This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

CMR-RD: Long-Tailed Adaptive VLM for Explainable CMR Diagnosis

Yansong Li[1], Zhongxi Qiu[2], Yun Tian[1]*, Zheng Jinyu[3], Shuo Li[4]
[1]Beijing Normal University, [2]Southern University of Science and Technology,
[3]Shanghai Jiao Tong University, [4]Case Western Reserve University
liyansong@mail.bnu.edu.cn, tianyun@bnu.edu.cn

Abstract

Cardiac magnetic resonance (CMR) is the clinical gold standard for assessing cardiovascular diseases, but its interpretation relies on expert experience and remains challenging, particularly for identifying rare diseases. Existing automated methods lack interpretable reasoning processes, limiting clinical adoption. Although vision-language models (VLMs) possess basic visual understanding and text generation capabilities, they still lack verifiable reasoning chains in medical diagnosis and underperform on minority classes in long-tail distributions. To address these challenges, we propose CMR-RD, to our knowledge the first VLM for interpretable diagnosis in CMR, capable of generating explicit diagnostic chains aligned with imaging evidence. We construct a CMR dataset that reflects real-world clinical distributions, comprising five disease categories (including two rare conditions) plus normal controls. Building on this, the general-purpose VLM is aligned to medical and CMR semantics using large-scale medical vision–text data, and cold-start training is used to enhance its understanding of medical concepts and basic reasoning. To enhance reasoning and performance on rare samples, we propose Group Phase Policy Optimization (GPPO), which combines online multi-stage reinforcement learning (RL) with adaptive sampling. GPPO enables the model to proactively explore rare and underperforming classes, thereby effectively mitigating long-tail bias. Experiments demonstrate that CMR-RD achieves highly competitive accuracy and reasoning-chain correctness compared with medical and general VLMs, shows stronger recognition of rare categories.

1. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of death worldwide [22]. Although cardiac magnetic resonance (CMR) is the gold standard for functional assessment and disease characterization [14], its clinical value is constrained by high barriers to interpretation and uneven resource distribution. Primary care facilities lack experienced subspecialty readers [1, 9], complex cases are time-consuming to interpret, and rare categories are prone to under-recognition, directly affecting patient prognosis and triage efficiency. The key clinical risks in CMR are concentrated in challenging samples and long-tail categories [20, 21], yet current automated models suffer from poor interpretability and weak recognition of rare classes, limiting clinical trust and large-scale adoption[28].

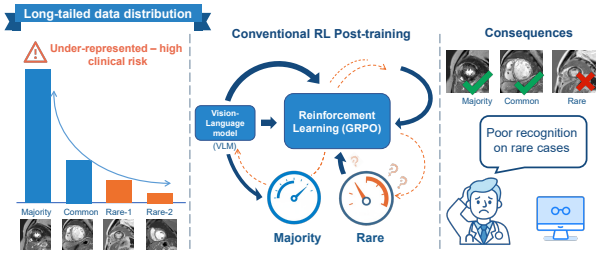


Figure 1. Long-tailed distributions can undermine the post-trained VLM's ability to recognize rare classes, which is especially critical in CMR diagnosis. Existing RL-based post-training methods (e.g., GRPO) remain vulnerable to class imbalance, leading to reward sparsity and degraded recognition of rare classes.

Currently, Vision-Language Models (VLMs) possess basic visual understanding and text generation capabilities and hold promise for tackling clinical visual diagnosis tasks [7, 13]. However, general VLMs underperform in cardiac disease diagnosis because they are not aligned with medical contexts and cardiovascular expertise. Moreover, they do not provide fully explicit reasoning processes (chain of thought, CoT) [34], falling short of the transparency required for clinical decision-making, which is critical for auditability and explainability in clinical practice. Medical data naturally exhibit a long-tailed distribution with scarce rare-class samples [44]. As shown in Fig. 1, existing reinforcement learning–based post-training methods (e.g., GRPO) [25, 30] are dominated by majority classes in

*Corresponding author.

both sampling and value estimation, biasing policy updates toward common classes and markedly weakening recognition and reasoning for rare classes [15, 26, 41, 43]. In particular, when training follows the empirical class distribution, majority classes are sampled more frequently, where advantage and value estimates for rare classes may exhibit higher variance and lower confidence, further diminishing their gradient contributions. Consequently, even if RL is used to strengthen medical alignment or explicit reasoning, the long-tailed data distribution causes the learned policy to favor common classes, limiting improvements on rare-class diagnosis and reasoning. Static class-balancing or data augmentation strategies lack stage-wise adaptivity during training and therefore struggle to achieve optimal performance across all categories [17, 27, 33, 37].

To address these key obstacles, we propose CMR-RD, the first VLM with explicit reasoning for explainable CMR diagnosis. It simulates clinicians' diagnostic processes to present a transparent diagnostic chain, while effectively identifying rare cardiac diseases. We constructed a CMR dataset reflecting real clinical distributions, including normal controls and five categories of cardiac diseases (two of which are rare), and designed a test set that evaluates both overall performance and long-tail robustness. We propose a two-stage training framework. In Stage 1, to ensure that a general VLM can produce answers that conform to clinical expression standards and maintain factual consistency, we align the general VLM to the medical context and CMR image semantics by jointly leveraging large-scale medical vision-language data and smaller CMR-specific data, thereby enhancing the model's understanding of medical concepts and reporting terminology. We also adopt a cold-start strategy to strengthen the model's basic reasoning capabilities. In Stage 2, to enable the model to self-evolve its reasoning ability using process-free supervision data and to reinforce recognition of rare categories, we propose the Group-Phased Policy Optimization (GPPO) algorithm. By decomposing the RL optimization into online multi-stage updates, GPPO guides the model to actively explore and to learn in a targeted manner for minority classes. We further introduce a dynamic sampling mechanism with stage-wise adaptive data mixing to mitigate long-tailed bias. To further reduce hallucinations and misleading descriptions, we incorporate lesion grounding into both the reward signal and the training objective, shifting the model from global image generalization to evidence-based, explainable decisions grounded in lesion findings, thereby improving the consistency between textual descriptions and imaging evidence. On the CMR diagnostic test set, CMR-RD outperforms other medical or general VLMs in terms of diagnostic accuracy and reasoning-chain correctness. Compared with alternative training methods, CMR-RD effectively addresses the long-tail problem, and attains highly competitive diagnostic performance. The contributions of this study are summarized as follows:

- We propose CMR-RD, the first VLM for explainable cardiac diagnosis. It can generate interpretable diagnostic reasoning chains while accurately identifying even rare cardiac conditions. Moreover, we constructed a new CMR dataset for training model.
- Based on the proposed GPPO algorithm, we have enhanced the reasoning capabilities of CMR-RD and strengthened rare and underperforming categories through an automatic exploration mechanism with an adaptive sampling strategy.
- CMR-RD significantly outperforms both medical and general VLM baselines in CMR diagnostic accuracy. By introducing a lesion grounding reward, it improves factual consistency between descriptions and images, thereby enhancing clinical usability.

2. Related Works

2.1. Medical Vision-Language Models

In recent years, VLMs have advanced rapidly for medical understanding and report generation. General models adapted via medical-domain instruction tuning and cross-modal alignment (e.g., Qwen-VL/Qwen2.5-VL, InternVL3) can acquire basic capabilities, yet clinical adaptation often remains at answer-level supervision with templated constraints, lacking systematic calibration for professional terminology, report structure, and consistency with imaging evidence [3, 7, 29, 38, 45].

Beyond these, several medical VLMs directly target clinical imaging–language alignment: Med-Flamingo and BiomedGPT extend open VLMs with biomedical corpora and radiology image-text alignment for zero/few-shot reporting [19, 23]; LLaVA-Med and LLaVA-Rad leverage curated instruction tuning on radiology dialogues and QA to enhance image-grounded clinical reasoning and captioning [5, 16]; RadFM and Radiology-GPT couple large-scale PACS-style image–report pretraining with structured prompt schemas for improved report generation and evidence grounding [18, 35]; and MIMIC-CXR-VQA and MEDIQA-Chat style benchmarks further drive progress in localized lesion reasoning and evidence-linked answers [2, 10], though gaps remain in consistent terminology control and verifiable evidence grounding for specialized modalities like CMR. For explicit reasoning, CoT and RL-based post-training (e.g., PathVLM-R1, MedVLM-R1, MedTVT-R1) have improved domain professionalism and consistency [15, 26, 36, 43]. Despite the benefits of post-training for model outputs, limited access to rare-disease samples constrains the recognition of long-tail distributions. Given the persistent sample imbalance in medical imaging, current post-training methods still necessitate refinement to im-

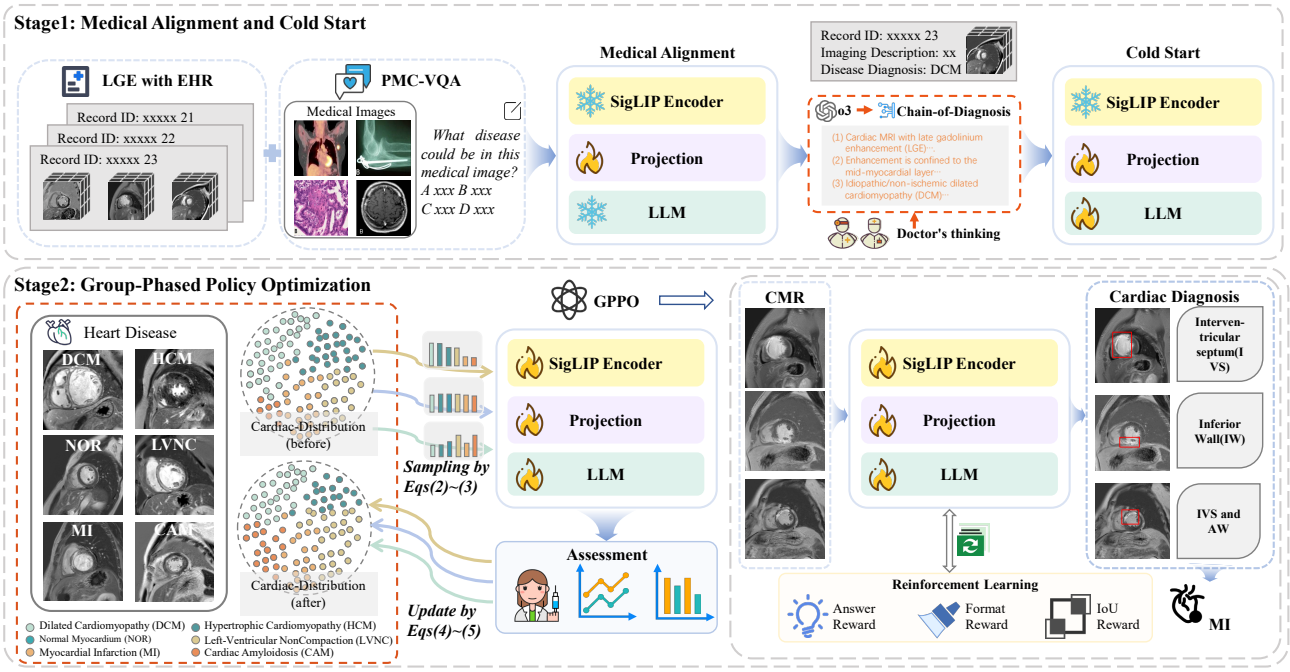


Figure 2. Two-stage training framework of CMR-RD. **Stage 1:** First perform cross-modal alignment and instruction tuning using public medical datasets (e.g., PMC-VQA and CMR image–text pairs); then introduce a small amount of supervised CoT cold-start data to teach the model to think through and reason about cardiac disease diagnosis. **Stage 2:** Introduce Group-Phased Policy Optimization and design a dynamic sampling mechanism, enabling the model to proactively and continuously optimize diagnostic performance for rare and under-represented classes under imbalanced distributions.

prove generalization across rare disease scenarios.

2.2. CMR Diagnosis and Long-Tailed Learning

CMR imaging datasets exhibit pronounced long-tailed distributions: common conditions dominate training, while rare yet high-risk categories are underrepresented, inducing majority-class bias in models [41, 42]. Existing mitigation strategies include balanced/weighted sampling, loss reweighting and invariance constraints, as well as feature-level relation modeling and virtual feature compensation [8, 17, 27, 37], which have been effective in general vision and some medical tasks. Yet these static strategies often misalign with the dynamic nature of RL post-training, creating tension between overall performance and minority-class recognition and hindering truly optimal performance across all categories. Consequently, there is a need for long-tail solutions that self-adapt to the reasoning-training phase and jointly optimize minority-class evidence alignment and reasoning-chain quality.

3. Proposed Method

To train CMR-RD, we first constructed a CMR dataset comprising scans from patients with five types of cardiac disease and from a healthy control group. As illustrated in Fig. 2, we then designed a two-stage training framework: Stage 1 carries out medical alignment, and Stage 2 enhances the model's reasoning capability. We introduce the GPPO algorithm, which improves the model's overall inference performance and increases its accuracy in identifying rare diseases.

3.1. Medical Alignment and Cold-Start

In this stage, we first perform domain alignment for the foundation VLM using the open-source medical corpus PMC-VQA [40] (227k question-answer pairs, 149k images) and CMR image-text pairs (7k). This process injects foundational medical imaging knowledge into the model and familiarizes it with the typical imaging features and professional terminology of CMR. To preserve the vision-language capabilities of the backbone and reduce training costs, both the vision backbone and the large language model (LLM) of VLM are frozen, with only the projector that maps visual features to the language space being fine-tuned.

Then, we constructed a small-scale dataset for "physician-style thinking." A cardiac radiologist selected representative cases from routine LGE-CMR reports. Subsequently, with the assistance of OpenAI's o3, the original reports were rewritten into reasoning templates that conform to the clinical decision-making process (including le-

sion observation, localization description and differential diagnosis). Finally, after quality control by clinicians, this collection served as the dataset for the cold-start [12]. At this stage, we froze the vision tower and trained only the projector and the LLM.

3.2. Group-Phased Policy Optimization

The CMR training data suffers from a severe long-tail distribution, where full-batch training can easily lead to significant factual bias. Furthermore, research has shown that a small amount of high-quality data can be sufficient to train a well-performing model [24, 39]. Therefore, we propose Group-Phased Policy Optimization, which decomposes the RL process into an online dynamic updating pipeline. By employing a dynamic sampling strategy, it adaptively adjusts the sampling weights of each disease category, thereby placing emphasis on strengthening learning for classes that are data-scarce or more difficult.

Prior and Posterior. To jointly address the data–model coupling induced by long-tail distributions and ongoing model evolution, we explicitly cast the data-sampling procedure as a Bayesian decision process under uncertainty and employ Thompson Sampling (TS) [4] to adaptively allocate per-class quotas at every training stage. Let the training set contain K disease categories:

$$\mathcal{C} = \{c_1, \ldots, c_K\}, \quad \mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}, y_i \in \mathcal{C}. \tag{1}$$

Let $\theta_c \in (0, 1)$ denote the model's true accuracy on class c, and assign a Beta prior to it for each class.

$$\theta_c \sim \mathrm{Beta}\left(\alpha_{c,0}, \beta_{c,0}\right), \quad \alpha_{c,0} = \beta_{c,0} = 1. \tag{2}$$

After the t-th training round, we record the counts on an independent test set.$n_{c,t}^{+}$ and $n_{c,t}^{-} = n_{c,t} - n_{c,t}^{+}$, where $n_{c,t}^{+}$ and $n_{c,t}^{-}$ respectively denote the numbers of correct and incorrect samples for class c at this stage.Then the posterior distribution satisfies:

$$\begin{aligned} \theta_c \mid \mathcal{D}_{\le t} &\sim \mathrm{Beta}\big(\alpha_{c,t}, \beta_{c,t}\big), \\ \alpha_{c,t} &= \alpha_{c,0} + \sum_{s=1}^{t} n_{c,s}^{+}, \\ \beta_{c,t} &= \beta_{c,0} + \sum_{s=1}^{t} n_{c,s}^{-}. \end{aligned} \tag{3}$$

Category weights generated by TS. Before the next round $t+1$ begins, independently sample for each category $\tilde{\theta}_{c,t} \sim \mathrm{Beta}\left(\alpha_{c,t}, \beta_{c,t}\right)$,and define its "difficulty" weight $w_{c,t} = 1 - \tilde{\theta}_{c,t}$.A larger weight indicates that the model is less proficient at that disease category or that its posterior is more uncertain.

Dynamic sampling. For a single-stage batch of size B, the quota for class c is given by Eq.4. We then apply a "bucket–refill" strategy: if the remaining pool size —$|\mathcal{P}_{c,t}| < \tau$ is smaller than τ (τ being a small threshold), the class is treated as a minority class and we randomly draw $min(q_{c,t}, |\mathcal{P}_{c,t}|)$ samples without removing them from the pool; otherwise, we draw $q_{c,t}$ samples without replacement and remove them from the pool. If an overall deficit still exists, the missing items are refilled one by one from the remaining majority-class pools.

$$q_{c,t} = \left\lfloor \frac{w_{c,t}}{\sum_{c' \in \mathcal{C}} w_{c',t}} B \right\rceil. \tag{4}$$

Model training and iteration. After the model is incrementally trained on $\mathcal{S}_{t+1}$, it is evaluated on the validation set to obtain $(n_{c,t+1}^{+}, n_{c,t+1}^{-})$. The Beta parameters are then updated according to Eq.3 This procedure is repeated until the convergence criterion is met. The expected weight:

$$\mathbb{E}[w_{c,t}] = 1 - \frac{\alpha_{c,t}}{\alpha_{c,t} + \beta_{c,t}}. \tag{5}$$

3.3. Reasoning enhancement

Several medical VLMs trained with GRPO can recognize lesion information in medical images. However, because VLMs are prone to hallucination, the generated textual descriptions often misalign with the actual lesion locations. To address this discrepancy, we introduce three reward functions into the reinforcement learning process, especially the IoU reward, which explicitly encourages the spatial consistency between the predicted text and the true lesion location.

1)Formatting Reward: We designed an answer template: <think>Thinking process</think>\<location\>[x1,y1,x2,y2]\</location>\<answer\>..\</answer>. When all tags are correctly matched, the reward is set to 1.5; otherwise, it is 0.

2)Answer Reward: Conditional on the formatting reward being correct, we perform answer verification. The answer within the <answer >..</answer>tags is extracted and compared with the ground truth. If the answers match, a reward of 2.0 is given; otherwise, the reward is 0. Unlike [15, 26], we opt for the model to answer with the actual disease category rather than options such as A, B, C, or D. We found during our experiments that the labels themselves contain semantic information, which is more conducive to model training.

3)IoU Reward: To further enhance the model's spatial awareness of lesions, we devise a sparse–dense hybrid reward based on the Intersection over Union (IoU) for the RL stage. The closer the bounding box predicted by the VLM is to the ground-truth lesion box, the larger the positive feedback it should receive. Given a predicted box $B_1 = [x_1, y_1, x_2, y_2]$ and a ground-truth box $B_2 = [x_1', y_1', x_2', y_2']$, IoU is defined as

$$\mathrm{IoU}(B_1, B_2) = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|}. \tag{6}$$

Because IoU is highly sparse in practice (IoU ≈ 0 for most steps), using it directly as a reward can slow RL convergence. We therefore apply a linear transformation with a threshold τ to obtain the IoU-based reward r_{IoU}:

$$r_{\text{IoU}} = \begin{cases} 0, & \text{if IoU} < \tau, \\ \dfrac{\text{IoU} - \tau}{1 - \tau}, & \text{otherwise.} \end{cases} \tag{7}$$

Let π_θ be the trainable policy (the VLM), $\pi_{\theta_{\text{old}}}$ the policy before the current update, and $G = \{g_1, \ldots, g_M\}$ a partition of the data (in our experiments each disease category constitutes one group). For every sampled image–prompt pair s_i the model generates an answer sequence $a_i \sim \pi_\theta(\cdot \mid s_i)$; the sequence is mapped to the sum of the three elementary rewards R_i.

Group-relative advantage. To optimize the policy without a dedicated value network, we employ the GRPO algorithm. For each input in a mini-batch, the model generates a group of sampled outputs $g \in G$. The baseline for each sample is derived dynamically from its group's collective performance. Specifically, we compute the group-mean reward:

$$\bar{R}_g = \frac{1}{|g|} \sum_{j \in g} R_j, \tag{8}$$

and define the group-relative advantage as the deviation of an individual reward R_i from its group mean:

$$A_i^{\text{GR}} = R_i - \bar{R}_{g(i)}. \tag{9}$$

Using this within-group baseline removes systematic bias and sharply reduces the high-variance rewards that typically hinder medical vision–language models.

For policy improvement we plug A_i^{GR} into a PPO-style clipped surrogate and add (i) a KL penalty to keep the new policy close to the old one and (ii) an entropy bonus to promote exploration. The final objective is

$$\mathcal{J}(\theta) = \mathcal{L}_{\text{clip}}^{\text{GRPO}}(\theta) + \beta_{\text{KL}}\, \text{KL}\big[\pi_{\theta_{\text{old}}} \,\|\, \pi_\theta\big] - \beta_{\text{ent}}\, \mathcal{H}\big[\pi_\theta\big], \tag{10}$$

where $\beta_{\text{KL}}, \beta_{\text{ent}} > 0$ are tunable coefficients and $\mathcal{H}[\pi_\theta]$ denotes policy entropy. During the RL fine-tuning phase, we jointly update all model components, including the vision encoder, the projection head, and the language model parameters θ. This GPPO scheme jointly enhances language accuracy and spatial reasoning while preserving the stability benefits of variance-reduced baselines.

4. Experiment

4.1. CMR-VQA

The CMR-VQA dataset consists primarily of two components. First, we randomly sampled 500 clinical cases and used OpenAI's o3 to generate question–answer pairs from the original reports for cold start; after rigorous quality control, 411 high-quality samples were retained to support CMR semantic alignment and reasoning fine-tune. The second component is training data centered on LGE-CMR, covering five cardiac diseases: cardiac amyloidosis (CAM, 146), dilated cardiomyopathy (DCM, 3,192), hypertrophic cardiomyopathy (HCM, 6,645), left ventricular noncompaction (LVNC, 465), myocardial infarction (MI, 2,833), and normal (NOR, 488). All images underwent strict quality control to exclude severe artifacts and low-quality scans. We randomly sampled 30 cases from each category to construct an independent test set. To strengthen lesion localization, we added expert-level annotations for myocardial infarction cases, using bounding boxes [x1, y1, x2, y2] to label infarct regions and microvascular obstruction. This clinically reviewed, comprehensive dataset provides a solid foundation for training CMR-RD and enhances diagnostic classification and lesion-level reasoning capabilities.

4.2. Implementation Details

CMR-RD uses Gemma-4B [32] as the base model and employs Verl [31]as the RL framework. The GPPO stage is divided into five sub-stages, each lasting two epochs, with the learning rate decaying at every stage. All images are resized to 256 × 256 and training is performed on 4 A800 GPUs, with a per-GPU batch size of 2 and 4 candidate rationales per sample of GPPO.

4.3. Metric

We quantitatively evaluate CMR-RD using three complementary metrics that capture both diagnostic and localization performance. Overall correctness is measured by Accuracy (ACC). The Area Under the ROC Curve (AUC) reflects the model's ability to rank positive and negative cases across all decision thresholds, making it less sensitive to the class imbalance in our long-tail dataset. We also report the F1-score, defined as the harmonic mean of precision and recall, to further account for distribution imbalance. To mitigate stochasticity in VLM generation and reinforcement learning, all reported results represent the mean values obtained across three independent runs.

4.4. Compared With Other VLMs

As shown in Table 1, we conducted a comparison between CMR-RD and two categories of baselines: (i) general-purpose VLMs such as GPT-4o [13], seed1.5-vl [11], Intervl3-2B [45], Qwen2.5-3B [3]; and (ii) VLMs that specifically on medical images such as Lingshu-7B [38], HuatuoGPT-Vision-7B [6], MedGemma-4B [29]. On the CMR detection task which includes six categories, our approach achieves the best performance on every evaluation metric—including ACC, AUC, and F1 score, thereby surpassing all competing models. These results underscore

the effectiveness and robustness of the proposed method for cardiac-MRI analysis.

To guarantee a fair assessment, all VLMs were queried with the same prompt, with restrictions placed on response length and output format to channel each model's reasoning capability. Test cases in which a model did not produce an explicit prediction were uniformly labeled as "Failed result" and excluded from metric calculation, thereby ensuring a consistent and unbiased evaluation of the reported performance.

To evaluate the model's reasoning capability, we employed GPT-4o and a radiology expert as impartial third-party reviewers. The procedure was as follows: we randomly sampled 50 representative cases from the test set. We then collected responses from both the baseline models and CMR-RD for each query and presented them side by side to the reviewers. GPT-4o scored the responses based on reasoning accuracy, coherence of the reasoning chain, and readability; meanwhile, the radiology expert assessed the professionalism and clinical reliability of the underlying logic. As shown in Fig.3, CMR-RD's reasoning ability surpassed that of other models according to both the AI's objective judgment and the expert's subjective evaluation.

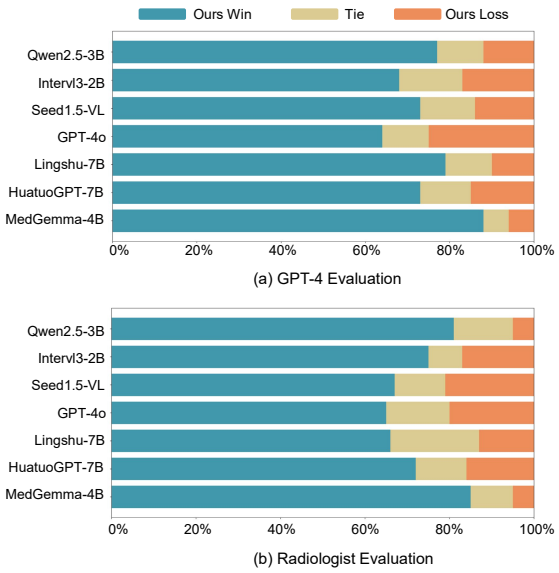


Figure 3. Reasoning process comparisons of CMR-RD with other VLMs via GPT-4 (a) and radiology expert evaluation (b).

To provide an intuitive comparison of how different medical VLMs respond to clinical queries, we present representative answers from several models in Table 3. The results show that, while these models already demonstrate a certain level of clinical reasoning, they remain susceptible to long-tail distribution and may introduce bias. When faced with clinically similar diseases, they often fail to reach accurate diagnostic conclusions. In contrast, CMR-RD adaptively learns the distribution of cardiac diseases and can still deliver good diagnostic performance even for uncommon conditions.

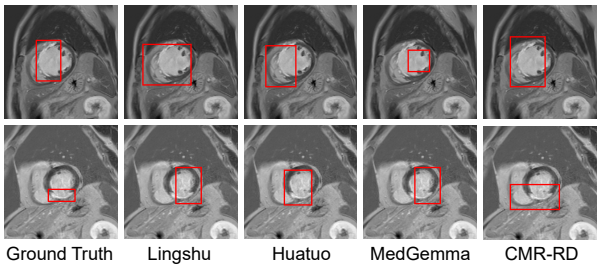


Figure 4. Lesion localization ability of different Medical VLMs. Although the positioning area of CMR-RD is larger than GT, it is more accurate than other medical VLMs.

Only HuotuoGPT and CMR-RD correctly diagnosed cardiac amyloidosis; Lingshu mistook it for myocardial infarction and MedGemma for dilated cardiomyopathy. CMR-RD cited diffuse late enhancement, ruled out ischemic, hypertrophic, and dilated cardiomyopathies, and justified its answer with a clear, checkable logic chain. Huotuo also chose the right diagnosis but used a largely enumerative approach, gave vague features, inserted factual errors (e.g., "HCM usually shows no LGE"), and even conflated HCM with arrhythmogenic right-ventricular dysplasia. Lingshu ignored the diffuse enhancement pattern typical of amyloidosis and followed an ischemic template, while MedGemma largely disregarded the critical LGE findings, focusing instead on apparent cardiac enlargement.

4.5. Localization enhancement

In the localization experiments, owing to the limited amount of annotated data, training was performed only on the MI dataset. As illustrated in Fig. 4, CMR-RD identifies and localizes lesions in LGE images more accurately than the other Medical VLMs. Nevertheless, it can be observed that although CMR-RD offers some localization capability, the VLM is not adept at target localization, so the resulting lesion bounding boxes are usually oversize.

Moreover, Table 4 further shows that incorporating the localization-learning mechanism enables CMR-RD to generate more reliable descriptions of lesion locations.In the base VLM trained without IoU constraints, lesion localization still shows deviations. For example, in this image the VLM mistakenly flags the posterior myocardial wall as a region with late enhancement, even though no lesion is actually present there. After introducing the localization-enhanced (IoU-optimized) strategy, the posterior wall is no longer mislabeled; the model now assigns high-confidence scores only to areas with true late enhancement, effectively

Table 1. Per-class performance (ACC/AUC/F1) of different VLMs on cardiac conditions.

	HCM			DCM			CAM			LVNC			MI			NOR		
Model	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
Qwen2.5-3B	0.261	0.615	0.208	0.178	0.441	0.131	0.033	0.450	0.040	0.000	0.496	0.000	0.586	0.539	0.385	0.035	0.497	0.057
MedGemma-4B	0.076	0.486	0.100	0.261	0.584	0.300	0.192	0.514	0.200	0.000	0.488	0.000	0.269	0.509	0.226	0.524	0.612	0.324
HuatuoGPT-7B	0.300	0.538	0.214	0.079	0.533	0.133	0.555	0.529	0.212	0.000	0.491	0.000	0.214	0.557	0.261	0.125	0.528	0.182
Lingshu-7B	0.214	0.544	0.251	0.192	0.512	0.192	0.300	0.406	0.161	0.000	0.500	0.000	0.300	0.447	0.180	0.100	0.467	0.103
Intervl3-2B	0.234	0.539	0.257	0.281	0.467	0.240	0.181	0.467	0.140	0.070	0.520	0.125	0.550	0.481	0.212	0.207	0.632	0.266
Seed1.5-vl	0.000	0.500	0.000	0.211	0.523	0.223	0.800	0.607	0.532	0.000	0.500	0.000	0.312	0.582	0.344	0.400	0.618	0.340
CMR-RD	0.641	0.723	0.655	0.622	0.712	0.645	0.574	0.688	0.622	0.582	0.654	0.601	0.611	0.699	0.641	0.633	0.739	0.667

Table 2. Comparison of results from different training and sampling methods.

	HCM			DCM			CAM			LVNC			MI			NOR		
Model	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1	ACC	AUC	F1
FullBatchSFT	0.421	0.512	0.477	0.311	0.523	0.353	0.100	0.498	0.132	0.000	0.500	0.000	0.322	0.531	0.374	0.389	0.541	0.412
FullBatchRL	0.664	0.721	0.683	0.613	0.697	0.635	0.100	0.457	0.151	0.070	0.495	0.115	0.672	0.733	0.698	0.330	0.576	0.341
Balance Sampling	0.512	0.643	0.511	0.483	0.625	0.527	0.533	0.621	0.566	0.523	0.638	0.547	0.561	0.656	0.545	0.571	0.662	0.619
Weighted Sampling	0.611	0.710	0.645	0.592	0.674	0.622	0.540	0.670	0.600	0.560	0.640	0.610	0.588	0.653	0.591	0.615	0.724	0.642
Thompson Sampling	0.641	0.723	0.655	0.622	0.712	0.645	0.574	0.688	0.622	0.582	0.654	0.601	0.611	0.699	0.641	0.633	0.739	0.667

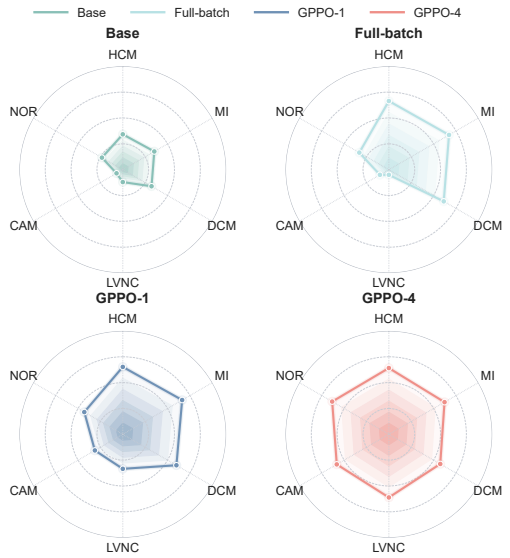


Figure 5. The Cardiac-Distribution Adaptive Learning based on TS algorithm can reduce the VLM bias introduced by long-tailed distributions.GPPO-x denotes the x-th round of GPPO.

eliminating false positives and improving both localization accuracy and clinical interpretability.

As illustrated in Fig. 5, when the VLM is trained with the plain full-batch scheme, LVNC and CAM samples make up only a tiny fraction of the dataset. The model therefore "sees" these cases only rarely within the limited number of iterations, fails to learn discriminative features for them, and the classification performance for both diseases quickly reaches a plateau. By introducing the dynamic TS strategy, the sampling probability is adjusted according to the class distribution. This significantly increases the presence of minority samples, particularly LVNC and CAM, in every mini-batch. Such repeated exposure enables the model to focus on fine-grained differences within minority classes and establishes clearer decision boundaries. Experimental results demonstrate that the TS algorithm not only markedly boosts recognition accuracy for LVNC and CAM but also improves overall robustness and generalization under long-tailed distributions.

4.6. Ablation Study

To assess the practical contribution of each training stage to the final performance of CMR-RD, we conducted a stepwise ablation study on Stage 1 and Stage 2 (results summarized in Table 5). The medical alignment and cold-start procedure first enables the model to "understand" medical images at a basic level, establishing a vision-language correspondence suited to the medical context. Cardiac distribution adaptive learning combines the strengths of supervised learning and reinforcement learning and is a key component for improving the model's diagnostic and reasoning capabilities. By dynamically switching between tasks of varying difficulty and objectives, the model continually refines its strategy and broadens its decision boundary, thereby achieving the most notable accuracy gains. For GPPO, as shown in table 2, we trained the model using Fullbatch's SFT and RL separately, and then applied three different sampling strategies to address the long-tail problem. We ultimately found that the TS algorithm effectively balances

Table 3. For the diagnoses of rare-sample diseases produced by several medical VLMs.

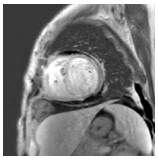
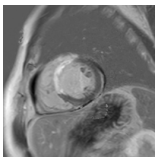
Image	
Question	Please analyze this Late Gadolinium Enhancement (LGE) image to determine whether the patient has heart disease. First, think about the reasoning process, then provide your final answer.
HuotuoGPT	<think>There appears to be a bright area within the LV wall . . . this hyperintense region suggests myocardial fibrosis... Likely cardiac amyloidosis...</think> the most likely diagnosis would be 'Cardiac amyloidosis'**<answer>CARDIAC AMELLOIDOSIS</answer>**
Lingshu	<think>...In this image, there is a distinct area of hyperenhancement visible in the myocardium, suggesting the presence of fibrotic tissue. This pattern is consistent with myocardial infarction, where necrotic tissue from a previous ischemic event leads to scar formation...Dilated cardiomyopathy would not typically show localized hyperenhancement unless secondary to other factors. Based on these observations, the most likely diagnosis is myocardial infarction. </think>**<answer>Myocardial infarction</answer>**
MedGemma	<think>... Given the enlarged heart and LGE, it is difficult to definitively diagnose without more information, but the findings are concerning for a cardiac condition. The image is not clear enough to rule out other possibilities. </think> **<answer>Dilated cardiomyopathy</answer>**
CMR-RD	<think>...It shows myocardial delayed enhancement suggestive of non-ischemic cardiomyopathy. The enhancement appears predominantly subendocardial, a pattern that can occur with myocardial infarction; however, infarction usually produces focal enhancement, whereas this image demonstrates diffuse late enhancement. Such a distribution is more consistent with cardiac amyloidosis...</think> **<answer>Cardiac amyloidosis</answer>**

Table 4. GPPO with an IoU reward improves the CMR-RD's ability to describe lesion regions, reduce the occurrence of incorrect descriptions.

Image	
Question	Please analyze this Late Gadolinium Enhancement (LGE) image to determine whether the patient has heart disease. First, think about the reasoning process, then provide your final answer.
CMR-RD w/o IoU	<think>This LGE shows delayed enhancement located on the anterior and lateral walls of the interventricular septum...**<answer>Myocardial infarction</answer>**
CMR-RD w/ IoU	<think>This is an LGE image that shows a certain degree of delayed enhancement, mainly distributed in the interventricular septum and some anterior wall positions... </think>**<location> [68, 80, 121, 184]</location> <answer>Myocardial infarction</answer>**

performance across common and rare conditions, yielding the most comprehensive accuracy across the entire dataset.

Table 5. Ablation study on CMR-RD. S1 and S2 stand for Stage1, Stage2.

Stage	Acc	F1 score	AUC
Neither S1 nor S2	0.212	0.244	0.494
Only S1	0.261	0.288	0.521
Only S2	0.493	0.512	0.683
S1 + S2	**0.610**	**0.639**	**0.703**

5. Conclusion

In this work, we propose CMR-RD, the VLM for clinical diagnostic reasoning in CMR. We design a two-stage training framework that first carries out medical alignment and then reasoning enhancement. The GPPO algorithm is leveraged to actively explore diseases on which the model performs poorly, while the TS algorithm adaptively adjusts the data ratio during each training stage to utilize the data more efficiently. Experiments on the expert-annotated test set show clear gains over other medical VLMs. Future research will focus on leveraging external tools to extend VLMs' capabilities so that we can develop intelligent agents for medical scenarios, thereby further enhancing the application value of VLMs in the medical domain.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2025YFC3508704) and the Fundamental Research Funds for the Central Universities (No. 2243200003).

References

[1] Giovanni Donato Aquaro, Carmelo De Gori, Lorenzo Faggioni, Maria Luisa Parisella, Dania Cioni, Riccardo Lencioni, and Emanuele Neri. Diagnostic and prognostic role of late gadolinium enhancement in cardiomyopathies. *European Heart Journal Supplements*, 25(Supplement_C): C130–C136, 2023. 1

[2] Seongsu Bae, Daeun Kyung, Jaehee Ryu, Eunbyeol Cho, Gyubok Lee, Sunjun Kweon, Jungwoo Oh, Lei Ji, Eric Chang, Tackeun Kim, et al. Ehrxqa: A multi-modal question answering dataset for electronic health records with chest x-ray images. *Advances in Neural Information Processing Systems*, 36:3867–3880, 2023. 2

[3] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2.5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 2, 5

[4] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24, 2011. 4

[5] Juan Manuel Zambrano Chaves, Shih-Cheng Huang, Yanbo Xu, Hanwen Xu, Naoto Usuyama, Sheng Zhang, Fei Wang, Yujia Xie, Mahmoud Khademi, Ziyi Yang, Hany Hassan Awadalla, Julia Gong, Houdong Hu, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Yu Gu, Cliff Wong, Mu-Hsin Wei, Tristan Naumann, Muhao Chen, Matthew P. Lungren, Serena Yeung-Levy, Curtis P. Langlotz, Sheng Wang, and Hoifung Poon. A clinically accessible small multimodal radiology model and evaluation metric for chest x-ray findings. *Nature Communications*, 16, 2024. 2

[6] Junying Chen, Chi Gui, Ruyi Ouyang, Anningzhe Gao, Shunian Chen, Guiming Hardy Chen, Xidong Wang, Zhenyang Cai, Ke Ji, Xiang Wan, et al. Towards injecting medical visual knowledge into multimodal llms at scale. In *Proceedings of the 2024 conference on empirical methods in natural language processing*, pages 7346–7370, 2024. 5

[7] Zhe Chen, Weiyun Wang, Hao Tian, Shenglong Ye, Zhangwei Gao, Erfei Cui, Wenwen Tong, Kongzhi Hu, Jiapeng Luo, Zheng Ma, et al. How far are we to gpt-4v? closing the gap to commercial multimodal models with open-source suites. *Science China Information Sciences*, 67(12):220101, 2024. 1, 2

[8] Wei Dai, Peilin Chen, Chanakya Ekbote, and Paul Pu Liang. Qoq-med: Building multimodal clinical foundation models with domain-aware grpo training. *arXiv preprint arXiv:2506.00711*, 2025. 3

[9] Ezequiel de La Rosa, Désiré Sidibé, Thomas Decourselle, Thibault Leclercq, Alexandre Cochet, and Alain Lalande. Myocardial infarction quantification from late gadolinium enhancement mri using top-hat transforms and neural networks. *Algorithms*, 14(8):249, 2021. 1

[10] John Giorgi, Augustin Toma, Ronald Xie, Sondra Chen, Kevin An, Grace Zheng, and Bo Wang. Wanglab at mediqa-chat 2023: Clinical note generation from doctor-patient conversations using large language models. In *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 323–334, 2023. 2

[11] Dong Guo, Faming Wu, Feida Zhu, Fuxing Leng, Guang Shi, Haobin Chen, Haoqi Fan, Jian Wang, Jianyu Jiang, Jiawei Wang, et al. Seed1. 5-vl technical report. *arXiv preprint arXiv:2505.07062*, 2025. 5

[12] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025. 4

[13] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. 1, 5

[14] Andre La Gerche, Guido Claessen, Alexander Van de Bruaene, Nele Pattyn, Johan Van Cleemput, Marc Gewillig, Jan Bogaert, Steven Dymarkowski, Piet Claus, and Hein Heidbuchel. Cardiac mri: a new gold standard for ventricular volume quantification during high-intensity exercise. *Circulation. Cardiovascular Imaging*, 6(2):329–338, 2012. 1

[15] 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. 2, 4

[16] 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. 2

[17] Hongda Liu, Yunlong Wang, Min Ren, Junxing Hu, Zhengquan Luo, Guangqi Hou, and Zhenan Sun. Balanced representation learning for long-tailed skeleton-based action recognition. *Machine Intelligence Research*, pages 1–18, 2025. 2, 3

[18] Zhengliang Liu, Yiwei Li, Peng Shu, Aoxiao Zhong, Hanqi Jiang, Yi Pan, Longtao Yang, Chao Ju, Zihao Wu, Chong Ma, et al. Radiology-gpt: a large language model for radiology. *Meta-Radiology*, page 100153, 2025. 2

[19] Yizhen Luo, Jiahuan Zhang, Siqi Fan, Kai Yang, Yushuai Wu, Mu Qiao, and Zaiqing Nie. Biomedgpt: Open multimodal generative pre-trained transformer for biomedicine. *arXiv preprint arXiv:2308.09442*, 2023. 2

[20] Lena Maier-Hein, Matthias Eisenmann, Annika Reinke, Sinan Onogur, Marko Stankovic, Patrick Scholz, Tal Arbel, Hrvoje Bogunovic, Andrew P Bradley, Aaron Carass, et al. Why rankings of biomedical image analysis competitions should be interpreted with care. *Nature communications*, 9 (1):5217, 2018. 1

[21] George Markousis-Mavrogenis, Aikaterini Giannakopoulou, Antonios Belegrinos, Maria Roser Pons, Maria Bonou, Vasiliki Vartela, Antigoni Papavasiliou, Aikaterini Christidi,

Soultana Kourtidou, Genovefa Kolovou, et al. Cardiovascular magnetic resonance imaging patterns in rare cardiovascular diseases. *Journal of Clinical Medicine*, 11(21):6403, 2022. 1

[22] Kevin Mc Namara, Hamzah Alzubaidi, and John Keith Jackson. Cardiovascular disease as a leading cause of death: how are pharmacists getting involved? *Integrated pharmacy research and practice*, pages 1–11, 2019. 1

[23] Michael Moor, Qian Huang, Shirley Wu, Michihiro Yasunaga, Cyril Zakka, Yashodhara Dalmia, Eduardo Pontes Reis, Pranav Rajpurkar, and Jure Leskovec. Medflamingo: a multimodal medical few-shot learner. *ArXiv*, abs/2307.15189, 2023. 2

[24] Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori Hashimoto. s1: Simple test-time scaling. *arXiv preprint arXiv:2501.19393*, 2025. 4

[25] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022. 1

[26] 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. 2, 4

[27] Li Pan, Yupei Zhang, Qiushi Yang, Tan Li, and Zhen Chen. Combat long-tails in medical classification with relation-aware consistency and virtual features compensation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 14–23. Springer, 2023. 2, 3

[28] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019. 1

[29] 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. 2, 5

[30] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. 1

[31] Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv: 2409.19256*, 2024. 5

[32] Gemma Team, Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, et al. Gemma 3 technical report. *arXiv preprint arXiv:2503.19786*, 2025. 5

[33] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021. 2

[34] 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. 1

[35] Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Hui Hui, Yanfeng Wang, and Weidi Xie. Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. *Nature Communications*, 16(1):7866, 2025. 2

[36] Jianyu Wu, Hao Yang, Xinhua Zeng, Guibing He, Zhiyu Chen, Zihui Li, Xiaochuan Zhang, Yangyang Ma, Run Fang, and Yang Liu. Pathvlm-r1: A reinforcement learning-driven reasoning model for pathology visual-language tasks. *arXiv preprint arXiv:2504.09258*, 2025. 2

[37] Peng Xia, Di Xu, Ming Hu, Lie Ju, and Zongyuan Ge. Lmpt: Prompt tuning with class-specific embedding loss for long-tailed multi-label visual recognition. In *Proceedings of the 3rd Workshop on Advances in Language and Vision Research (ALVR)*, pages 26–36, 2024. 2, 3

[38] Weiwen Xu, Hou Pong Chan, Long Li, Mahani Aljunied, Ruifeng Yuan, Jianyu Wang, Chenghao Xiao, Guizhen Chen, Chaoqun Liu, Zhaodonghui Li, et al. Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. *arXiv preprint arXiv:2506.07044*, 2025. 2, 5

[39] Yixin Ye, Zhen Huang, Yang Xiao, Ethan Chern, Shijie Xia, and Pengfei Liu. Limo: Less is more for reasoning. *arXiv preprint arXiv:2502.03387*, 2025. 4

[40] Xiaoman Zhang, Chaoyi Wu, Ziheng Zhao, Weixiong Lin, Ya Zhang, Yanfeng Wang, and Weidi Xie. Pmc-vqa: Visual instruction tuning for medical visual question answering. *arXiv preprint arXiv:2305.10415*, 2023. 3

[41] Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. Deep long-tailed learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(9):10795–10816, 2023. 2, 3

[42] Yundi Zhang, Paul Hager, Che Liu, Suprosanna Shit, Chen Chen, Daniel Rueckert, and Jiazhen Pan. Towards cardiac mri foundation models: Comprehensive visual-tabular representations for whole-heart assessment and beyond. *Medical image analysis*, 106:103756, 2025. 3

[43] Yuting Zhang, Kaishen Yuan, Hao Lu, Yutao Yue, Jintai Chen, and Kaishun Wu. Medtvt-r1: A multimodal llm empowering medical reasoning and diagnosis. *arXiv preprint arXiv:2506.18512*, 2025. 2

[44] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5):820–838, 2021. 1

[45] Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, et al. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. *arXiv preprint arXiv:2504.10479*, 2025. 2, 5