare-20** (20 rare diseases; $15 < N \leq 75$ images/class) and **Retina-31** by adding 11 common diseases ($N > 200$ images/class). We use a 70/15/15 train/val/test split; for retrieval-based methods, the retrieval corpus is restricted to the training split to prevent test leakage.

Table 1: Comparison with foundation models under zero-shot and linear probing settings. Performance is reported as mean±std for linear probing evaluation.

Method	Rare-20			Retina-31		
	Acc[%]	F1-score[%]	Sensitivity[%]	ACC[%]	F1-score[%]	Sensitivity[%]
Zero-shot Evaluation:						
BiomedCLIP [21]	2.91	0.29	3.75	0.39	0.10	2.49
MedCLIP [19]	4.85	3.77	7.08	1.08	0.65	3.58
EyeCLIP [15]	23.95	10.10	11.50	21.60	12.79	13.13
FLAIR [16]	20.39	15.86	21.59	39.00	24.15	31.08
RetiZero [18]	25.24	26.37	29.35	11.53	12.86	19.34
Linear Probing Evaluation:						
EyeCLIP [15]	27.64 ± 5.01	14.48 ± 2.10	18.90 ± 3.74	30.49 ± 1.72	17.28 ± 1.33	19.15 ± 3.96
FLAIR [16]	27.51 ± 1.12	8.54 ± 0.36	12.60 ± 0.50	55.16 ± 0.40	22.81 ± 0.22	24.79 ± 0.23
RetiZero [18]	26.86 ± 2.02	10.15 ± 0.84	13.05 ± 1.05	36.62 ± 0.31	13.86 ± 0.28	15.56 ± 0.13
Evo-RAD (Ours)	**46.28 ± 0.56**	**40.99 ± 0.53**	**42.43 ± 0.21**	**65.33 ± 0.28**	**51.45 ± 0.43**	**49.53 ± 0.52**

Table 2: Comparison with retrieval and PEFT methods (mean±std). RetiZero is the foundation model. Params is number of trainable parameters on Retina-31.

METHOD	Params	Rare-20			Retina-31		
		Acc[%]	F1-score[%]	Sensitivity[%]	Acc[%]	F1-score[%]	Sensitivity[%]
Zero-shot	0	25.24	26.37	29.35	11.53	12.86	19.34
Static Retrieval	0	42.72	35.13	36.64	60.41	43.93	41.21
CoOp [22]	3K	29.13 ± 0.35	25.31 ± 0.13	27.52 ± 0.75	20.95 ± 1.05	15.34 ± 0.72	17.44 ± 0.41
CLIP-Adapter [5]	263K	35.40 ± 0.30	23.39 ± 0.50	26.95 ± 0.28	20.70 ± 0.11	14.44 ± 0.31	18.71 ± 0.25
Tip-Adapter [20]	0	37.86 ± 0.00	24.45 ± 0.00	27.00 ± 0.00	22.97 ± 0.00	15.57 ± 0.00	19.12 ± 0.00
RAC [11]	23.8K	39.48 ± 1.48	35.39 ± 2.55	40.06 ± 2.36	51.81 ± 0.78	36.71 ± 1.37	45.60 ± 1.71
TDA [7]	0	30.10 ± 1.12	25.73 ± 0.13	30.08 ± 0.11	15.62 ± 0.15	14.33 ± 0.32	20.72 ± 0.52
XCoOp [2]	12.3K	36.57 ± 0.97	31.75 ± 0.91	31.97 ± 1.19	44.28 ± 1.62	25.53 ± 1.25	25.08 ± 1.09
DPC [10]	24.6K	30.74 ± 1.48	25.66 ± 0.74	27.02 ± 1.28	34.77 ± 0.31	19.79 ± 3.03	19.32 ± 2.39
BiomedCoOp [9]	12.3K	42.72 ± 0.97	36.89 ± 1.89	37.05 ± 1.74	46.20 ± 0.65	26.03 ± 0.57	25.58 ± 0.51
Evo-RAD (Ours)	**116.4K**	**46.28 ± 0.56**	**40.99 ± 0.53**	**42.43 ± 0.21**	**65.33 ± 0.28**	**51.45 ± 0.43**	**49.53 ± 0.52**

Rare-20 contains 524/103/103 train/val/test images, and Retina-31 contains 4,737/1,001/1,023.

Baselines and Implementation Details. We compare Evo-RAD against three groups of methods. (1) *Foundation models*: general medical VLMs (BiomedCLIP [21], MedCLIP [19]) and retina-specific models (EyeCLIP [15], FLAIR [16], RetiZero [18]). (2) *PEFT methods*: CoOp [22], CLIP-Adapter [5], XCoOp [2], BiomedCoOp [9], TDA [7], Tip-Adapter [20], and DPC [10]. (3) *Retrieval-based methods*: Static Retrieval and RAC [11]. We implement Evo-RAD atop RetiZero [18] with a candidate buffer $|\mathcal{B}| = 100$ and initial retrieval size $K = 8$. The policy allows at most 10 actions min_size=2 and is trained via GRPO with $G = 8$. We report 3-seed averages using an NVIDIA RTX 4090 (24GB).

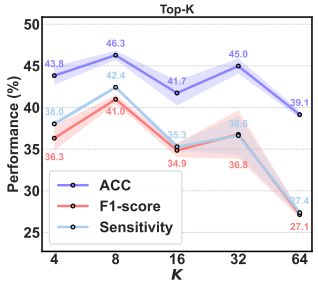
3.2 Main Results

Comparison with Foundation Models. Table 1 shows that Evo-RAD consistently outperforms both general medical VLMs and retina-specific foundation models. On Rare-20, Evo-RAD achieves 46.28% ACC / 40.99% macro-F1 / 42.43% sensitivity, outperforming the strongest baseline (RetiZero zero-shot) by

Table 3: Ablation study on state representation, step-wise rewards, and terminal rewards.

Setting	Acc [%]	F1-score [%]	Sensitivity [%]
Full (all components)	**46.28 ± 0.56**	**40.99 ± 0.53**	**42.43 ± 0.21**
(a) Ablation on State Representation:			
− *Mean deviations*	42.74 ± 0.42	35.46 ± 0.54	36.21 ± 0.63
− *Base metrics*	44.56 ± 0.43	36.21 ± 0.58	38.41 ± 0.44
(b) Ablation on Trajectory-Level Rewards:			
− R_{acc}	43.68 ± 0.63	35.82 ± 0.68	36.40 ± 0.63
− R_{purity}	40.14 ± 1.23	33.11 ± 1.92	33.30 ± 0.90
− R_{density}	43.89 ± 1.60	36.06 ± 1.28	36.80 ± 0.94
(c) Ablation on Step-Level Rewards:			
− r_{ins}	39.69 ± 0.84	34.84 ± 0.61	34.85 ± 0.42
− r_{del}	40.61 ± 0.64	35.04 ± 0.54	35.65 ± 0.36

Fig. 2: Impact of inital retrieval size K on Rare-20.



a wide margin. On Retina-31, Evo-RAD reaches 65.33% ACC / 51.45% macro-F1, substantially exceeding linear-probing results of FLAIR and RetiZero, indicating stronger generalization across both rare and common diseases.

Comparison with Retrieval and PEFT Methods. As reported in Table 2, Evo-RAD improves over retrieval, and PEFT approaches built on the same foundation model. Notably, compared with *Static Retrieval*, Evo-RAD yields consistent gains on both Rare-20 and Retina-31, and it also outperforms RAC and prompt tuning baselines, demonstrating the advantage of the self-evolving agentic retrieval framework over fixed retrieval or lightweight prompt tuning.

Efficiency Analysis. Evo-RAD contains only 116.4K trainable parameters (Table 2), yet achieves substantially better performance. This indicates the gains mainly come from evolving retrieval rather than heavy fine-tuning, making Evo-RAD efficient to train and deploy.

3.3 Ablation Study

Ablation on State Representation. Table 3(a) shows that removing either component reduces ACC. Dropping the *mean deviations* (set-wise consistency) lowers ACC to 42.74%, and removing *base metrics* (individual similarity) reduces it to 44.56%. This confirms the synergy between individual relevance and group-level consensus.

Ablation on Trajectory-Level Rewards. As shown in Table 3(b), removing any trajectory-level reward reduces ACC: dropping R_{acc} lowers ACC to 43.68%, while among auxiliary terms R_{purity} is most critical (40.14% ACC without it). Removing R_{density} also degrades ACC to 43.89%, suggesting that a well-connected reference set benefits evidence evolution.

Ablation on Step-Level Rewards. As shown in Table 3(c), removing step-wise shaping rewards hurts ACC, dropping to 39.69% without r_{ins} and 40.61% without r_{del}, confirming the value of dense feedback during evidence evolution.

Ablation on Initial Reference Set Size K. As shown in Fig. 2, performance peaks at $K = 8$. Smaller K provides insufficient evidence, whereas larger K introduces more redundant/noisy references. We therefore use $K = 8$ by default.

4 Conclusion

We presented Evo-RAD, a self-evolving agentic retrieval framework for rare retinal disease diagnosis that formulates evidence refinement as sequential DELETE/ INSERT/ TERMINATE decisions and learns an evidence-evolution policy via GRPO with a homogeneity-aware reward. By explicitly suppressing hub-driven distractors and consolidating clinically coherent support sets before prediction, Evo-RAD delivers consistent improvements over foundation-model baselines, standard retrieval, and PEFT variants. Future work will extend Evo-RAD to multi-label and co-morbidity settings, enabling evidence evolution under realistic mixed-pathology supervision.

Acknowledgments. This work was partially supported by PolyU Undergraduate Research and Innovation Scheme (No. P0058439), RGC Collaborative Research Fund (No. C5055-24G), the Start-up Fund of The Hong Kong Polytechnic University (No. P0045999), the Seed Fund of the Research Institute for Smart Ageing (No. P0050946), and Tsinghua-PolyU Joint Research Initiative Fund (No. P0056509), and PolyU UGC funding (No. P0053716 and P0058848).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. American Society of Retina Specialists (ASRS): Retina image bank. https://www.asrs.org/clinical/retina-image-bank (2026), accessed: 2026-02-25. An open-access library of nearly 30,000 downloadable retina images.
2. Bie, Y., Luo, L., Chen, Z., Chen, H.: Xcoop: Explainable prompt learning for computer-aided diagnosis via concept-guided context optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 773–783 (2024)
3. Croskerry, P.: A universal model of diagnostic reasoning. Academic medicine **84**(8), 1022–1028 (2009)
4. Du, Y., Yu, N., Wang, S.: Medical knowledge intervention prompt tuning for medical image classification. IEEE Transactions on Medical Imaging (2025)
5. Gao, P., Geng, S., Zhang, R., Ma, T., Fang, R., Zhang, Y., Li, H., Qiao, Y.: Clip-adapter: Better vision-language models with feature adapters. International journal of computer vision **132**(2), 581–595 (2024)
6. Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., et al.: Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. Nature **645**(8081), 633–638 (2025)
7. Karmanov, A., Guan, D., Lu, S., El Saddik, A., Xing, E.: Efficient test-time adaptation of vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14162–14171 (2024)
8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (2017)
9. Koleilat, T., Asgariandehkordi, H., Rivaz, H., Xiao, Y.: Biomedcoop: Learning to prompt for biomedical vision-language models. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14766–14776 (2025)

10. Li, H., Wang, L., Wang, C., Jiang, J., Peng, Y., Long, G.: Dpc: Dual-prompt collaboration for tuning vision-language models. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 25623–25632 (2025)
11. Long, A., Yin, W., Ajanthan, T., Nguyen, V., Purkait, P., Garg, R., Blair, A., Shen, C., van den Hengel, A.: Retrieval augmented classification for long-tail visual recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6959–6969 (2022)
12. Luo, Y., Li, W., Chen, C., Li, X., Liu, T., Niu, T., Yuan, Y.: Llm-guided decoupled probabilistic prompt for continual learning in medical image diagnosis. IEEE Transactions on Medical Imaging (2025)
13. Phillips, C., Parkinson, A., Namsrai, T., Chalmers, A., Dews, C., Gregory, D., Kelly, E., Lowe, C., Desborough, J.: Time to diagnosis for a rare disease: managing medical uncertainty. a qualitative study. Orphanet journal of rare diseases **19**(1), 297 (2024)
14. Radovanović, M., Nanopoulos, A., Ivanović, M.: Hubs in space: Popular nearest neighbors in high-dimensional data. J. Mach. Learn. Res. **11**, 2487–2531 (2010)
15. Shi, D., Zhang, W., Yang, J., Huang, S., Chen, X., Xu, P., Jin, K., Lin, S., Wei, J., Yusufu, M., Liu, S., Zhang, Q., Ge, Z., Xu, X., He, M.: A multimodal visual–language foundation model for computational ophthalmology. npj Digital Medicine **8**(1), 381 (6 2025)
16. Silva-Rodríguez, J., Chakor, H., Kobbi, R., Dolz, J., Ben Ayed, I.: A foundation language-image model of the retina (flair): encoding expert knowledge in text supervision. Medical Image Analysis **99**, 103357 (2025)
17. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction, vol. 1. MIT press Cambridge (1998)
18. Wang, M., Lin, T., Lin, A., Yu, K., Peng, Y., Wang, L., Chen, C., Zou, K., Liang, H., Chen, M., Yao, X., Zhang, M., Huang, B., Zheng, C., Zhang, P., Chen, W., Luo, Y., Chen, Y., Xia, H., Fu, H.: Enhancing diagnostic accuracy in rare and common fundus diseases with a knowledge-rich vision-language model. Nature Communications **16**, 5528 (07 2025)
19. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 3876–3887 (2022)
20. Zhang, R., Zhang, W., Fang, R., Gao, P., Li, K., Dai, J., Qiao, Y., Li, H.: Tip-adapter: Training-free adaption of clip for few-shot classification. In: European Conference on Computer Vision (ECCV). pp. 493–510 (2022)
21. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., Wong, C., Tupini, A., Wang, Y., Mazzola, M., Shukla, S., Liden, L., Gao, J., Crabtree, A., Piening, B., Bifulco, C., Lungren, M.P., Naumann, T., Wang, S., Poon, H.: A multimodal biomedical foundation model trained from fifteen million image–text pairs. NEJM AI **2**(1), AIoa2400640 (2025)
22. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International journal of computer vision **130**(9), 2337–2348 (2022)