This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Med3D-R1: Mitigating Narrative Bias and Enhancing Reasoning Consistency in 3D Medical Vision-Language Models

Haoran Lai[1,2], Zihang Jiang[2,*], Kun Zhang[2], Qingsong Yao[3], Rongsheng Wang[2], Zhiyang He[4], Xiaodong Tao[4], Wei Wei[5], and Shaohua Kevin Zhou[1,6,7,8,9,*]

[1] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei, Anhui, China 230026
[2] Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China
[3] Renmin University of China, Beijing 100872, China
[4] Medical Business Department, iFlytek Co., Ltd., Hefei 230088, China
[5] The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230001, China
[6] Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE) Lab, YRD-RIGHT, USTC Suzhou Institute for Advanced Research, Suzhou 215123, China
[7] Jiangsu Provincial Key Laboratory of Multimodal Digital Twin Technology, Suzhou 215123, China
[8] Biomedical Basic Research Center (BBRC) of Jiangsu Province, Suzhou 215123, China
[9] State Key Laboratory of Precision and Intelligent Chemistry, Hefei 230026, China
* Co-corresponding authors: jzh0103@ustc.edu.cn, skevinzhou@ustc.edu.cn

Abstract. Developing 3D vision-language models (VLMs) with robust clinical reasoning remains challenging due to the high-dimensional complexity of volumetric imaging and biases in clinical narratives. We propose Med3D-R1, a unified framework that enhances diagnostic accuracy and interpretability for 3D CT-based abnormality diagnosis. First, we identify a structural narrative bias in standardized radiology reports and introduce an Abnormality Re-Weighting (ARW) strategy to counteract positional priors, together with a Residual Alignment Mechanism (RAM) that stabilizes cross-modal mapping between volumetric features and textual embeddings. Furthermore, we implement a reasoning-aware reinforcement learning stage with a consistency reward that encourages intermediate reasoning traces to align with report-supported evidence under report-based supervision. Evaluated on CT-RATE and the external RAD-ChestCT Multiple-choice Visual Question Answer benchmark, Med3D-R1 achieves state-of-the-art performance and improves diagnostic consistency and reasoning transparency in 3D medical VLMs. Code is available at: https://github.com/laihaoran/Med3D-R1.

Keywords: 3D Medical Vision–Language Models · Cross-Modal Alignment · Structural Narrative Bias · Reinforcement Learning · Clinical Reasoning

1 Introduction

With the increasing availability of volumetric imaging, computed tomography (CT)-based 3D diagnosis has become essential in clinical decision-making [7,16,4]. Recent progress in large language models (LLMs) has enabled 3D vision-language models (VLMs) to generate radiology-style interpretations [17,13]. However, current 3D VLMs remain far from deployable, as high-dimensional 3D visual features make cross-modal alignment unstable and final-answer-oriented training may under-constrain intermediate reasoning under report supervision.

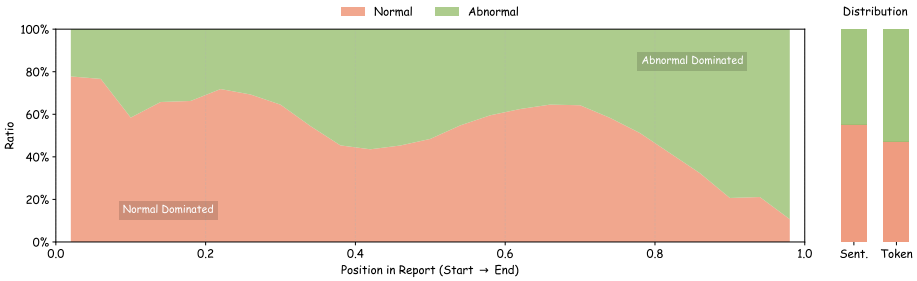

Fig. 1. Positional and overall distribution of normal and abnormal content in CT-RATE reports. Normal findings dominate early positions, whereas abnormal findings appear later, revealing an observed positional regularity that can induce prediction inertia during training.

This reporting order is clinically meaningful rather than inherently problematic, since radiologists often describe normal anatomical structures systematically before abnormal findings. Our analysis is limited to the CT-RATE training set, where this observed regularity may be learned by autoregressive models as a positional prior that favors normal findings before later abnormalities.

Existing 3D medical VLMs suffer from three major limitations. (1) The semantic gap between heterogeneous 3D volumetric features and compact textual embeddings leads to unstable cross-modal alignment [15,21]. (2) Structured radiology reports introduce an observed *structural narrative bias*: normal sentences consistently appear at early positions, while abnormalities concentrate toward later sections (Figure 1). This positional regularity may become a harmful prior during autoregressive training, resulting in "prediction inertia" and reduced abnormality sensitivity. (3) Existing medical VLMs [17,13,14] using reinforcement learning (RL) mainly optimize answer accuracy, and final-answer-oriented rewards may leave intermediate reasoning under-constrained by the available report supervision.

To address these issues, we propose **Med3D-R1**, a 3D medical VLM for report-grounded and interpretable reasoning under report-based supervision. Our contributions are threefold: **(1) Residual Alignment Mechanism (RAM)**,

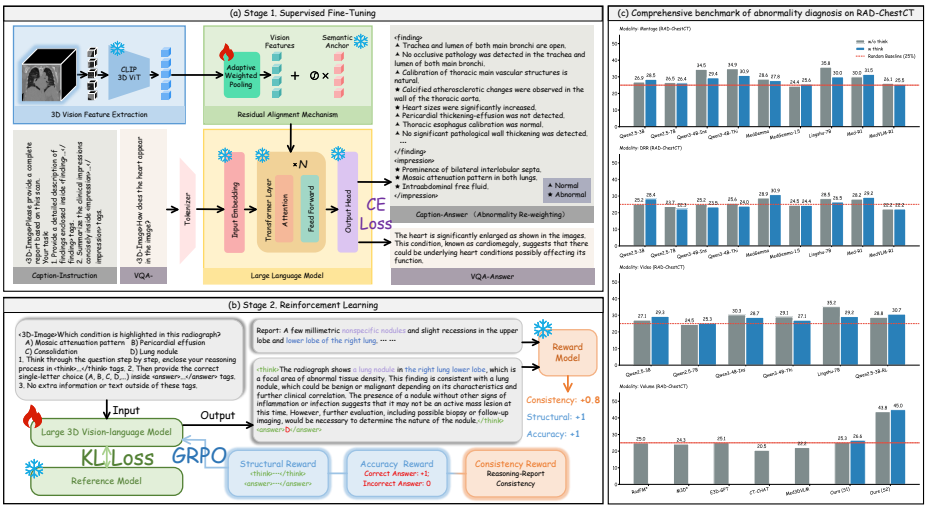

Fig. 2. Overview of Med3D-R1. (a) Stage 1 supervised fine-tuning with RAM and ARW. (b) Stage 2 reinforcement learning with structure, accuracy, and consistency rewards. (c) Comparison with representative baselines on the RAD-ChestCT Multiple-choice VQA test set, showing Med3D-R1 after SFT (S1) and RL (S2). Asterisks (*) mark outputs post-processed due to invalid choice formats. Models without valid `<think>`–`<answer>` outputs are omitted from accuracy. Compared methods include general-purpose 2D VLMs [3,26], medical 2D VLMs [19,25,13,17]—two of which also adopt GRPO-based reinforcement learning—and medical 3D VLMs [22,2,11,10,23].

a residual semantic anchor that regularizes volumetric tokens toward a reference embedding, stabilizing 3D–text projection; **(2) Abnormality Re-Weighting (ARW)**, a token-level strategy that mitigates structural narrative bias and improves abnormality sensitivity during supervised fine-tuning; and **(3) reasoning-aware reinforcement learning**, which introduces a *consistency reward* to align generated reasoning traces (`<think>`) with report-supported evidence and penalize unsupported statements. Experiments on CT-RATE [10] and RAD-ChestCT [6] demonstrate state-of-the-art abnormality diagnosis and improved report-grounded reasoning transparency over existing 3D VLMs.

2 Method

Med3D-R1 follows a two-stage training paradigm consisting of supervised cross-modal alignment and reasoning-aware RL. The model integrates three key components: a RAM for stabilizing 3D–text feature projection, an ARW strategy to mitigate structural narrative bias during SFT, and a consistency reward to regulate the model's reasoning process in RL. An overview of the framework is shown in Figure 2 (a) (b), and we detail each module below.

2.1 Residual Alignment Mechanisms

Direct projection of heterogeneous 3D CT features into the low-variance linguistic embedding space often leads to unstable optimization, even after adaptive weighted pooling [11]. To stabilize cross-modal alignment, RAM reformulates the projection as a residual interpolation toward a fixed semantic reference.

Let $\mathbf{E} \in \mathbb{R}^{V \times D}$ denote the token embedding matrix of the pretrained language model, which defines the reference semantic space. We compute a non-trainable semantic anchor as

$$\mathbf{s}_{\text{anchor}} = \frac{1}{V} \sum_{j=1}^{V} \mathbf{E}_j. \tag{1}$$

Given a projected image token $\mathbf{x}_i \in \mathbb{R}^D$, RAM produces the stabilized representation

$$\mathbf{z}_i = (1 - \sigma_i)\, \mathbf{x}_i + \sigma_i\, \mathbf{s}_{\text{anchor}}, \tag{2}$$

where the interpolation coefficient is obtained via

$$\sigma_i = \sigma(\text{MLP}([\mathbf{x}_i;\, \mathbf{s}_{\text{anchor}}]))\,, \tag{3}$$

and $\sigma(\cdot)$ denotes the sigmoid function. This residual regularization reduces feature drift and yields more stable 3D–text alignment.

2.2 Abnormality Re-Weighting

Structured radiology reports follow a clinically meaningful anatomical ordering, in which descriptions of normal structures typically appear earlier, while abnormal findings are documented later in the sequence. This observed regularity can act as a positional prior during autoregressive training, biasing generation toward normal patterns and reducing sensitivity to clinically meaningful abnormalities.

To mitigate this effect, ARW applies token-level re-weighting based on sentence-level normal/abnormal categorization. Each report is segmented into sentences using rule-based preprocessing. DeepSeek-V3 [9] is then used to label each sentence as normal, abnormal, or uncertain. The labels are used to adjust training weights and do not modify the report content. The SFT objective becomes:

$$\mathcal{L} = -\sum_{t=1}^{T} \alpha_t \log P(\hat{y}_t = y_t), \tag{4}$$

$$\alpha_t = \begin{cases} \lambda, & \text{if } y_t \text{ lies in an abnormal sentence,} \\ 1, & \text{otherwise,} \end{cases} \tag{5}$$

where $\lambda > 1$ increases the contribution of clinically informative abnormal tokens, improving sensitivity under structurally biased training data.

2.3 Rewarding Reasoning Consistency

Existing RL-based medical VLMs [13,17] primarily optimize final-answer accuracy, and such final-answer-oriented rewards may leave intermediate reasoning under-constrained, allowing generated `<think>` traces to contain statements that are not explicitly supported by the available report supervision. We introduce a consistency reward that evaluates whether the generated reasoning aligns with evidence in the reference radiology report.

Task Formulation. The abnormality diagnosis task is formulated as a Multiple-choice Visual Question Answering (MVQA) problem. Given a CT volume and a multiple-choice query, the model outputs a rationale enclosed in `<think>` tags followed by the final choice in `<answer>` tags, allowing reasoning and answer accuracy to be evaluated separately (Figure 2 (b)).

Consistency Reward. A reward model computes a continuous semantic similarity score between the generated rationale $\mathtt{t}_g$ and the reference report $\mathtt{t}_r$:

$$s = \mathrm{RM}(\mathtt{t}_g, \mathtt{t}_r). \tag{6}$$

Directly using this raw score as reward produces high-variance policy updates. To improve stability, we discretize s into the ordinal set $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ via a quantization operator $Q(\cdot)$:

$$\mathcal{R}_{\text{consistency}} = Q(s). \tag{7}$$

This variance-controlled shaping preserves the relative ordering of similarity scores while preventing unstable gradients, which in turn supports a more stable and consistent RL optimization trajectory during policy learning.

Overall Objective. The total reward combines structural, accuracy, and consistency components:

$$\mathcal{R}_{\text{total}} = \mathcal{R}_{\text{structural}} + \mathcal{R}_{\text{accuracy}} + \mathcal{R}_{\text{consistency}}. \tag{8}$$

We optimize this objective using GRPO [20] with group-normalized advantages and a KL penalty that constrains the updated policy to remain close to the SFT reference model. This formulation improves answer accuracy while encouraging the generated reasoning to remain evidence-consistent with the reference report under report-based supervision.

3 Experiment

3.1 Dataset

CT-RATE [10] The CT-RATE dataset contains 50,188 non-contrast chest CT volumes from 21,304 patients, each paired with a radiology report and 18 automatically extracted abnormality labels. We follow the official split with 47,149

volumes for training and 3,039 for validation. During SFT, we jointly train on report generation and 713,150 simple VQA pairs generated by GPT-4 [1]. These VQA samples contain only direct question-answer pairs and do not include any structured output template or reasoning traces. In the RL stage, the task is reformulated as MVQA with one true abnormality and three distractors. A curated subset covering 5% of patients (1,065 cases) is used for RL, while the official validation set is used for all evaluations.

RAD-ChestCT [6] RAD-ChestCT provides 3,630 CT volumes with 84 abnormality labels but no reports. It is used only as an external MVQA test set. For consistency with CT-RATE, 27 labels are mapped to its 18 abnormalities, following [12]. No model is trained or fine-tuned on RAD-ChestCT.

3.2 Implementation Details

We use the 3D ViT encoder from BrgSA [12] as the visual backbone and Qwen2.5-3B-Instruct [26] as the language decoder. In Stage 1, the 3D encoder and the language model are frozen; only the projection layer and the trainable components of RAM are updated during supervised fine-tuning. In Stage 2, RL is applied on the MVQA subset. The policy model—including the language model, projection layer, and 3D encoder—is jointly optimized using GRPO [20] with a KL penalty that constrains the policy to stay close to its SFT reference. RM-Mistral-7B [24] serves as the reward model for computing the consistency reward. The abnormality re-weighting factor is set to $\lambda = 1.10$.

3.3 Comparison with State-of-the-art Methods

To comprehensively evaluate Med3D-R1, we compare it with a diverse set of state-of-the-art VLMs spanning three major categories: (1) general-purpose 2D VLMs, applied to CT via montage [18] or DRR [5] projections; (2) medical-domain 2D VLMs, pretrained or aligned using radiology data but still operating on 2D representations; (3) medical 3D VLMs, which directly process volumetric CT inputs and represent the most relevant competitors. These categories cover the full spectrum from general to domain-specific and from 2D to true 3D reasoning, ensuring a fair and comprehensive comparison.

As shown in Figure 2(c), general-purpose and medical 2D VLMs achieve similar zero-shot performance when applied to CT via 2D projections. Despite relying only on montage or DRR representations rather than full volumetric inputs, these models retain strong cross-modal priors learned from large-scale image–text pretraining, which allows them to maintain step-by-step reasoning even when their overall accuracy remains modest.

Med3D-R1 (S1) also lies near this range (accuracy 26.6%) while producing structured reasoning traces on volumetric inputs. After applying reasoning-aware RL, Med3D-R1 (S2) improves substantially to 45% accuracy, surpassing both general-purpose and medical 2D VLMs (maximum 35.2%) and achieving the best combination of diagnostic accuracy and evaluated reasoning scores in this comparison.

3.4 Ablation Studies

We evaluate the contributions of the three key components of Med3D-R1: the RAM, the ARW, and the consistency reward used in the RL stage. All ablations are performed on the CT-RATE validation set: RAM and ARW are assessed through the report-generation task, while the consistency reward is evaluated using the MVQA task.

Table 1. Ablation study of RAM and ARW on CT-RATE (report-generation task). B-Avg and R-Avg denote the averaged BLEU-1/2/3/4 and ROUGE-1/2/L scores. * indicates $p < 0.05$ and † indicates $p < 0.005$ vs. the full model (paired t-test).

RAM	ARW	B-Avg	R-Avg	METEOR	CIDEr	BERT_F1	LLM_Score
✓	✓	**29.66**	**43.66**	**36.18**	**23.86**	**88.70**	**80.48**
✓	×	27.70*	42.88*	35.16*	20.22*	88.47*	79.50†
×	✓	28.56*	42.87†	35.49*	22.51*	88.41†	79.49†
×	×	26.82†	41.87†	34.43†	22.07*	88.21†	78.48†

Effectiveness of RAM and ARW. Table 1 shows that removing either module causes statistically significant performance drops across all metrics. Removing ARW yields the largest decline in CIDEr, indicating reduced sensitivity to rare abnormal terms. Removing RAM primarily affects BERT_F1 and LLM_Score, reflecting weakened semantic grounding. Disabling both modules results in the lowest overall performance, confirming that ARW improves abnormal sensitivity while RAM stabilizes global 3D–text alignment.

Table 2. Reasoning evaluation with and without the consistency reward on the CT-RATE validation set. The metrics Clinical, Clarity, Differential Diagnosis (Diff. Dx), and Conciseness (Concis) are scored on a 0–5 scale. * indicates $p < 0.0001$ compared with the baseline.

Judge	Consistency Reward	Clinical	Clarity	Diff. Dx	Concis
Radiologist	×	3.13	3.14	3.08	3.26
Radiologist	✓	**3.53***	**3.48***	**3.35***	**3.50***
Deepseek-V3	×	3.53	3.53	2.64	3.69
Deepseek-V3	✓	**4.29***	**4.08***	**3.54***	**3.90***

Effectiveness of the Consistency Reward. We further evaluate the effect of the consistency reward on report-grounded reasoning using two independent judges: a radiologist and DeepSeek-V3. As shown in Table 2, adding the reward increases the scores for Clinical, Clarity, Differential Diagnosis, and Conciseness,

with all gains statistically significant ($p < 0.0001$). These gains suggest that the consistency reward encourages generated `<think>` traces to align with report-supported evidence and discourages unsupported statements under report-based supervision.

Table 3. Evaluation of different λ on report generation. B-Avg and R-Avg denote averaged BLEU and ROUGE scores.

λ	B-Avg	R-Avg	METEOR	CIDEr	BERT_F1	LLM_Score
1.06	28.53	43.18	36.10	23.55	88.69	79.88
1.08	28.75	43.34	35.86	23.07	88.49	80.24
1.10	**29.66**	**43.32**	**36.18**	**23.86**	**88.70**	**80.48**
1.12	25.03	41.62	32,92	14.78	88.20	79.14
1.14	29.22	42.72	35.76	21.14	88.27	79.06

Effect of λ in ARW We vary the abnormal-token weight λ in ARW from 1.06 to 1.14 and report results in Table 3. Performance peaks at $\lambda = 1.10$, which provides the best balance between emphasizing abnormal content and maintaining text quality. Larger values overemphasize abnormalities and reduce generalization, while smaller values underweight abnormal terms. These results confirm that moderate re-weighting is beneficial for addressing structural narrative bias.

Table 4. Impact of reward-output quantization on consistency-reward evaluation. All metrics are scored on a 0–5 scale by the Deepseek-V3 automatic evaluator. An asterisk (*) indicates $p < 0.0001$ versus w/o quantization.

Judge	quantization	Clinical	Clarity	Diff. Dx	Concis
Deepseek-V3	×	2.82	3.28	2.31	3.27
Deepseek-V3	✓	**4.29***	**4.08***	**3.54***	**3.90***

Effectiveness of Reward Quantization Because the reward model is pretrained on general-domain text, its raw similarity scores exhibit a domain mismatch when applied to medical reasoning. This leads to biased reward magnitudes and unstable optimization during RL. Quantizing the scores into ordinal levels mitigates these effects, producing more stable updates and clear improvements across all four reasoning metrics (Table 4, $p < 0.0001$).

Reward Model Comparison We also compare different reward models in Table 5. Qwen-2.5-7B and Llama-3-8B provide inconsistent or even harmful supervision, whereas the discriminative RM-Mistral-7B yields the highest judged reasoning scores. This suggests that discriminative reward models are better suited for assessing report-supported rationales than generative LLMs.

Table 5. Average evaluation scores across different reward models in consistency reward. All metrics range from 0 to 5 and are assessed by Deepseek-V3. An asterisk (*) indicates $p < 0.0001$ compared with RM-Mistral-7B.

Reward Model	Clinical	Clarity	Diff. Dx	Conciseness	Total
Qwen-2.5-7B [26]	3.11*	3.52*	2.85*	3.42*	12.90
Llama-3-8B [8]	3.06*	3.45*	3.39*	3.06*	12.96
RM-Mistral-7B [24]	**4.29**	**4.08**	**3.54**	**3.90**	**15.81**

4 Conclusion

We propose Med3D-R1, a 3D medical VLM that improves CT abnormality diagnosis while encouraging report-grounded and evidence-consistent intermediate reasoning under report-based supervision. RAM stabilizes 3D–text alignment, ARW enhances abnormality sensitivity, and the consistency reward guides `<think>` traces toward report-supported evidence rather than unsupported statements. Experiments on CT-RATE and RAD-ChestCT demonstrate clear gains over prior 2D and 3D VLMs. Med3D-R1 provides a concise and effective framework for 3D medical vision–language modeling with more report-aligned reasoning traces. To the best of our knowledge, we are the first to identify the positional bias inherent in structured radiology reports and to introduce ARW as an effective remedy for this issue.

Acknowledgments. Supported by Natural Science Foundation of China under Grant 62271465, National Key R&D Program of China under Grant 2025YFC3408300, Natural Science Foundation of Jiangsu Province (No. BK20255001), and Suzhou Basic Research Program under Grant SYG202338.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
2. Bai, F., Du, Y., Huang, T., Meng, M.Q.H., Zhao, B.: M3d: Advancing 3d medical image analysis with multi-modal large language models. arXiv preprint arXiv:2404.00578 (2024)
3. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., Zhong, H., Zhu, Y., Yang, M., Li, Z., Wan, J., Wang, P., Ding, W., Fu, Z., Xu, Y., Ye, J., Zhang, X., Xie, T., Cheng, Z., Zhang, H., Yang, Z., Xu, H., Lin, J.: Qwen2.5-vl technical report. arXiv preprint arXiv:2502.13923 (2025)
4. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., Madabhushi, A.: Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nature reviews Clinical oncology **16**(11), 703–715 (2019)

5. Dhont, J., Verellen, D., Mollaert, I., Vanreusel, V., Vandemeulebroucke, J.: Realdrr–rendering of realistic digitally reconstructed radiographs using locally trained image-to-image translation. Radiotherapy and Oncology **153**, 213–219 (2020)
6. Draelos, R.L., Dov, D., Mazurowski, M.A., Lo, J.Y., Henao, R., Rubin, G.D., Carin, L.: Machine-learning-based multiple abnormality prediction with large-scale chest computed tomography volumes. Medical image analysis **67**, 101857 (2021)
7. Ginat, D.T., Gupta, R.: Advances in computed tomography imaging technology. Annual review of biomedical engineering **16**(1), 431–453 (2014)
8. Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., et al.: The llama 3 herd of models. arXiv preprint arXiv:2407.21783 (2024)
9. Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al.: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
10. Hamamci, I.E., Er, S., Wang, C., Almas, F., Simsek, A.G., Esirgun, S.N., Dogan, I., Durugol, O.F., Hou, B., Shit, S., et al.: Generalist foundation models from a multimodal dataset for 3d computed tomography. Nature Biomedical Engineering pp. 1–19 (2026)
11. Lai, H., Jiang, Z., Yao, Q., Wang, R., He, Z., Tao, X., Wei, W., Lv, W., Zhou, S.K.: E3d-gpt: Enhanced 3d visual foundation for medical vision-language model. arXiv preprint arXiv:2410.14200 (2024)
12. Lai, H., Jiang, Z., Yao, Q., Wang, R., He, Z., Tao, X., Wei, W., Lv, W., Zhou, S.K.: Bridged semantic alignment for zero-shot 3d medical image diagnosis. arXiv preprint arXiv:2501.03565 (2025)
13. 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)
14. Li, Y., Tang, F., Li, Y., Zhou, S.K.: Medreason-r1: Learning to reason for ct diagnosis with reinforcement learning and local zoom. arXiv preprint arXiv:2510.19626 (2025)
15. Lin, J., Xia, Y., Zhang, J., Yan, K., Lu, L., Luo, J., Zhang, L.: Ct-glip: 3d grounded language-image pretraining with ct scans and radiology reports for full-body scenarios. arXiv preprint arXiv:2404.15272 (2024)
16. Oikonomou, E.K., Marwan, M., Desai, M.Y., Mancio, J., Alashi, A., Centeno, E.H., Thomas, S., Herdman, L., Kotanidis, C.P., Thomas, K.E., et al.: Non-invasive detection of coronary inflammation using computed tomography and prediction of residual cardiovascular risk (the crisp ct study): a post-hoc analysis of prospective outcome data. The Lancet **392**(10151), 929–939 (2018)
17. 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)
18. Qi, S., Xu, C., Li, C., Tian, B., Xia, S., Ren, J., Yang, L., Wang, H., Yu, H.: Dr-mil: deep represented multiple instance learning distinguishes covid-19 from community-acquired pneumonia in ct images. Computer Methods and Programs in Biomedicine **211**, 106406 (2021)
19. 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)

20. 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)
21. Shui, Z., Zhang, J., Cao, W., Wang, S., Guo, R., Lu, L., Yang, L., Ye, X., Liang, T., Zhang, Q., et al.: Large-scale and fine-grained vision-language pre-training for enhanced ct image understanding. arXiv preprint arXiv:2501.14548 (2025)
22. Wu, C., Zhang, X., Zhang, Y., Hui, H., Wang, Y., Xie, W.: Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. Nature Communications **16**(1), 7866 (2025)
23. Xin, Y., Ates, G.C., Gong, K., Shao, W.: Med3dvlm: An efficient vision-language model for 3d medical image analysis. arXiv preprint arXiv:2503.20047 (2025)
24. Xiong, W., Dong, H., Ye, C., Wang, Z., Zhong, H., Ji, H., Jiang, N., Zhang, T.: Iterative preference learning from human feedback: Bridging theory and practice for rlhf under kl-constraint. ICML (2024)
25. Xu, W., Chan, H.P., Li, L., Aljunied, M., Yuan, R., Wang, J., Xiao, C., Chen, G., Liu, C., Li, Z., et al.: Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. arXiv preprint arXiv:2506.07044 (2025)
26. Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al.: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)