Skin-R1: Clinical Knowledge-Guided Dermatological Diagnosis Using Vision-Language Models

Zehao Liu[1], Weijieying Ren[2], Jipeng Zhang[3], Tianxiang Zhao[1], Jingxi Zhu[1], Xiaoting Li[1], and Vasant G Honavar[1]

[1] Pennsylvania State University, University Park, PA 16802, USA
{zml5418,jqz5678,vuh14}@psu.edu, {txiangchao,xiaotili0123}@gmail.com
[2] Stanford University, Stanford, CA 94305, USA
wjyren@stanford.edu
[3] Hong Kong University of Science and Technology, Hong Kong, China
jzhanggr@connect.ust.hk

Abstract. Vision–language models (VLMs) have recently shown promise for assisting clinical reasoning in dermatological diagnosis. However, their trustworthiness and clinical utility remain limited by three key challenges: heterogeneous datasets with inconsistent diagnostic labels and concept annotations, the lack of grounded diagnostic rationales for reliable reasoning supervision, and limited scalability when transferring knowledge from small, densely annotated datasets to large collections with sparse labels.

To address these challenges, we propose Skin-R1, a dermatology-oriented VLM that integrates textbook-grounded clinical reasoning supervision with reinforcement learning (RL) to improve the accuracy and robustness of diagnostic prediction. First, we construct a textbook-based reasoning generator that synthesizes hierarchy-aware and differential-diagnosis (DDx) diagnostic trajectories derived from authoritative dermatology knowledge. Second, these trajectories are used for supervised fine-tuning (SFT), establishing a clinically grounded reasoning foundation for the model. Finally, we introduce an RL training framework that incorporates the hierarchical structure of dermatological diseases into the reward design, enabling the model to generalize grounded diagnostic reasoning to large-scale datasets with sparse annotations.

Extensive experiments across multiple dermatology benchmarks demonstrate that Skin-R1 consistently improves diagnostic accuracy and robustness compared to state-of-the-art Med-VLM baselines. Ablation studies further highlight the critical role of grounded reasoning supervision introduced during the SFT stage.

Code and model weight are available at https://github.com/1593191569/Skin-R1.

Keywords: Dermatology · Medical AI · Vision–Language Models · Reinforcement Learning · GRPO

Table 1: Comparison of representative Medical VLMs in terms of key diagnostic reasoning abilities. **Grounded Rationale** indicates whether the model is designed to produce rationales grounded in authoritative clinical sources; **DDx Awareness** reflects awareness of differential diagnoses through comparative reasoning; **Hierarchical Awareness** captures understanding of disease taxonomy and multi-level relations; and **Sparse-Supervision Learning** measures the ability to learn and generalize from sparsely annotated data.

Type	Model	Grounded Rationale	DDx Awareness	Hierarchical Awareness	Sparse-Supervision Learning
RL-based	Med-R1 [22]	✗	✗	✗	✓
	MedVLM-R1 [34]	✗	✗	✗	✓
	MedCCO [39]	✗	✗	✗	✓
	RARL [35]	RL reward	✗	✗	✓
	[65]	RL reward	✗	✗	✓
SFT-based	SkinVL-PubMM [59]	grounding SFT	✗	✗	✗
	OmniV-Med [17]	grounding SFT	✓	✗	✗
	LLaVA-Med [24]	distillation	✓	✗	✗
	HuatuoGPT-Vision [4]	distillation	✓	✗	✗
Hybrid	MedGemma [40]	grounding SFT, distillation, RL reward	✓	✗	✓
	Skin-R1	**grounding SFT**	✓	✓	✓

1 Introduction

Dermatological diagnosis is fundamentally a visual reasoning task: clinicians interpret subtle visual patterns, compare them against hierarchical disease taxonomies, and perform differential diagnosis (DDx) to distinguish visually similar conditions. Recent advances in multimodal vision–language models (VLMs) suggest that AI systems could assist this process by jointly interpreting medical images and clinical knowledge. However, enabling such systems to perform reliable and scalable diagnostic reasoning remains an open challenge.

Each year, millions of people worldwide experience delayed or incorrect diagnosis of skin conditions such as melanoma, eczema, and psoriasis—often due to limited access to dermatologists [9, 11, 38] or subjective interpretation of visual symptoms. Advances in computer vision, deep learning, and multimodal learning now offer the possibility of AI-assisted systems capable of analyzing dermoscopic and clinical images with expert-level precision, supporting clinicians in triage, screening, and treatment planning. By combining clinical expertise with AI, dermatology-oriented vision models have the potential to democratize access to high-quality care and reduce the global burden of preventable skin disease [6, 29, 58, 62].

Recent multimodal vision foundation models and medical vision–language models (Med-VLMs) [13, 20, 21, 52, 54, 55] have demonstrated promising performance in dermatological image understanding and diagnosis. However, existing approaches still face important limitations. For example, models such as SkinGPT-4 [62] rely heavily on labor-intensive human annotations and often lack clinically grounded diagnostic rationales.

Building a reliable dermatology vision–language model (VLM) presents three major challenges.

1. **Data heterogeneity.** Real-world dermatology datasets vary widely in their diagnostic ontologies and concept vocabularies. For example, SkinCon [7]

provides annotations for 29 dermatological concepts (e.g., *erythema*, *plaque*) but only two coarse diagnostic labels (benign vs. malignant). In contrast, DermNet [8] contains images spanning more than 600 diseases but lacks structured concept annotations. Such discrepancies lead to inconsistent supervision and noisy training signals.

2. **Lack of expert-like reasoning supervision.** Accurate dermatological diagnosis requires structured reasoning that respects hierarchical disease taxonomies and supports differential diagnosis (DDx). However, existing datasets rarely provide supervision that captures clinically grounded diagnostic reasoning processes.
3. **Limited scalability of training paradigms.** Models trained on small datasets with dense annotations often struggle to generalize to large datasets with sparse labels, making it difficult to transfer grounded diagnostic reasoning to large-scale heterogeneous data.

These challenges reflect a deeper issue: dermatological diagnosis is fundamentally a reasoning problem over a hierarchical disease space, where visually similar conditions must be distinguished through structured clinical reasoning and differential diagnosis. Existing training paradigms struggle to capture such reasoning processes while remaining scalable to large heterogeneous datasets.

Motivated by this observation, a key open question is how to develop dermatology VLMs capable of performing *expert-like clinical reasoning* while remaining scalable to large heterogeneous datasets. Existing approaches typically rely either on small datasets with dense annotations that provide limited scale, or on large weakly labeled datasets that lack reliable reasoning supervision. Bridging this gap requires a framework that can first acquire grounded clinical reasoning patterns from reliable sources and then effectively generalize them across diverse datasets.

To address this problem, we propose Skin-R1, which, to the best of our knowledge, is the first dermatology-oriented VLM that integrates *textbook-grounded clinical reasoning* with reinforcement learning–based adaptation. Our framework unifies expert-aligned reasoning and scalable knowledge transfer within a single end-to-end training pipeline. Figure 1 illustrates the overall architecture and training procedure of Skin-R1.

First, we construct a grounded dataset of hierarchy-aware and differential-diagnosis (DDx)-informed diagnostic trajectories derived from authoritative dermatology textbooks. We then perform supervised fine-tuning (SFT) on these trajectories to establish a clinically grounded reasoning foundation. Finally, we scale this reasoning capability to large datasets with sparse annotations through a reinforcement learning (RL) framework guided by hierarchical diagnostic accuracy and structured output regularization rewards, encouraging the model to generalize grounded diagnostic reasoning across heterogeneous data sources.

The main contributions of this paper are summarized as follows:

1. We introduce Skin-R1, a novel training paradigm for dermatology vision–language models that integrates textbook-grounded clinical reasoning with

reinforcement learning to improve diagnostic accuracy and robustness across heterogeneous datasets.
2. We design a textbook-grounded trajectory generation framework that synthesizes hierarchy-aware and differential-diagnosis (DDx)-informed diagnostic reasoning traces, providing scalable supervision for clinically consistent reasoning.
3. We develop a hierarchical reward design that incorporates disease taxonomy structure into reinforcement learning, enabling grounded reasoning patterns learned from dense data to generalize effectively to large-scale sparsely annotated datasets.
4. We conduct extensive experiments across multiple dermatology benchmarks demonstrating that Skin-R1 consistently outperforms state-of-the-art baselines in diagnostic accuracy and robustness.

More broadly, our results suggest that grounding vision-language models in expert-level reasoning supervision may provide a promising path toward increasing the reliability of AI systems for medical applications.

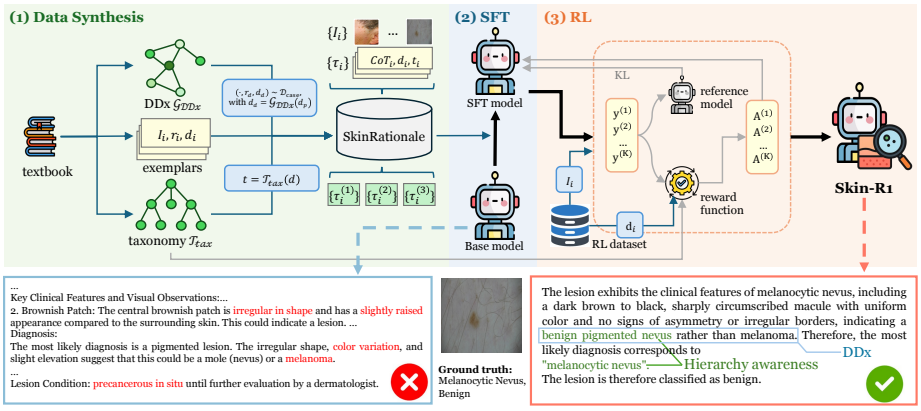

Fig. 1: Top: Overview of the proposed training framework, illustrating (1) the synthesis of the SkinRationale dataset, (2) the SFT stage performed on SkinRationale, and (3) the GRPO-based RL stage.
Bottom: Two representative cases comparing diagnostic responses from Skin-R1 and baseline models. The response from Skin-R1 is concise, accurate, and aligned with expert diagnostic reasoning. Red text highlights incorrect or hallucinated content; Green text denotes *hierarchical-aware* reasoning; and content enclosed within the blue box indicates the *DDx comparison* derived from prior differential diagnostic reasoning.

2 Related Work

Medical Vision–Language Models. Vision–language models (VLMs) [1, 2, 27, 57, 63] extend large language models (LLMs) [32, 43, 47] to multimodal

reasoning by enabling joint processing of visual and textual inputs. Medical VLMs (Med-VLMs) [37, 46, 61] adapt these architectures for clinical applications by developing domain-specific visual encoders, integrating them with LLMs, and aligning multimodal representations across diverse biomedical datasets.

Early efforts focused on improving visual understanding in specialized medical domains through medical foundation models, including radiology [50], pathology [30], and dermatology [20, 52–54]. Subsequent approaches introduced encoder–decoder Med-VLM architectures that combine medical visual encoders with language models. For example, XrayGPT [45] integrates MedCLIP [50] with a fine-tuned LLM, while LLaVA-Med [24] improves multimodal alignment using PubMed image–text pairs. BiomedGPT [31] adopts a BERT–GPT hybrid architecture trained on biomedical corpora, and HuatuoGPT-Vision [4] scales this paradigm using the large PubMedVision dataset. More recently, HealthGPT [25] unifies medical image understanding and generation within a single autoregressive framework, and SkinGPT-4 [62] applies instruction tuning to improve dermatological reasoning.

Despite these advances, most Med-VLMs rely on large volumes of high-quality annotated data, limiting their applicability in domains where expert annotations are scarce. In contrast, Skin-R1 learns grounded diagnostic reasoning from a small set of high-quality expert-derived trajectories and then scales this capability using reinforcement learning.

RL for Reasoning in Med-VLMs. Recent work has demonstrated that reinforcement learning (RL) can induce reasoning abilities in large language and vision–language models [3, 10, 15, 51, 60]. Methods based on RL with verifiable rewards (RLVR), such as DeepSeek-R1 [10], employ Group Relative Policy Optimization (GRPO) to encourage structured reasoning behaviors without explicit chain-of-thought (CoT) supervision.

RL has also been applied to improve reasoning in Med-VLMs for diagnostic tasks [22, 28, 34, 35, 39, 65]. Because high-quality annotated medical reasoning data are scarce, many of these approaches bypass supervised fine-tuning (SFT) and instead apply RL directly using GRPO-style objectives. This RL-only paradigm has shown promising improvements in medical visual question answering and cross-domain generalization.

For example, MedVLM-R1 [34] demonstrates that GRPO can induce emergent chain-of-thought reasoning without explicit reasoning supervision. Similarly, Med-R1 [22] reports that models trained without explicit reasoning supervision can sometimes outperform those trained with CoT-style supervision, suggesting that conventional SFT may introduce *memorization shortcuts*. However, we argue that this phenomenon arises because existing approaches lack clinically grounded reasoning supervision, often leading to heuristic or hallucinated rationales [16, 23].

Unlike prior work, Skin-R1 first establishes a grounded diagnostic reasoning foundation using textbook-derived diagnostic trajectories and then employs RL to amplify and generalize this reasoning capability. Ablation experiments

(Sec. 4.2) confirm that this combination of grounded supervision and RL significantly improves diagnostic reasoning performance.

3 Methods

Pipeline Overview. The top panel of Figure 1 illustrates the overall training pipeline of Skin-R1. Our framework consists of three stages. First, we construct **SkinRationale**, a textbook-derived dataset that captures hierarchy-aware and differential diagnosis (DDx)-informed reasoning trajectories grounded in authoritative dermatology knowledge. Second, we perform supervised fine-tuning (SFT) on these trajectories to initialize the model with clinically consistent diagnostic reasoning. Finally, we apply reinforcement learning (RL) using Group-Relative Policy Optimization (GRPO) to generalize this reasoning ability to large-scale dermatology datasets with sparse annotations. Through this pipeline, the model learns to produce structured diagnostic reasoning while maintaining robustness across heterogeneous clinical data distributions.

In the following sections, we first describe the construction of SkinRationale, and then present the two-stage training strategy consisting of SFT and GRPO.

3.1 Training Data for Grounded Reasoning

Reliable dermatological diagnosis requires resources that capture multi-step clinical reasoning. However, existing dermatology datasets primarily focus on visual classification or lesion segmentation and rarely include diagnostic rationales linking visual concepts to clinical conclusions. To address this limitation, we introduce **SkinRationale**, a textbook-derived dataset designed to provide clinically grounded reasoning supervision.

Extracting Textbook-Grounded Knowledge We extract three complementary components from the dermatology textbook of [18]: (1) Diagnostic exemplars $\mathcal{D} = \{(I_i, r_i, d_i)\}_{i=1}^{N}$, each represented as a triplet (I_i, r_i, d_i) consisting of a clinical image I_i, a textual rationale r_i, and a diagnostic label d_i, (2) A differential diagnosis graph $\mathcal{G}_{\text{DDx}}$, where edges connect clinically confusable diseases; and (3) A disease taxonomy $\mathcal{T}_{\text{tax}}$, where parent–child relations represent hierarchical disease categories.

In total, we collect 220 diagnostic exemplars, construct a DDx graph with 211 nodes and 245 edges (covering 61.36% of the collected exemplars), and build a disease taxonomy containing 458 nodes and 473 edges.

The knowledge extraction pipeline largely follows the automated procedure proposed in [20]. The pipeline consists of five stages: (1) image and text extraction, (2) image clustering and filtering, (3) text pairing and filtering, (4) diagnostic rule extraction, and (5) DDx and taxonomy confirmation and refinement. Human involvement is limited to selecting representative image clusters. Further implementation details are provided in Appendix.

Constructing Reasoning Trajectories Generating reliable reasoning trajectories for cold-start supervision is challenging. Prior work often relies on large reasoning models (LRMs) such as QwQ-32B [44] to produce reasoning traces. However, such model-generated rationales may contain partial errors, potentially leading to model collapse [42] or hallucination propagation [49].

Instead, we construct reasoning trajectories from textbook-derived rationales, which provide stable and expert-curated supervision signals.

Two structural challenges arise when converting textbook knowledge into training trajectories:

1. **Label granularity inconsistency**: diagnoses may appear at different specificity levels (e.g., melanoma vs. superficial spreading melanoma); and
2. **Differential diagnosis ambiguity**: dermatological reasoning requires distinguishing visually similar diseases.

To address these issues, we introduce two complementary mechanisms.

Hierarchical diagnosis completion. We construct a dermatology diagnosis tree capturing parent–child relationships among disease entities (e.g., melanoma → superficial spreading melanoma). Each final label is augmented with its ancestral and sibling nodes. This process mitigates inconsistencies in label granularity and encourages taxonomic coherence across diagnostic levels.

Differential diagnosis reasoning enrichment. Using the DDx graph, we synthesize reasoning trajectories that explicitly describe how visually similar conditions are differentiated. These DDx-informed rationales allow the model not only to identify the correct diagnosis but also to provide interpretable justification for excluding alternatives.

The resulting reasoning trajectories are defined as

$$\{\tau_i\}_{i=1}^{K} = \{(CoT_i, d_i, t_i)\}_{i=1}^{K}, \tag{1}$$

where each trajectory contains a reasoning trace CoT_i, a diagnosis label d_i, and hierarchical diagnostic metadata t_i. In total, we construct $K = 2020$ diagnostic trajectories. Details of trajectory synthesis are provided in Appendix.

3.2 Enhancing Grounded Medical Reasoning

Problem Formulation. We formulate dermatological diagnosis as a multimodal reasoning task. Given a clinical image I and an instruction prompt p, the model generates a structured diagnostic response consisting of: (i) a reasoning trajectory CoT, (ii) a predicted disease label $\hat{\ell}$, and (iii) a lesion category $\hat{b}$ (e.g., benign, malignant, or precancerous). Formally, the policy model π_θ learns a conditional distribution

$$\pi_\theta(y \mid I, p), \tag{2}$$

where $y = (CoT, \hat{\ell}, \hat{b})$ denotes the structured output. The objective is to learn a policy that produces diagnostically correct predictions while maintaining clinically coherent reasoning trajectories.

Supervised Fine-Tuning on SkinRationale We first train the policy model π_θ using SFT on hierarchy-aware and DDx-informed diagnostic trajectories. Given a clinical image I, an instruction prompt p, and the trajectory $\tau = (x_0, x_1, \ldots, x_n)$, the training objective is

$$L = -\sum_{i=1}^{n} \log \pi_\theta(x_i \mid x_{<i}, I, p). \tag{3}$$

This objective allows the model to imitate expert diagnostic reasoning across hierarchical disease categories and establishes a robust initialization for RL optimization.

Reinforcement Learning via GRPO Although SFT instills grounded reasoning ability, it remains limited to the curated trajectories. To generalize reasoning across large-scale sparsely annotated datasets, we formulate diagnosis as a reinforcement learning problem and optimize the model using Group-Relative Policy Optimization (GRPO) [41].

Given input $x = (I, q)$, the policy generates K candidate responses

$$\{y^{(j)}\}_{j=1}^{K} = \{(CoT^{(j)}, \hat{\ell}^{(j)}, \hat{b}^{(j)})\}_{j=1}^{K}, \tag{4}$$

where $CoT^{(j)}$ denotes the reasoning trace, $\hat{\ell}^{(j)}$ the predicted disease label, and $\hat{b}^{(j)}$ the predicted lesion category. Each candidate response receives reward $r_j = R_{\text{total}}(\mathcal{P}, y^{(j)})$, where $\mathcal{P}$ denotes the ground-truth diagnostic path in the taxonomy. Rewards are normalized within groups, expressed as $A_j = (r_j - \mu)/\sigma$. The GRPO objective is defined as

$$\mathcal{L}_{\text{GRPO}}(\theta) = \mathbb{E}_x \left[\frac{1}{K} \sum_{j=1}^{K} \min(\rho_j A_j, \text{clip}(\rho_j, 1-\varepsilon, 1+\varepsilon) A_j) \right] - \beta \text{KL}(\pi_\theta(\cdot|x) \| \pi_{\text{ref}}(\cdot|x)). \tag{5}$$

Reward Function. The reward function encourages diagnostically correct and structurally consistent outputs:

$$R_{\text{total}} = R_{\text{format}} + R_{\text{gran}} + R_{\text{malignancy}}. \tag{6}$$

Format Compliance Reward.

$$R_{\text{format}} = \begin{cases} 1 & \text{if required output tags are present,} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Granularity-wise Reward.

$$R_{\text{gran}} = \begin{cases} 0.75 w_{i^*}, & \text{if } \hat{\ell} \in \mathcal{P}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

where $w_i = i/L$ represents the normalized depth of the taxonomy path. Details of R_{gran} and an example are given in Appendix.

Malignancy Discrimination Reward.

$$R_{\text{malignancy}} = \begin{cases} 0.25, & \text{if } \hat{b} = b^\star, \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

where $b^\star, \hat{b} \in \{\text{benign}, \text{malignant}, \text{precancerous in situ}\}$ denote the ground-truth and predicted malignancy categories, respectively, which correspond to a fundamental clinical classification task.

In summary, Skin-R1 integrates textbook-grounded reasoning supervision with reinforcement learning-based generalization. The combination of hierarchy-aware reasoning trajectories, supervised initialization, and GRPO-based policy refinement enables the model to produce clinically interpretable diagnostic predictions while remaining scalable to heterogeneous dermatology datasets.

4 Experiments and Results

Results Summary. Across a diverse set of dermatology benchmarks, Skin-R1 consistently achieves the strongest diagnostic performance among all evaluated models on both in-distribution (ID) and out-of-distribution (OOD) datasets while maintaining balanced performance on lesion condition classification. Ablation studies further show that trajectory-based SFT provides a crucial foundation for effective RL training, and that the proposed hierarchy-aware reward improves generalization to unseen disease types. Additional targeted evaluations confirm that Skin-R1 demonstrates stronger differential diagnosis (DDx) reasoning and improved awareness of disease taxonomies compared with existing VLM baselines.

Key Findings.

- **State-of-the-art diagnostic accuracy.** Skin-R1 achieves the highest accuracy on both ID and OOD disease diagnosis benchmarks.
- **Balanced clinical decision performance.** The model obtains the best combined accuracy and Macro-F1 on lesion condition classification.
- **Grounded reasoning improves RL training.** Trajectory-based SFT significantly improves structural alignment and raises the achievable performance ceiling for RL.
- **Hierarchy-aware rewards improve generalization.** The proposed granularity-wise reward yields stronger OOD performance by encouraging hierarchical reasoning across disease categories.
- **Improved differential diagnosis capability.** Targeted DDx and taxonomy-based evaluations confirm that Skin-R1 more reliably distinguishes visually similar dermatological conditions.

Training & Testing Datasets. We use six dermatology datasets for RL training and evaluation: PAD-UFES-20 [33], DermNet [8], BCN20000 [5], DERM12345 [56], Derm7pt [19], and HAM10000 [48]. To assess cross-dataset generalization, we

additionally evaluate the trained model on the dermoscopy subset of OmniMed-VQA [14]. Detailed descriptions of all datasets and dataset preprocessing are provided in Appendix.

Task Description. We design a visual question answering (VQA) evaluation framework to assess diagnostic reasoning under multiple distributional scenarios. The evaluation includes three settings:

1. **In-distribution disease diagnosis**, where the model predicts the disease category for dermoscopic images drawn from the training domain;
2. **Out-of-distribution (OOD) disease diagnosis**, which assesses generalization to unseen disease types outside the training distribution; and
3. **Lesion condition classification**, where the model classifies each case as `benign`, `malignant`, or `precancerous in situ`.

All tasks are formulated as multiple-choice VQA problems and evaluated primarily using accuracy. For lesion condition classification, Macro-F1 is additionally reported to account for class imbalance. To ensure fair comparison, an answer extraction protocol is applied to baseline models lacking the required response format. Details of the multiple-choice setup, distractor design, and answer extraction protocol are provided in Appendix.

Implementation Details. We adopt Qwen2.5-VL-7B-Instruct [2] as the backbone model. Both SFT and RL stages are trained using LoRA [12], with hyperparameters `lora_r=64`, `lora_alpha=32`, and `lora_dropout=0.1`. Further implementation details are provided in Appendix.

Baselines. We compare Skin-R1 with a diverse set of recent VLMs spanning general, medical, and dermatology-specific models. General-purpose VLMs include Qwen2.5-VL-7B and -32B [2], Qwen3-VL-8B and -32B [36], InternVL3-8B [64], and LLaVA-v1.6-7B and -13B [26]. Medical-domain VLMs include LLaVA-Med-7B [24], HuatuoGPT-Vision-7B [4], and MedGemma-4B [40]. We also include a dermatology-focused VLM, SkinVL-PubMM [59].

Methods such as Med-R1 [22] and MedVLM-R1 [34] primarily rely on GRPO-style RL training. Therefore, the **RL without SFT** configuration in Sec. 4.2 can be viewed as their adaptation to dermatological diagnosis.

4.1 Skin-R1 Compared with SOTA Benchmarks

As shown in Tab. 2, **Skin-R1 outperforms all baseline models on disease diagnosis**. It achieves the highest average accuracy on both ID datasets (0.6385, +0.1955 over the next-best model) and OOD datasets (0.7171, +0.0284), demonstrating strong diagnostic accuracy and generalization capability.

For lesion condition classification (Tab. 3), **Skin-R1 achieves the most balanced performance**, obtaining the highest average accuracy (0.6928, +0.0297) and Macro-F1 score (0.4287, +0.0633).

Table 2: Comparison of Skin-R1 with baseline LVLMs on in-distribution and OOD disease diagnostic tasks. **Bold** and underlined values indicate the best and second-best results, respectively.

Type	Model	In-distribution							OOD					
		BCN20k	HAM10k	PAD	Derm12345	Derm7pt	DermNet	Avg.	ISBI16	ISIC18	ISIC19	ISIC20	Monk22	Avg.
General	Qwen2.5-VL-7B-Instruct	0.4911	0.5162	0.5033	0.3521	0.2962	0.4852	0.4367	0.3655	0.6301	0.6359	0.3325	**0.7078**	0.5027
	Qwen3-VL-8B-Instruct	0.2947	0.2085	0.3275	0.2386	0.2354	0.2899	0.2585	0.6996	0.2890	0.3934	**0.8837**	0.4610	0.5939
	InternVL3-8B	0.4996	0.4705	0.6009	0.3453	0.3364	0.5473	0.4430	0.4580	0.6474	**0.7746**	0.3531	0.6948	0.5841
	LLaVA-v1.6-7B	0.2299	0.2416	0.3080	0.3046	0.2354	0.2885	0.2636	0.5819	0.1908	0.3134	0.5899	0.3312	0.4351
	LLaVA-v1.6-13B	0.3933	0.3779	0.4469	0.4121	0.3376	0.4193	0.3960	0.2563	0.1676	0.3415	0.1584	0.2143	0.2560
	Qwen2.5-VL-32B-Instruct	0.4746	0.4057	0.4469	0.3815	0.3596	**0.5763**	0.4307	0.5966	0.2601	0.3445	0.6040	0.3247	0.4587
	Qwen3-VL-32B-Instruct	0.3156	0.2363	0.3688	0.2801	0.2536	0.3304	0.2892	**0.6996**	0.3353	0.5596	0.8408	0.6688	0.6651
Medical	LLaVA-Med-7b	0.2508	0.2091	0.2039	0.2503	0.2135	0.2870	0.2400	0.2542	0.2601	0.3928	0.5314	0.4545	0.4170
	HuatuoGPT-Vision-7B	0.4799	0.4944	0.4295	0.3638	0.2561	0.4719	0.4241	0.4202	0.5896	0.5657	0.4505	0.5779	0.5101
	MedGemma-4B	0.4130	0.2627	0.3753	0.3767	0.2924	0.3240	0.3548	0.5168	0.5318	0.7214	0.7508	0.5584	0.6887
Dermatology	SkinVL-PubMM	-	-	-	-	-	-	-	-	-	-	-	-	-
	Skin-R1	**0.6345**	**0.7214**	**0.6573**	**0.6608**	**0.4955**	0.5370	**0.6385**	0.6828	**0.7630**	0.7630	0.6708	0.6494	**0.7171**

Table 3: Comparison of Skin-R1 with baseline LVLMs on in-distribution skin lesion diagnostic tasks. **Bold** and underlined values indicate the best and second-best results, respectively.

Type	Model	In-distribution													
		BCN20k		HAM10k		PAD		Derm12345		Derm7pt		DermNet		Avg.	
		Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
General	Qwen2.5-VL-7B-Instruct	0.4308	0.3346	0.2905	0.2449	0.4100	0.3632	0.3046	0.2045	0.3661	0.3468	0.5399	0.4390	0.3698	0.2910
	Qwen3-VL-8B-Instruct	0.4581	0.3105	0.6592	0.3601	0.3124	0.2891	0.7437	0.3716	0.6494	0.5858	0.5059	0.3503	0.5924	0.3649
	InternVL3-8B	0.4795	**0.3768**	0.3514	0.2639	0.5033	**0.4932**	0.4769	0.3091	0.4334	0.4343	0.5932	**0.5366**	0.4619	0.3610
	LLaVA-v1.6-7B	0.4014	0.3036	0.5897	0.3669	0.2907	0.2743	0.6414	0.3285	0.5356	0.5361	0.4601	0.3109	0.5174	0.3428
	LLaVA-v1.6-13B	0.3704	0.3070	0.4057	0.2888	0.3536	0.3458	0.3847	0.2330	0.3739	0.4181	0.4267	0.3524	0.3849	0.2978
	Qwen2.5-VL-32B-Instruct	0.4799	0.3507	0.4917	0.3390	0.3514	0.3027	0.6093	0.3442	0.4903	0.4948	0.5896	0.4511	0.5231	0.3654
	Qwen3-VL-32B-Instruct	0.4481	0.2452	0.7598	0.3508	0.2560	0.1968	0.8716	0.3923	0.6986	0.4991	0.5333	0.3092	0.6491	0.3337
Medical	LLaVA-Med-7b	0.4465	0.2058	**0.7968**	0.2956	0.2299	0.1246	**0.8994**	0.3157	0.6986	0.4113	0.5459	0.2354	0.6631	0.2714
	HuatuoGPT-Vision-7B	0.4489	0.3525	0.3349	0.2858	**0.5098**	0.4725	0.3682	0.2390	0.3842	0.3957	0.5562	0.4680	0.4105	0.3267
	MedGemma-4B	0.0817	0.0746	0.0668	0.0558	0.3384	0.2302	0.0692	0.0574	0.0517	0.0813	0.1746	0.1531	0.0942	0.0816
Dermatology	SkinVL-PubMM	0.4368	0.2373	0.7187	0.3459	0.2560	0.1733	0.8740	0.3196	0.6779	0.5491	0.5592	0.2816	0.6392	0.3100
	Skin-R1	**0.5370**	0.3486	0.7472	**0.4039**	0.3536	0.2888	0.8918	**0.4786**	**0.7089**	**0.6589**	**0.6257**	0.4271	**0.6928**	**0.4287**

Several baseline models exhibit notable biases. MedGemma-4B shows a strong bias toward the rare class `C: precancerous in situ` (7,860 of 8,390 predictions), leading to poor overall performance. LLaVA-Med-7B demonstrates a bias toward the majority class `A: benign` (5,563 of 8,390 predictions), resulting in low F1 scores. SkinVL-PubMM also shows a mild bias toward the `A: benign` category.

Moreover, SkinVL-PubMM rarely produces meaningful diagnostic responses on both ID and OOD tasks even when prompts are simplified. Consequently, its metrics are reported as "–", indicating that the model fails to produce meaningful answers rather than being affected by the answer extraction protocol.

4.2 Ablation Study

We summarize ablation results evaluating trajectory SFT, RL stages, and reward design in Tab. 4 and Tab. 5.

1. **Stage 1 (Trajectory SFT) Improves Diagnostic Ability and Structural Alignment.** The **SFT** model outperforms the **base model** Qwen2.5-VL-7B-Instruct overall. The base model also frequently fails to follow the required output format. Under strict answer extraction, it fails to produce valid answers in approximately 30% of ID diagnosis cases, compared with only 2%

for the SFT model. These results highlight the critical role of trajectory-based SFT in establishing grounded diagnostic reasoning and structural alignment.

2. **Stage 2 (RL) Significantly Boosts Performance and Generalization.** The full Skin-R1 pipeline achieves markedly higher performance than earlier checkpoints and the base model. The proposed reward design encourages structured hierarchical reasoning and lesion differentiation. Strong OOD performance further confirms improved generalization. We select the checkpoint at 1,500 RL steps as the representative model.
3. **High-Quality SFT is Essential for Effective RL.** The complete Skin-R1 pipeline consistently outperforms the **RL without SFT** variant. Although RL without SFT improves upon the base model, its best performance remains significantly lower (Fig. 2). This indicates that RL-only approaches lack the grounded knowledge necessary for robust reasoning.
4. **Granularity-wise Reward Improves Generalization Ability.** Replacing the granularity-aware reward with a standard binary reward leads to weaker OOD performance. This suggests that the proposed reward provides more informative learning signals by encouraging hierarchical reasoning across the disease taxonomy.

Table 4: Ablation study of Skin-R1 across in-distribution and OOD datasets. **Bold** and underlined indicate the best and second-best performance, respectively.

Model	In-distribution							Out-of-distribution (OOD)					
	BCN20k	HAM10k	PAD	Derm12345	Derm7pt	DermNet	Avg.	ISBI16	ISIC18	ISIC19	ISIC20	Monk22	Avg.
Qwen2.5-VL-7B-Instruct	0.4911	0.5162	0.5033	0.3521	0.2962	0.4852	0.4367	0.3655	0.6301	0.6359	0.3325	0.7078	0.5027
SFT	0.4899	0.6122	0.5358	0.4181	0.2937	0.4793	0.4743	0.5693	0.4509	0.5651	0.4249	0.6039	0.5153
RL without SFT	0.5898	0.6671	0.5813	0.6056	0.3855	0.5296	0.5843	0.4223	0.7514	0.7208	0.3663	**0.7597**	0.5674
RL with standard reward	0.6252	**0.7227**	0.6182	**0.6628**	**0.5265**	**0.5444**	0.6379	**0.6870**	0.7399	0.6885	0.5528	0.5195	0.6386
Skin-R1 (500 steps)	0.6139	0.6870	0.6095	0.6020	0.3351	0.4882	0.5875	0.6744	0.6474	0.6805	**0.6848**	0.5519	0.6742
Skin-R1 (1,000 steps)	0.6083	0.7009	0.6508	0.6487	0.4502	0.4956	0.6156	0.6660	0.7168	**0.7648**	0.5982	0.5909	0.6870
Skin-R1 (1,500 steps)	**0.6345**	0.7214	**0.6573**	0.6608	0.4955	0.5370	**0.6385**	0.6828	**0.7630**	0.7630	0.6708	0.6494	**0.7171**

Table 5: Ablation study of Skin-R1 on skin lesion diagnostic task across in-distribution datasets. **Bold** and underlined indicate the best and second-best performance, respectively.

Model	In-distribution													
	BCN20k		HAM10k		PAD		Derm12345		Derm7pt		DermNet		Avg.	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Qwen2.5-VL-7B-Instruct	0.4308	0.3346	0.2905	0.2449	0.4100	0.3632	0.3046	0.2045	0.3661	0.3468	0.5399	0.4390	0.3698	0.2910
SFT	0.4972	0.3228	0.6373	0.3567	0.3688	0.2970	0.8072	0.4037	0.6223	0.5876	0.6006	0.4590	0.6271	0.3868
RL without SFT	0.4819	**0.3654**	0.3812	0.3037	**0.4338**	**0.3740**	0.4748	0.2839	0.3946	0.3824	0.5962	0.4792	0.4602	0.3414
RL with standard reward	**0.5407**	0.3633	0.6810	0.3849	0.3991	0.3232	0.8378	0.4480	0.6624	0.6416	**0.6450**	**0.4841**	0.6658	0.4254
Skin-R1(500 steps)	0.5229	0.3305	**0.7565**	0.3867	0.3362	0.2638	**0.8994**	0.4584	**0.7413**	0.6571	0.6036	0.4215	0.6928	0.4123
Skin-R1(1,000 steps)	0.5342	0.3528	0.6962	0.3880	0.3601	0.2916	0.8519	0.4421	0.6792	0.6386	0.6228	0.4325	0.6684	0.4150
Skin-R1(1,500 steps)	0.5370	0.3486	0.7472	**0.4039**	0.3536	0.2888	0.8918	**0.4786**	0.7089	**0.6589**	0.6257	0.4271	**0.6928**	**0.4287**

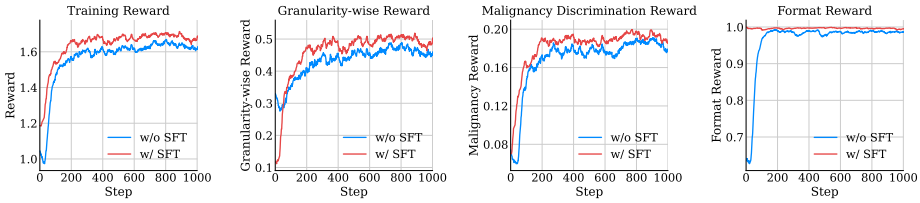

Fig. 2: Training reward curve comparison of Skin-R1 with or without SFT stage.

4.3 Targeted DDx and Hierarchical Evaluation

To evaluate DDx and hierarchy-aware reasoning, we construct two targeted evaluation tasks: (1) *DDx diagnosis* and (2) *hierarchical diagnosis*. These tasks modify the standard ID benchmark by replacing distractors using the disease taxonomy $\mathcal{T}_{tax}$ and the differential diagnosis graph $\mathcal{G}_{DDx}$.

Evaluation uses held-out images never seen during training. The taxonomy and DDx graph are used only to generate distractors and are not provided to the model during inference.

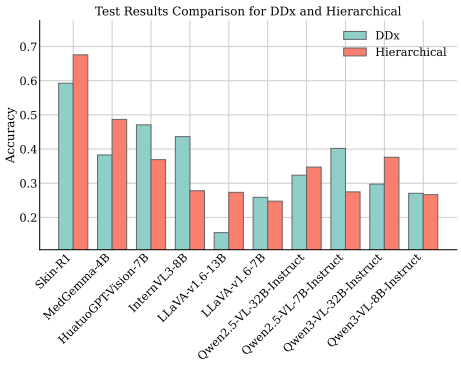

Fig. 3: Comparison on DDx and hierarchical diagnosis tasks.

- For hierarchical diagnosis, distractors correspond to ancestor nodes of the ground-truth diagnosis.
- For DDx diagnosis, distractors correspond to confusable neighboring diseases in $\mathcal{G}_{DDx}$.

As shown in Fig. 3, Skin-R1 achieves the highest accuracy in both settings, demonstrating stronger differentiation of visually similar diseases and improved awareness of taxonomic relationships.

4.4 Qualitative Analysis

Qualitative comparisons (Appendix) show that Skin-R1 produces responses that are more concise, accurate, and consistent with expert diagnostic reasoning compared with the base Qwen2.5-VL-7B model.

For example, in Open-ended Case 1 of Appendix, Skin-R1 correctly concludes that the observed features "indicate a benign pigmented nevus rather than melanoma", demonstrating precise DDx reasoning. In contrast, the base model fails to capture this critical clinical distinction.

While the base model tends to produce longer descriptive outputs, it often sacrifices diagnostic precision. Skin-R1, in contrast, produces more focused and reliable diagnoses, reflecting stronger grounded diagnostic reasoning capabilities.

5 Summary and Discussion

Summary. We addressed three major challenges that limit the trustworthiness and clinical utility of dermatological vision–language models: dataset heterogeneity, the scarcity of grounded reasoning supervision, and the difficulty of transferring complex diagnostic reasoning to large-scale datasets with sparse annotations.

We introduced Skin-R1, a dermatology-oriented VLM trained through a three-stage framework that integrates textbook-grounded supervision with reinforcement learning. First, we construct a textbook-based reasoning generator that synthesizes hierarchy-aware and differential-diagnosis (DDx)-informed diagnostic trajectories, providing expert-aligned supervision. Second, these trajectories are used for supervised fine-tuning (SFT), establishing a grounded reasoning foundation. Third, we introduce a reinforcement learning framework that leverages hierarchical diagnostic rewards to generalize these reasoning patterns to large-scale sparsely annotated datasets.

Extensive experiments across multiple dermatology benchmarks show that Skin-R1 consistently outperforms state-of-the-art baselines in diagnostic accuracy and robustness. Ablation studies further highlight the critical role of trajectory-based SFT in enabling effective reinforcement learning and improving generalization.

Overall, our results highlight the importance of grounded reasoning supervision for developing trustworthy medical VLMs and suggest a promising direction toward clinically reliable multimodal diagnostic systems.

Discussion. Although Skin-R1 improves diagnostic reasoning and fine-grained prediction, several limitations remain before reliable real-world deployment.

First, the reward signal and evaluation is answer-level: models are scored by multiple-choice accuracy, while the correctness and faithfulness of the generated reasoning traces are examined only qualitatively; quantifying rationale quality in more open-ended, clinically realistic settings, together with confidence calibration, abstention or referral under high uncertainty, and severity-aware error analysis for malignant conditions, is left to future work.

Second, our pipeline extracts a taxonomy and DDx graph offline and keeps them fixed during training; since these structures also generate the distractors for the in-distribution and targeted evaluations, those results partially reflect consistency with this fixed clinical knowledge, whereas the out-of-distribution evaluation (OmniMedVQA), whose options are independent of our taxonomy, is not subject to this coupling. This static design also limits adaptation to evolving or dataset-specific ontologies.

Third, Skin-R1 is instantiated only on Qwen2.5-VL-7B; whether the same SFT+RL recipe transfers to other backbones, including medical ones, and how it compares with additional dermatology-specific foundation models, remains to be studied.

Finally, we observe modest degradation in certain Fitzpatrick skin types (Appendix), likely reflecting training-data bias, and our outcome-level objective does

not explicitly reward intermediate reasoning steps; fairness-aware training, domain adaptation, and process-level supervision are promising directions; however, defining reliable reward signals for clinical reasoning remains challenging.

6 Acknowledgements.

The work of Vasant G Honavar and Zehao Liu was supported in part by grants from the National Science Foundation (2226025, 2225824), the National Center for Advancing Translational Sciences, and the National Institutes of Health (UL1 TR002014).

Bibliography

[1] Awais, M., Naseer, M., Khan, S., Anwer, R.M., Cholakkal, H., Shah, M., Yang, M.H., Khan, F.S.: Foundation models defining a new era in vision: a survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)

[2] Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., et al.: Qwen2. 5-vl technical report. arXiv preprint arXiv:2502.13923 (2025)

[3] Chen, J., He, Q., Yuan, S., Chen, A., Cai, Z., Dai, W., Yu, H., Yu, Q., Li, X., Chen, J., et al.: Enigmata: Scaling logical reasoning in large language models with synthetic verifiable puzzles. arXiv preprint arXiv:2505.19914 (2025)

[4] Chen, J., Gui, C., Ouyang, R., Gao, A., Chen, S., Chen, G.H., Wang, X., Zhang, R., Cai, Z., Ji, K., et al.: Huatuogpt-vision, towards injecting medical visual knowledge into multimodal llms at scale. arXiv preprint arXiv:2406.19280 (2024)

[5] Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A.C., Puig, S., et al.: Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)

[6] Cruz-Roa, A.A., Arevalo Ovalle, J.E., Madabhushi, A., González Osorio, F.A.: A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In: International conference on medical image computing and computer-assisted intervention. pp. 403–410. Springer (2013)

[7] Daneshjou, R., Yuksekgonul, M., Cai, Z.R., Novoa, R., Zou, J.Y.: Skincon: A skin disease dataset densely annotated by domain experts for fine-grained debugging and analysis. Advances in Neural Information Processing Systems **35**, 18157–18167 (2022)

[8] Dermnet Team: Dermnet. https://www.kaggle.com/datasets/shubhamgoel27/dermnet (2020)

[9] Feng, H., Berk-Krauss, J., Feng, P.W., Stein, J.A.: Comparison of dermatologist density between urban and rural counties in the united states. JAMA dermatology **154**(11), 1265–1271 (2018)

[10] 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)

[11] Hay, R.J., Johns, N.E., Williams, H.C., Bolliger, I.W., Dellavalle, R.P., Margolis, D.J., Marks, R., Naldi, L., Weinstock, M.A., Wulf, S.K., et al.: The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. Journal of investigative dermatology **134**(6), 1527–1534 (2014)

[12] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In:

International Conference on Learning Representations (2022), https://openreview.net/forum?id=nZeVKeeFYf9
[13] Hu, L., Lai, S., Chen, W., Xiao, H., Lin, H., Yu, L., Zhang, J., Wang, D.: Towards multi-dimensional explanation alignment for medical classification. Advances in Neural Information Processing Systems **37**, 129640–129671 (2024)
[14] Hu, Y., Li, T., Lu, Q., Shao, W., He, J., Qiao, Y., Luo, P.: Omnimedvqa: A new large-scale comprehensive evaluation benchmark for medical lvlm. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22170–22183 (2024)
[15] Hu, Z., Wang, Y., Dong, H., Xu, Y., Saha, A., Xiong, C., Hooi, B., Li, J.: Beyond'aha!': Toward systematic meta-abilities alignment in large reasoning models. arXiv preprint arXiv:2505.10554 (2025)
[16] Huang, J., Chen, X., Mishra, S., Zheng, H.S., Yu, A., Song, X., Zhou, D.: Large language models cannot self-correct reasoning yet. In: International conference on learning representations. vol. 2024, pp. 32808–32824 (2024)
[17] Jiang, S., Wang, Y., Song, S., Zhang, Y., Meng, Z., Lei, B., Wu, J., Sun, J., Liu, Z.: Omniv-med: Scaling medical vision-language model for universal visual understanding. arXiv preprint arXiv:2504.14692 (2025)
[18] Kang, S., Amagai, M., Bruckner, A.L., Enk, A.H., Margolis, D.J., McMichael, A.J., Orringer, J.S. (eds.): Fitzpatrick's Dermatology. McGraw-Hill Education, New York, 9 edn. (2019)
[19] Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE Journal of Biomedical and Health Informatics **23**(2), 538–546 (2019). https://doi.org/10.1109/JBHI.2018.2824327
[20] Kim, C., Gadgil, S.U., DeGrave, A.J., Omiye, J.A., Cai, Z.R., Daneshjou, R., Lee, S.I.: Transparent medical image ai via an image–text foundation model grounded in medical literature. Nature medicine **30**(4), 1154–1165 (2024)
[21] Koh, P.W., Nguyen, T., Tang, Y.S., Mussmann, S., Pierson, E., Kim, B., Liang, P.: Concept bottleneck models. In: International conference on machine learning. pp. 5338–5348. PMLR (2020)
[22] Lai, Y., Zhong, J., Li, M., Zhao, S., Yang, X.: Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. arXiv preprint arXiv:2503.13939 (2025)
[23] Lanham, T., Chen, A., Radhakrishnan, A., Steiner, B., Denison, C., Hernandez, D., Li, D., Durmus, E., Hubinger, E., Kernion, J., et al.: Measuring faithfulness in chain-of-thought reasoning. arXiv preprint arXiv:2307.13702 (2023)
[24] Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for biomedicine in one day. Advances in Neural Information Processing Systems **36**, 28541–28564 (2023)
[25] Lin, T., Zhang, W., Li, S., Yuan, Y., Yu, B., Li, H., He, W., Jiang, H., Li, M., Song, X., et al.: Healthgpt: A medical large vision-language model

for unifying comprehension and generation via heterogeneous knowledge adaptation. arXiv preprint arXiv:2502.09838 (2025)
[26] Liu, H., Li, C., Li, Y., Li, B., Zhang, Y., Shen, S., Lee, Y.J.: Llavanext: Improved reasoning, ocr, and world knowledge (2024)
[27] Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. arXiv:2304.08485 (2023)
[28] Liu, L., Chen, L., Zhang, Z., Tian, M., Cui, H., Li, R., Liu, Z., Ju, Q., Li, Q., Zhou, H.Y.: Skinflow: Efficient information transmission for open dermatological diagnosis via dynamic visual encoding and staged rl. arXiv preprint arXiv:2601.09136 (2026)
[29] Liu, Y., Jain, A., Eng, C., Way, D.H., Lee, K., Bui, P., Kanada, K., de Oliveira Marinho, G., Gallegos, J., Gabriele, S., et al.: A deep learning system for differential diagnosis of skin diseases. Nature medicine **26**(6), 900–908 (2020)
[30] Lu, M.Y., Chen, B., Williamson, D.F., Chen, R.J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L.P., Gerber, G., et al.: A visual-language foundation model for computational pathology. Nature medicine **30**(3), 863–874 (2024)
[31] Luo, Y., Zhang, J., Fan, S., Yang, K., Hong, M., Wu, Y., Qiao, M., Nie, Z.: Biomedgpt: An open multimodal large language model for biomedicine. IEEE Journal of Biomedical and Health Informatics (2024)
[32] OpenAI: Gpt-4 technical report (2023)
[33] Pacheco, A.G., Krohling, R.A.: The impact of patient clinical information on automated skin cancer detection. Computers in biology and medicine **116**, 103545 (2020)
[34] Pan, J., Liu, C., Wu, J., Liu, F., Zhu, J., Li, H.B., Chen, C., Ouyang, C., Rueckert, D.: 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. pp. 337–347. Springer (2025)
[35] Pham, T.H., Ngo, C.: Rarl: Improving medical vlm reasoning and generalization with reinforcement learning and lora under data and hardware constraints. arXiv preprint arXiv:2506.06600 (2025)
[36] Qwen AI and Alibaba Cloud: Qwen3-vl: Sharper vision, deeper thought, broader action. Blog post (Oct 2025), https://qwen.ai/blog?id=99f0335c4ad9ff6153e517418d48535ab6d8afef&from=research.latest-advancements-list, accessed: 2025-11-04
[37] Ren, W., Zhu, J., Liu, Z., Zhao, T., Honavar, V.: A comprehensive survey of electronic health record modeling: From deep learning approaches to large language models. arXiv preprint arXiv:2507.12774 (2025)
[38] Resneck Jr, J., Kimball, A.B.: The dermatology workforce shortage. Journal of the American Academy of Dermatology **50**(1), 50–54 (2004)
[39] Rui, S., Chen, K., Ma, W., Wang, X.: Improving medical reasoning with curriculum-aware reinforcement learning. arXiv preprint arXiv:2505.19213 (2025)

[40] Sellergren, A., Kazemzadeh, S., Jaroensri, T., Kiraly, A., Traverse, M., Kohlberger, T., Xu, S., Jamil, F., Hughes, C., Lau, C., et al.: Medgemma technical report. arXiv preprint arXiv:2507.05201 (2025)
[41] Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
[42] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., Gal, Y.: Ai models collapse when trained on recursively generated data. Nature **631**(8022), 755–759 (2024)
[43] Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J.B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M., Hauth, A., et al.: Gemini: a family of highly capable multimodal models. arXiv preprint arXiv:2312.11805 (2023)
[44] Team, Q.: Qwq-32b: Embracing the power of reinforcement learning (March 2025), https://qwenlm.github.io/blog/qwq-32b/
[45] Thawkar, O., Shaker, A., Mullappilly, S.S., Cholakkal, H., Anwer, R.M., Khan, S., Laaksonen, J., Khan, F.S.: Xraygpt: Chest radiographs summarization using medical vision-language models. arXiv preprint arXiv:2306.07971 (2023)
[46] Tian, D., Jiang, S., Zhang, L., Lu, X., Xu, Y.: The role of large language models in medical image processing: a narrative review. Quantitative Imaging in Medicine and Surgery **14**(1), 1108–1121 (2024)
[47] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
[48] Tschandl, P.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions (2018), https://doi.org/10.7910/DVN/DBW86T
[49] Wang, K., Zhang, G., Zhou, Z., Wu, J., Yu, M., Zhao, S., Yin, C., Fu, J., Yan, Y., Luo, H., et al.: A comprehensive survey in llm (-agent) full stack safety: Data, training and deployment. arXiv preprint arXiv:2504.15585 (2025)
[50] Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing. vol. 2022, p. 3876 (2022)
[51] Xie, T., Gao, Z., Ren, Q., Luo, H., Hong, Y., Dai, B., Zhou, J., Qiu, K., Wu, Z., Luo, C.: Logic-rl: Unleashing llm reasoning with rule-based reinforcement learning. arXiv preprint arXiv:2502.14768 (2025)
[52] Yan, S., Li, X., Hu, M., Jiang, Y., Yu, Z., Ge, Z.: Make: Multi-aspect knowledge-enhanced vision-language pretraining for zero-shot dermatological assessment. arXiv preprint arXiv:2505.09372 (2025)
[53] Yan, S., Li, X., Mo, D., Tschandl, P., Jiang, Y., Wang, Z., Hu, M., Ju, L., Vico-Alonso, C., Zheng, Y., et al.: A vision-language foundation model for zero-shot clinical collaboration and automated concept discovery in dermatology. arXiv preprint arXiv:2602.10624 (2026)

[54] Yan, S., Yu, Z., Primiero, C., Vico-Alonso, C., Wang, Z., Yang, L., Tschandl, P., Hu, M., Ju, L., Tan, G., et al.: A multimodal vision foundation model for clinical dermatology. Nature Medicine pp. 1–12 (2025)
[55] Yan, S., Yu, Z., Zhang, X., Mahapatra, D., Chandra, S.S., Janda, M., Soyer, P., Ge, Z.: Towards trustable skin cancer diagnosis via rewriting model's decision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11568–11577 (2023)
[56] Yilmaz, A., Yasar, S.P., Gencoglan, G., Temelkuran, B.: Derm12345: A large, multisource dermatoscopic skin lesion dataset with 40 subclasses. Scientific Data **11**(1), 1302 (2024)
[57] Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., Chen, E.: A survey on multimodal large language models. arXiv preprint arXiv:2306.13549 (2023)
[58] Yuan, Y., Chao, M., Lo, Y.C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE transactions on medical imaging **36**(9), 1876–1886 (2017)
[59] Zeng, W., Sun, Y., Ma, C., Tan, W., Yan, B.: Mm-skin: Enhancing dermatology vision-language model with an image-text dataset derived from textbooks. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 3769–3778 (2025)
[60] Zhang, J., Huang, J., Yao, H., Liu, S., Zhang, X., Lu, S., Tao, D.: R1-vl: Learning to reason with multimodal large language models via step-wise group relative policy optimization. arXiv preprint arXiv:2503.12937 (2025)
[61] Zhou, H., Liu, F., Gu, B., Zou, X., Huang, J., Wu, J., Li, Y., Chen, S.S., Zhou, P., Liu, J., et al.: A survey of large language models in medicine: Progress, application, and challenge. arXiv preprint arXiv:2311.05112 (2023)
[62] Zhou, J., He, X., Sun, L., Xu, J., Chen, X., Chu, Y., Zhou, L., Liao, X., Zhang, B., Afvari, S., et al.: Pre-trained multimodal large language model enhances dermatological diagnosis using skingpt-4. Nature Communications **15**(1), 5649 (2024)
[63] Zhu, D., Chen, J., Shen, X., Li, X., Elhoseiny, M.: Minigpt-4: Enhancing vision-language understanding with advanced large language models. arXiv preprint arXiv:2304.10592 (2023)
[64] Zhu, J., Wang, W., Chen, Z., Liu, Z., Ye, S., Gu, L., Tian, H., Duan, Y., Su, W., Shao, J., et al.: Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. arXiv preprint arXiv:2504.10479 (2025)
[65] Zhu, W., Dong, X., Li, X., Qiu, P., Chen, X., Razi, A., Sotiras, A., Su, Y., Wang, Y.: Toward effective reinforcement learning fine-tuning for medical vqa in vision-language models. arXiv preprint arXiv:2505.13973 (2025)