HEATMAP-INFORMED DIRECT PREFERENCE OPTIMIZATION FOR MITIGATING HALLUCINATIONS OF MEDICAL LVLMS ON SUBTLE LESIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Medical large vision-language models (Med-LVLMs) have shown strong capabilities in clinical tasks such as medical VQA and report generation, but remain prone to hallucinations—textual output inconsistent with the corresponding images, which can lead to misdiagnoses or overlooked findings. Existing Direct Preference Optimization (DPO) methods, relying on coarse-grained vision language alignment and synthetic text-based preference data, often fail to capture subtle lesions, as hallucinations frequently arise from insufficient fine-grained alignment and preference data that do not faithfully reflect visual content. To address these challenges, we propose Heatmap-informed Direct Preference Optimization (HDPO), which integrates lesion-level heatmaps to mitigate hallucinations of Med-LVLMs on subtle lesions. HDPO leverages heatmaps to guide preference data curation by explicitly modeling misdiagnosis, false positives, and false negatives, and employs a lesion-weighted DPO loss to emphasize clinically salient regions, allowing fine-grained visual-textual alignment and improved analysis of subtle lesions. Extensive experiments on four radiology datasets demonstrate that HDPO consistently outperforms the latest baselines, achieving up to 3% improvement in VQA accuracy and 2% gains in report generation metrics, particularly for subtle lesions, confirming its effectiveness in reducing hallucinations and enhancing factual accuracy in Med-LVLMs.

1 INTRODUCTION

The field of medical artificial intelligence (AI) has advanced substantially, particularly in applications such as pathology detection, interactive diagnosis, and report generation(Jin et al., 2024; Wolleb et al., 2022; Xia et al., 2024b;c; Wang et al., 2025a; Zhu et al., 2024; Ding et al., 2025; Yang et al., 2025). With the rapid emergence of large vision–language models (LVLMs), medical LVLMs (Med-LVLMs) have become a promising paradigm that integrates visual and textual information to enhance clinical understanding and reasoning(Kurz et al., 2025; Hu et al., 2024; Wang et al., 2025b; Lin et al., 2025; Liu et al., 2024). Despite their strong capabilities, Med-LVLMs remain vulnerable to hallucinations: textual descriptions inconsistent with or unsupported by medical images(Xia et al., 2024a; Zhu et al., 2024). Such errors can lead to misdiagnosis or overlooked pathologies, compromising the reliability and safety of AI-assisted healthcare(Chen et al., 2024; Gupta et al., 2024).

Recent studies have sought to address hallucinations in Med-LVLMs by investigating their causes and developing mitigation strategies, including improving vision–language alignment, fine-tuning with high-quality medical data, and employing preference optimization(Xia et al., 2024b;c; Zhu et al., 2024; Gupta et al., 2024; Ding et al., 2025; Lan et al., 2024). However, these methods often adapt techniques from natural image domains without considering challenges specific to medical images, such as subtle abnormalities and sparse visual cues. These characteristics hinder the reliable extraction of clinically relevant features and exacerbate hallucinations(Weese & Lorenz, 2016; Cheplygina et al., 2019; Zemouri et al., 2019). To bridge this gap, recent efforts have incorporated domain-specific features, such as clinical relevance scores, into Med-LVLMs(Zhu et al., 2024). Although this approach improves coarse-grained supervision, it fails to explicitly find fine-grained disease-relevant regions, risking omission of subtle but clinically significant findings. For example, as shown in Figure 1(I), when part of the cardiac contour is masked, the model generates the same response as before, indicating that lesion-level evidence is ignored and hallucinations arise. Moreover, preference data are often constructed using synthetic dispreferred answers generated by large language models(Xia et al., 2024c;b). As illustrated in Figure 1(II), real hallucinations differ markedly from synthetic ones: the model misidentifies the "right lung" instead of the correct "left lung", while the LLM-synthesized dispreferred answer ("right breast") does not capture the true visual–text misalignment. These findings underscore the need for fine-grained and lesion-aware supervision that explicitly aligns the textual findings with the corresponding visual evidence.

To address these challenges, we propose Heatmap-informed Direct Preference Optimization (HDPO), a framework that integrates lesion-level visual attribution with preference optimization to mitigate hallucinations of Med-LVLMs on subtle lesions and improve factuality in medical VQA and report generation tasks. Unlike prior methods that rely solely on medical images, HDPO incorporates lesion-aware heatmaps into preference data curation by modeling three common failure modes: misdiagnosis, false positives, and false negatives. By aligning textual keywords with their most relevant visual regions, HDPO promotes clinically meaningful vision–language associations rather than coarse and LLM-designed dispreferred answers. In addition, we calculate a heatmap alignment score between salient heatmap regions and disease-related keywords to quantify each preference pair. Finally, we introduce a Heatmap-guided preference fine-tuning strategy to scale each preference pair using the heatmap alignment score, guiding the model to prioritize clinically critical findings and reduce hallucinations from overlooked or misinterpreted lesions.

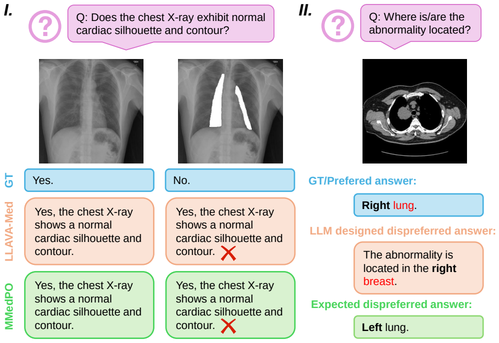

I.
Q: Does the chest X-ray exhibit normal cardiac silhouette and contour?
GT
Yes.
No.
LLAVA-Med
Yes, the chest X-ray shows a normal cardiac silhouette and contour.
Yes, the chest X-ray shows a normal cardiac silhouette and contour.
MMedPO
Yes, the chest X-ray shows a normal cardiac silhouette and contour.
Yes, the chest X-ray shows a normal cardiac silhouette and contour.
II.
Q: Where is/are the abnormality located?
GT/Prefered answer:
Right lung.
LLM designed dispreferred answer:
The abnormality is located in the right breast.
Expected dispreferred answer:
Left lung.

Figure 1: Limitations of existing Med-LVLMs. (I) When a part of the cardiac contour is masked, Med-LVLMs generate identical answers as before, indicating that models fail to capture fine-grained visual cues of the lesion. (II) Example showing that real hallucination arises from laterality error (left vs. right lung), whereas the LLM-generated dispreferred answer (right breast) fails to reflect the true visual–text mismatch.

The primary contribution of this paper is Heatmap-informed Direct Preference Optimization (HDPO), which improves factuality and reduces hallucinations in Med-LVLMs, particularly in cases involving subtle lesions. By incorporating a heatmap-guided preference data curation strategy and a lesion-weighted DPO framework that prioritizes clinically relevant regions, HDPO effectively aligns textual findings with corresponding subtle visual evidence and mitigates hallucinations of Med-LVLMs. Our method consistently outperforms the latest baselines in four radiology datasets, achieving improvements of up to 3% in VQA accuracy and 2% in report generation metrics. These findings underscore the importance of fine-grained, lesion-informed heatmaps to improve the reliability of medical vision–language models.

2 Preliminaries

In this section, we will provide a brief overview of Med-LVLMs and preference optimization.

Medical Large Vision Language Models. Med-LVLMs are specialized models designed to process medical images alongside associated textual inputs. They typically integrate a large language model (LLM) with a visual encoder that extracts features from medical images and converts them into a representation compatible with the language component. Given a medical image x_v and a clinical query x_t, the combined input is represented as $x = (x_v, x_t)$. The model then generates the response y through autoregressive decoding based on the fused multimodal input.

Preference Optimization. Preference optimization has emerged as an effective approach for fine-tuning large language models (LLMs), enabling stronger alignment between the model's output and the intended objectives. In this framework, for a given input x, the model policy π_θ defines a conditional distribution $\pi_\theta(y|x)$, where y denotes a possible textual response. A representative technique,

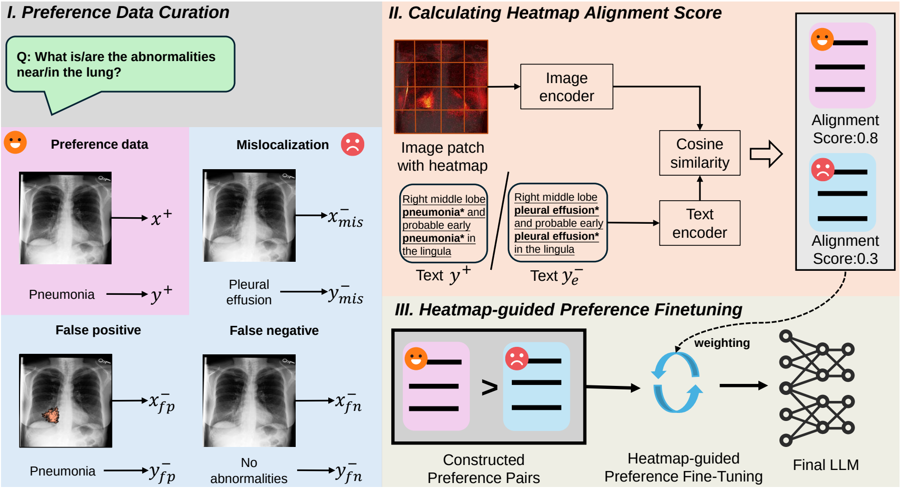

I. Preference Data Curation
Q: What is/are the abnormalities near/in the lung?
Preference data
Mislocalization
x^+
x^-_{mis}
Pneumonia
y^+
Pleural effusion
y^-_{mis}
False positive
False negative
x^-_{fp}
x^-_{fn}
Pneumonia
y^-_{fp}
No abnormalities
y^-_{fn}
II. Calculating Heatmap Alignment Score
Image patch with heatmap
Image encoder
Cosine similarity
Text encoder
Right middle lobe pneumonia* and probable early pneumonia* in the lingula
Text y^+
Right middle lobe pleural effusion* and probable early pleural effusion* in the lingula
Text y^-_e
Alignment Score:0.8
Alignment Score:0.3
III. Heatmap-guided Preference Finetuning
>
weighting
Constructed Preference Pairs
Heatmap-guided Preference Fine-Tuning
Final LLM

Figure 2: Overview of the proposed Heatmap-informed Direct Preference Optimization (HDPO) framework. HDPO integrates lesion-level heatmaps into DPO by (I) preference data curation, constructing non-preferred samples via misdiagnosis, false positives, and false negatives; (II) calculating heatmap alignment score, matching image patches with heatmaps and textual keywords to quantify each preference pair; (III) heatmap-guided preference finetuning, guiding the model to prioritize clinically critical findings and reduce hallucinations using the heatmap alignment score.

Direct Preference Optimization (DPO)(Rafailov et al., 2023), utilizes paired preference data to guide the model toward preferred behaviors. The preference data are defined as $\mathcal{D} = \left\{x^{(i)}, y_w^{(i)}, y_l^{(i)}\right\}_{i=1}^{N}$, where $y_w^{(i)}$ is the favored output and $y_l^{(i)}$ is the less desirable alternative for the same input $x^{(i)}$. The likelihood of preferring y_w over y_l is modeled as $p\left(y_w \succ y_l\right) = \sigma\left(r\left(x, y_w\right) - r\left(x, y_l\right)\right)$, where $\sigma(\cdot)$ is the sigmoid function. In DPO, optimization can be formulated as classification loss over the preference data as

$$\mathcal{L}_{DPO}\left(\pi_\theta; \pi_{\text{ref}}\right) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\left(\alpha\log\frac{\pi_\theta\left(y_w \mid x\right)}{\pi_{\text{ref}}\left(y_w \mid x\right)} - \alpha\log\frac{\pi_\theta\left(y_l \mid x\right)}{\pi_{\text{ref}}\left(y_l \mid x\right)}\right)\right]. \qquad (1)$$

Here, π_θ represents the reference policy, which is the fine-tuned LLM through supervised fine-tuning.

3 Methodology

As illustrated in Figure 2, we propose the Heatmap-informed Direct Preference Optimization (HDPO) framework to improve Med-LVLMs by using fine-grained image-text alignment via heatmaps. We first describe the preference data curation strategy by modeling common failures, such as false positives, false negatives, and misdiagnoses. Next, we detail the calculation of the heatmap alignment score by matching image patches with heatmaps and textual keywords. Finally, we introduce the heatmap-guided preference fine-tuning strategy, which guides the model to prioritize clinically critical findings and reduces hallucinations by using the heatmap alignment score.

3.1 Preference Data Curation

High-quality preference data is essential for HDPO, providing fine-grained supervision that enforces consistency between visual evidence and textual reports. For each input x, we construct preference quadruples (x^+, y^+, x_e^-, y_e^-), where (x^+, y^+) is the evidence-aligned preferred pair, and (x_e^-, y_e^-) is a dispreferred variant generated from one of three error types: misdiagnosis, false positive, or false negative. This process is guided by lesion-level heatmaps, which identify salient regions and their associated disease keywords. By replacing, masking, or removing these keywords and regions, we systematically produce dispreferred responses that misrepresent or omit critical lesion information. Compared to LLM-based curation methods(Zhu et al., 2024), this heatmap-guided perturbation provides structured, clinically meaningful supervision for preference optimization.

3.1.1 Heatmap-Guided Keyword Extraction

To explicitly capture lesion-level evidence, we use heatmaps generated by the DAug model(Jin et al., 2024) to weight visual patches. Each input image I is divided into N non-overlapping patches $\{p_1, p_2, \ldots, p_N\}$ with corresponding heatmap response $\{H(p_i)\}$. After normalization, we have

$$\tilde{H}(p_i) = \frac{H(p_i)}{\sum_{j=1}^{N} H(p_j)}, \tag{2}$$

which reflects the contribution of the patch p_i to the localization of the lesion. The weighted visual embedding is then obtained by scaling the patch features extracted from the image encoder $f_v(\cdot)$:

$$\mathbf{v} = \sum_{i=1}^{N} \tilde{H}\left(p_i\right) \cdot f_v\left(p_i\right). \tag{3}$$

In parallel, the ground-truth text answer y_{gt} is segmented into M semantic chunks $\{t_1, t_2, \ldots, t_M\}$. Each segment embedding is computed via a text encoder $f_t(\cdot)$:

$$\mathbf{u}_j = f_t(t_j), \quad j = 1, \ldots, M. \tag{4}$$

We then compute the cosine similarity between the heatmap-guided visual embedding $\mathbf{v}$ and each textual fragment $\mathbf{u}_j$

$$\mathbf{s}_j = \cos\left(\mathbf{v}, \mathbf{u}_j\right) = \frac{\mathbf{v}^T \mathbf{u}_j}{||\mathbf{v}|| \cdot |||\mathbf{u}_j|}. \tag{5}$$

The chunk with the highest value

$$k^* = \arg\max_j \mathbf{s}_j \tag{6}$$

is considered the disease keyword most strongly supported by lesion-level visual evidence.

Building on this alignment, we construct preference pairs by perturbing or replacing lesion-related keywords k^* to simulate different error types (misdiagnosis, false positive, and false negative). This design ensures that perturbations are localized, clinically significant, and directly related to visual evidence, providing high-quality supervision signals for HDPO training.

3.1.2 Heatmap-Guided Preference Data Construction

To construct preference pairs for DPO, we exploit the alignment between heatmap-salient image regions and lesion-related textual keywords. From an original image–text pair (x^+, y^+), we generate perturbed counterparts (x^-_e, y^-_e) by explicitly modeling three common error modes in medical vision–language reasoning: misdiagnosis, false positive, and false negative. These perturbations preserve the clinical fidelity of preferred data while ensuring dispreferred data misrepresent or omit lesion evidence.

Misdiagnosis. Misdiagnosis occurs when an image contains a pathological finding, but the corresponding text incorrectly labels it with an incorrect disease keyword. To synthesize a misdiagnosis sample (x^-_{mis}, y^-_{mis}), we replace $k*$ with an alternative $\tilde{k}$ drwan from the medical vocabulary $\mathcal{K}$, ensuring $\tilde{k} \neq k^*$. This produces

$$x^-_{mis} = x^+, \quad y^-_{mis} = y^+ \backslash \{k^*\} \cup \{\tilde{k}\}. \tag{7}$$

This modification preserves the sentence structure and grammaticality of the original response but semantically introduces an incorrect diagnosis attribution. As a result, the non-preferred response (x^-_{mis}, y^-_{mis}) reflects a clinically invalid diagnosis, while (x^+, y^+) remains evidence-consistent data.

False positive. False positives simulate cases in which the response asserts a finding unsupported by the image. Given a disease keyword k^*, we first use the heatmap to localize its corresponding salient region $\Omega^* \subset \mathbb{R}^{H \times W}$. To generate a false positive sample, a binary mask $M\left(\Omega^*\right)$ is applied to the salient region, producing a corrupted image while preserving the original textual description.

$$x^-_{fp} = x^+ \odot \left(1 - M\left(\Omega^*\right)\right), \quad y_{fp} = y^+. \tag{8}$$

This creates a controlled mismatch between the image and text, so the model learns to avoid generating findings that lack visual support.

False negative. False negatives represent under-reporting errors, that is, a lesion present in the image but omitted or misrepresented in the textual description. For the keyword k^* identified by heatmap-guided patch–keyword alignment, a non-preferred response is generated while the image remains unchanged $x^-_{fn} = x^+$. We exemplify two strategies: (i) neutralization, replacing the keyword lesion with a generic negation phrase, such as “no abnormalities”, to explicitly deny the finding,

$$y^-_{fn} = y^+ \setminus \{k^*\} \cup \{\text{”no abnormalities”}\}; \tag{9}$$

(ii) deletion, removing the lesion keyword from the text while retaining the rest of the report, i.e.,

$$y^-_{fn} = y^+ \setminus \{k^*\}\,. \tag{10}$$

These manipulations produce dispreferred pairs (x^-_{fn}, y^-_{fn}) in which the textual description underreports or omits the true evidence of the lesion. During preference optimization, such examples guide the model in assigning higher likelihoods to reports that accurately describe the salient findings, thus improving lesion-aware and evidence-based responses.

3.2 Calculating Heatmap Align Score

After constructing preference pairs, we quantify each pair using lesion-level heatmaps, as errors involving critical lesions should weigh more heavily in model optimization. To do this, a heatmap alignment score is calculated for each pair, measuring how well the predicted visual embeddings of the model attend to lesion-critical regions.

Formally, for a given image-text pair, we first extract the visual embedding $\mathbf{v}$ and the textual embeddings k_j corresponding to the set of differential keywords $\Delta(y^+, y^-_e)$ that distinguish the preferred response y^+ from the non-preferred response y^-_e. The heatmap alignment score is then defined as

$$w(x^+, y^+, x^-_e, y^-_e) = 1 + \lambda \cdot \max_{k_j \in \Delta(y^+, y^-_e)} \cos(\mathbf{v}, f_t(k_j)). \tag{11}$$

where λ is a scaling factor. This score quantifies the alignment between the model's attention and clinically important regions: preference pairs that involve critical lesions receive higher scores, whereas less important discrepancies are down-weighted. These weights capture lesion-specific importance at a fine-grained level, providing the foundation for a subsequent preference fine-tuning.

3.3 Heatmap-Guided Preference Fine-tuning

Once the heatmap alignment score is computed, they are integrated into the HDPO framework to guide model fine-tuning. Given a data set of preference quadruples

$$\mathcal{D} = \left\{ \left(x^+, y^+, x^-_e, y^-_e\right) | e \in \{mis, fp, fn\} \right\}, \tag{12}$$

the HDPO loss is formulated as

$$\mathcal{L}_{HDPO}\left(\pi_\theta; \pi_{\mathrm{ref}}\right) = -\mathbb{E}_{\mathcal{D}} \left[w \log \sigma \left(\alpha \log \frac{\pi_\theta\left(y^+ \mid x^+\right)}{\pi_{\mathrm{ref}}\left(y^+ \mid x^+\right)} - \alpha \log \frac{\pi_\theta\left(y^-_e \mid x^-_e\right)}{\pi_{\mathrm{ref}}\left(y^-_e \mid x^-_e\right)} \right) \right]. \tag{13}$$

This formulation ensures that errors involving heatmap-aligned lesion keywords generate stronger gradient signals, directing the model to prioritize clinical accuracy. In practice, lesion-weighted loss accelerates preference alignment for critical lesions while reducing hallucinations of overlooked or misinterpreted lesions.

4 Experiment

In this section, we conducted extensive experiments to assess the effectiveness of the proposed Heatmap-informed Direct Preference Optimization (HDPO) framework. We benchmark HDPO in four widely used radiology datasets that cover both VQA and report generation tasks and compared against recent fine-tuned Med-LVLM baselines.

4.1 Experimental Setups

Implementation Details. We employ LLaVA-Med-1.5 7B (Li et al., 2023) as the backbone model. The lesion-level heatmaps are curated using the DAug model(Jin et al., 2024). We adopt a Vision Transformer (ViT-B/16)(Dosovitskiy et al., 2020) pretrained on large-scale medical datasets as image encoder, while BioClinicalBERT(Alsentzer et al., 2019) is used as text encoder to extract keyword representations. During the preference optimization stage, we apply the LoRA fine-tuning (Hu et al., 2022) on LLaVA-Med-1.5 7B with a batch size of 4, a learning rate of 1×10^{-7}, and train for 3 epochs. All experiments are carried out using PyTorch 2.1.2 on four NVIDIA RTX A100 GPUs, with a total training time of approximately 2–3 hours.

Baseline Methods. We compare HDPO with Direct Preference Optimization (DPO)(Rafailov et al., 2023)and several recent variants. These include the self-rewarding method(Yuan et al., 2024), which generates its own responses to construct preference pairs; STLLaVA-Med(Sun et al., 2024), which refines preference selection through advanced LLM, and MMDPO(Zhu et al., 2024), which incorporates clinical relevance to improve optimization. Additionally, we benchmark three VLM preference fine-tuning methods originally developed for natural images: POVID(Zhou et al., 2024), FiSAO (Cui et al., 2024), and SIMA(Wang et al., 2024). All methods are also evaluated on models previously trained with supervised fine-tuning (SFT) using the corresponding datasets, enabling direct comparison.

Evaluation Datasets. To evaluate the effectiveness of HDPO in improving factuality and clinical reliability, we adopt four widely used medical vision–language datasets that cover VQA and report generation in X-ray and CT modalities. For VQA, we used VQA-RAD (Lau et al., 2018) and SLAKE (Liu et al., 2021), which provide fine-grained question–answer pairs linked to radiology images. For report generation, we used two large-scale chest X-ray corpora: MIMIC-CXR (Johnson et al., 2019), which includes more than 377,000 images with clinical reports, and IU-Xray (Demner-Fushman et al., 2015), a benchmark dataset with paired images and reports.

Evaluation Metrics. Following (Xia et al., 2024b;c), we evaluate the medical VQA task using accuracy and recall metrics. For the report generation task, we adopt BLEU(Papineni et al., 2002), ROUGE-L(Lin, 2004), METEOR(Banerjee & Lavie, 2005) as evaluation metrics.

4.2 Main Results

Comparison with Baseline Methods. As shown in Table 1, HDPO consistently outperforms all baselines in four radiology datasets: SLAKE, VQA-RAD, IU-Xray, and MIMIC-CXR. Without supervised fine-tuning (SFT), it achieves the highest accuracy on both open- and closed-ended questions in SLAKE (54.68 and 74.59) and VQA-RAD (38.14 and 68.53), surpassing preference optimization methods such as DPO, STLLaVA-Med, and MMedPO. It also outperforms the VLM fine-tuning methods developed for natural images, including POVID, FiSAO, and SIMA, underscoring the importance of lesion-aware supervision. For report generation, HDPO achieves substantial gains on IU-Xray, reaching 24.58 METEOR, 31.12 BLEU, and 35.98 ROUGE-L, outperforming the best baseline MMedPO (23.49, 29.52, 34.16). On the large-scale MIMIC-CXR dataset, it further sets a new state-of-the-art with 13.87 METEOR, 12.54 BLEU, and 11.59 ROUGE-L.

When combined with SFT, the improvement of HDPO becomes more pronounced, achieving the best performance in the four datasets and exceeding other preference optimization methods by a clear margin. These results validate the core design of HDPO: leveraging heatmap-guided lesion supervision in preference construction explicitly grounds textual descriptions in clinically relevant visual evidence, thereby reducing misdiagnosis, false positives, and false negatives while producing more clinically factual outputs than prior approaches.

Effect on Medical Data with Subtle Lesions To evaluate the capacity of HDPO to capture fine-grained lesion evidence, we construct subtle lesion subsets from SLAKE, VQA-RAD, IU-Xray, and MIMIC-CXR using annotated segmentation masks. Lesions occupying less than 5% of the image area are defined as small. We evaluated our method on these subsets separately. As shown in Table 2, HDPO achieves the largest performance gains in small-lesion cases relative to existing methods. This improvement reflects the effectiveness of heatmap-guided data curation and preference fine-tuning to improve the description of subtle abnormalities often overlooked by baselines. These results underscore the clinical reliability of HDPO and its ability to mitigate hallucinations arising from neglected or misinterpreted lesions.

Table 1: Performance comparison on medical VQA and report generation tasks covering four radiology datasets: SLAKE, VQA-RAD, IU-Xray, and MIMIC-CXR. Recall is reported for open-ended questions (Open), and accuracy for closed-ended questions (Closed). The BLEU denotes the average of BLEU-1/2/3/4. +SFT indicates that the model was first fine-tuned with SFT before applying the corresponding baselines.

Models	**SLAKE**		**VQA-RAD**		**IU-Xray**			**MIMIC-CXR**		
	Open	Closed	Open	Closed	METEOR	BLEU	ROUGE-L	METEOR	BLEU	ROUGE-L
LLaVA-Med v1.5	44.26	61.30	29.24	63.97	14.56	10.31	10.95	10.25	9.38	7.71
+ Self-Rewarding	42.63	61.30	33.29	64.17	14.20	10.38	10.52	10.78	9.27	7.73
+ DPO	49.30	62.02	29.76	64.70	16.08	12.95	17.13	11.19	9.45	7.80
+ POVID	52.43	70.35	31.77	65.07	20.80	24.33	30.05	11.21	9.66	7.84
+ SIMA	51.77	69.10	31.23	64.80	17.11	22.87	29.10	11.16	9.58	7.49
+ FiSAO	52.69	70.46	32.70	64.11	21.06	25.72	30.82	11.32	9.68	7.62
+ STLLaVA-Med	48.65	61.75	30.17	64.38	16.11	10.58	10.51	11.11	9.29	7.72
+ MMedPO	53.99	73.08	36.36	66.54	23.49	29.52	34.16	12.85	11.13	10.03
+ HDPO(Ours)	**54.68**	**74.59**	**38.14**	**68.53**	**24.58**	**31.12**	**35.98**	**13.87**	**12.54**	**11.59**
+ SFT	50.45	65.62	31.38	64.26	22.75	28.86	33.66	12.39	10.21	8.75
+ Self-Rewarding	50.62	65.89	32.69	65.89	22.89	28.97	33.93	12.15	10.05	8.77
+ DPO	53.50	69.47	32.88	64.33	23.07	29.97	34.89	12.37	10.38	9.10
+ POVID	52.18	70.67	32.95	64.97	23.95	29.75	34.63	11.85	10.45	9.05
+ SIMA	51.75	69.28	32.50	64.08	23.90	29.41	34.45	12.44	10.25	9.02
+ FiSAO	52.80	70.82	32.94	65.77	23.57	29.88	35.01	12.97	10.69	9.39
+ STLLaVA-Med	52.72	66.69	33.72	64.70	22.79	28.98	34.05	12.21	10.12	8.98
+ MMedPO	55.23	75.24	34.03	67.64	24.00	30.13	35.17	13.28	13.22	10.20
+ HDPO(Ours)	**55.47**	**75.17**	**35.41**	**67.54**	**24.49**	**30.37**	**35.86**	**13.69**	**13.94**	**12.97**

Table 2: Performance comparison on the full subtle-lesion subsets across four radiology datasets for medical VQA and report generation shows that HDPO achieves larger gains on subtle-lesion cases than state-of-the-art methods, highlighting its advantage in describing subtle abnormalities.

Models	**SLAKE**		**VQA-RAD**		**IU-Xray**			**MIMIC-CXR**		
	Open	Closed	Open	Closed	METEOR	BLEU	ROUGE-L	METEOR	BLEU	ROUGE-L
LLaVA-Med v1.5	43.89	59.14	31.11	62.67	12.59	9.67	9.84	8.51	7.99	6.12
+ MMedPO	51.64	70.26	34.57	64.00	21.84	27.51	31.69	10.85	8.98	8.07
+ HDPO(Ours)	**53.97**	**73.41**	**37.17**	**67.79**	**23.46**	**29.97**	**33.82**	**12.67**	**11.03**	**10.42**

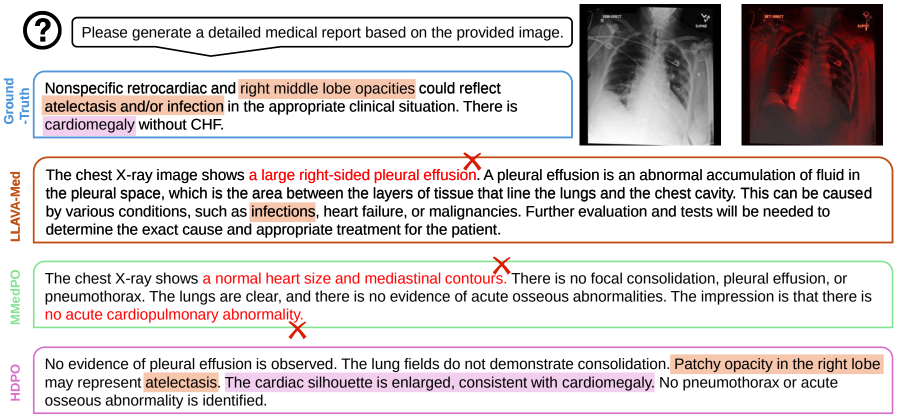
4.3 Ablation Study

Here, we performed ablation studies to evaluate the contribution of key components in our proposed framework–preference data curation, lesion-aware loss weighting, and heatmap source.

Table 3: Ablation studies for preference data curation strategy in HDPO.

misdiagnosis	**false positive**	**false negtive**	**SLAKE**	**VQA-RAD**	**IU-Xray**	**MIMIC-CXR**
✗	✗	✗	61.47	50.69	27.87	10.83
✓	✗	✗	62.57	52.41	29.11	11.73
✗	✓	✗	63.18	52.89	29.74	11.88
✗	✗	✓	63.57	53.04	30.02	12.09
✓	✓	✓	**64.64**	**53.34**	**30.56**	**12.66**

Impact of Preference Data Curation Strategy. To assess the role of different error types in preference data curation, we perform ablations by removing misdiagnoses, false positives, and false negatives. As shown in Table 3, excluding any single type consistently degrades performance across all four datasets. In particular, omitting false negatives yields the largest drop, suggesting that under-reporting errors are especially harmful in clinical scenarios. However, incorporating all three types of error achieves the highest scores, confirming that comprehensive modeling hallucination of misdiagnosis, overlooked, and overdiagnosis provides the most effective supervision in Med-LVLM to improve both VQA accuracy and report factuality.



Please generate a detailed medical report based on the provided image.

Nonspecific retrocardiac and right middle lobe opacities could reflect atelectasis and/or infection in the appropriate clinical situation. There is cardiomegaly without CHF.

The chest X-ray image shows a large right-sided pleural effusion. A pleural effusion is an abnormal accumulation of fluid in the pleural space, which is the area between the layers of tissue that line the lungs and the chest cavity. This can be caused by various conditions, such as infections, heart failure, or malignancies. Further evaluation and tests will be needed to determine the exact cause and appropriate treatment for the patient.

The chest X-ray shows a normal heart size and mediastinal contours. There is no focal consolidation, pleural effusion, or pneumothorax. The lungs are clear, and there is no evidence of acute osseous abnormalities. The impression is that there is no acute cardiopulmonary abnormality.

No evidence of pleural effusion is observed. The lung fields do not demonstrate consolidation. Patchy opacity in the right lobe may represent atelectasis. The cardiac silhouette is enlarged, consistent with cardiomegaly. No pneumothorax or acute osseous abnormality is identified.

Figure 3: Illustration of HDPO's ability of lesion-aware grounding, reduction of hallucinations, and improved clinical factuality.

Effect of Lesion-Aware Loss Weighting. We further investigate the impact of lesion-aware weighting by setting $\lambda = 0$, which reduces our loss to the standard DPO formulation without focusing on clinically salient regions.

Table 4: Ablation studies for lesion-aware weighting in HDPO.

Dataset	DPO Loss	HDPO Loss
SLAKE	62.17	**64.64**
VQA-RAD	51.46	**53.34**
IU-Xray	28.55	**30.56**
MIMIC-CXR	10.08	**12.66**

As shown in Table 4, this leads to a consistent performance degradation across all datasets, with drops of more than 2 points on SLAKE and IU-Xray. In contrast, HDPO loss achieves the best results on all benchmarks, confirming that weighting preference pairs according to the alignment of the lesion and keyword provides stronger supervision and is essential for learning reliable visual-textual associations in medical reasoning tasks.

Table 5: Ablation studies for heatmap source selection in HDPO.

Dataset	GradCAM	MedKLIP	DAug
SLAKE	58.94	60.31	**64.64**
VQA-RAD	47.37	49.87	**53.34**
IU-Xray	24.91	26.76	**30.56**
MIMIC-CXR	8.67	9.32	**12.66**

Effect of Heatmap Source. We also study the effect of different heatmap sources on HDPO. As shown in Table 5, using Grad-CAM(Selvaraju et al., 2017), a CAM-based method, produces the weakest performance due to its coarse and often noisy activation maps. MedKLIP(Wu et al., 2023), which leverages attention-based attribution, provides stronger signals but still fails to locate the fine-grained lesion. In contrast, DAug-generated heatmaps deliver the most precise lesion-level supervision, resulting in significant improvements across all datasets (e.g., SLAKE 64.64 vs. 58.94 with GradCAM). Overall, the comparison demonstrates that accurate lesion attribution is essential for HDPO: while coarse heatmaps can only provide weak guidance, fine-grained lesion-aware maps allow the model to learn precise and clinically meaningful visual-textual associations, thereby improving both medical VQA and report generation tasks.

4.4 Case Study

To further illustrate the effectiveness of HDPO in reducing hallucinations on subtle lesions and improving visual-text alignment, we present the representative case study on a chest X-ray image.

Visualization and Grounding. In Figure 3, we show the original chest radiograph images alongside heatmaps. The heatmaps highlight lesion-aware regions, such as atelectasis, infection, and cardiomegaly, which correspond to the critical findings mentioned in the ground-truth reports. This visualization demonstrates the ability of a lesion-aware heatmap to ground textual outputs in the correct visual evidence, enhancing both interpretability and clinical reliability.

Reduction of Hallucinations. In Figure 3, LLAVA-Med erroneously reported a "large right-sided pleural effusion" while failing to detect the cardiomegaly present. Similarly, MMedPO misdiagnosed "normal heart size and mediastinal contours" and overlooked lobe opacities. In contrast, HDPO avoided these hallucinations, accurately reporting "The cardiac silhouette is enlarged, consistent with cardiomegaly" and correctly identifying atelectasis from the observed opacity, supported by a precise lesion heatmap. Across multiple cases, HDPO consistently reduces spurious findings, particularly for subtle abnormalities, demonstrating its ability to capture clinical visual evidence.

Improved Clinical Factuality. HDPO further improves the specificity and clinical accuracy of its outputs. As shown in Figure 3, for cases with small lobe opacities, HDPO provides precise descriptions specifying the exact location, for example, "Patchy opacity in the right lobe," while baseline methods often omit such details or produce vague, potentially misleading statements. By aligning the textual output with salient image regions, HDPO ensures accurate capture of critical diagnostic information, enhancing its utility for clinical decision support.

5 REALTED WORK

Factuality in Med-LVLMs. The rapid development of Large Vision Language Models (LVLMs) has accelerated progress in medical applications(Kurz et al., 2025; Xia et al., 2024c; Lin et al., 2025), demonstrating strong capabilities across diverse imaging modalities and clinical tasks(Ding et al., 2025; Yang et al., 2025; Wang et al., 2025a). Despite these advances, existing Med-LVLMs often struggle with factual consistency(Zhu et al., 2024; Chen et al., 2024; Gupta et al., 2024), failing to reason effectively in complex medical scenarios and generating hallucinated outputs unsupported by the corresponding images. Such errors compromise the reliability and safety of AI-assisted healthcare, potentially causing misdiagnoses or missed pathologies. Recent benchmarking studies(Xia et al., 2024b;c; Zhu et al., 2024; Kurz et al., 2025) have highlighted these ongoing challenges in tasks such as medical VQA and report generation.

Preference Optimization in Med-LVLMs. Preference optimization is essential to develop effective, safe, and trustworthy models while mitigating hallucinations in medical applications (Gorbatovski et al., 2024; Gao et al., 2023; Xu et al., 2024). Standard approaches, such as RLHF (Ouyang et al., 2022), rely on human-labeled preference data to train a reward model, but this adds complexity and potential instability. Direct Preference Optimization (DPO) (Rafailov et al., 2023) simplifies training by fine-tuning directly on pairwise preference data without explicit reward modeling. MMedPO (Zhu et al., 2024) extends this to medical models using clinically relevant preferences, but focuses on coarse textual and visual alignment and can miss fine-grained pathological regions. To overcome this, we propose HDPO, a preference optimization framework designed to capture detailed disease-specific characteristics in medical images.

Lesion-Aware-Heatmap Supervision for Assistance in Medical Imaging. Medical image analysis depends on subtle pathological cues, but global visual or textual preferences often overlook critical lesions, resulting in hallucinations in Med-LVLM output. Incorporating visual attributions, such as class activation (Selvaraju et al., 2017) or attention maps (Wu et al., 2023) as supervision directs models to relevant regions, improving accuracy and interpretability. However, these approaches typically produce coarse heatmaps. To address this, DAug (Zhu et al., 2024) used generative models to generate lesion-level heatmaps. Motivated by the ability of lesion-level annotations to reduce localized hallucinations, we propose HDPO, which integrates lesion-aware supervision into preference data curation and fine-tuning to improve diagnostic factuality in Med-LVLMs.

6 CONCLUSION

In this work, we introduce Heatmap-informed Direct Preference Optimization (HDPO) to address the persistent issue of hallucinations in medical large vision-language models. By incorporating lesion-level heatmaps into both preference data construction and optimization, HDPO achieves fine-grained vision-language alignment and effectively reduces errors such as misdiagnosis, false positives, and false negatives. Extensive experiments on four radiology datasets demonstrate the effectiveness of our approach. HDPO consistently outperforms the latest baselines, particularly in medical data with subtle lesions. Beyond mitigating hallucinations, HDPO highlights the potential of integrating visual interpretability signals into preference-based training, paving the way for more reliable and reliable medical AI systems.

Acknowledgement

This work is supported by the National Key R&D Program of China (Grant No. 2024YFF0505703).

LLM Use Statement

During the preparation of this paper, we used large language models (LLMs) solely for language polishing. All content in the paper was developed and verified by the authors.

References

Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*, 2019.

Satanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pp. 65–72, 2005.

Jiawei Chen, Dingkang Yang, Tong Wu, Yue Jiang, Xiaolu Hou, Mingcheng Li, Shunli Wang, Dongling Xiao, Ke Li, and Lihua Zhang. Detecting and evaluating medical hallucinations in large vision language models. *arXiv preprint arXiv:2406.10185*, 2024.

Veronika Cheplygina, Marleen De Bruijne, and Josien PW Pluim. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, 54:280–296, 2019.

Chenhang Cui, An Zhang, Yiyang Zhou, Zhaorun Chen, Gelei Deng, Huaxiu Yao, and Tat-Seng Chua. Fine-grained verifiers: Preference modeling as next-token prediction in vision-language alignment. *arXiv preprint arXiv:2410.14148*, 2024.

Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2015.

Meidan Ding, Jipeng Zhang, Wenxuan Wang, Haiqin Zhong, Xiaoqin Wang, Xinheng Lyu, Wenting Chen, and Linlin Shen. Eagle: Expert-guided self-enhancement for preference alignment in pathology large vision-language model. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 14603–14619, 2025.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, pp. 10835–10866. PMLR, 2023.

Alexey Gorbatovski, Boris Shaposhnikov, Alexey Malakhov, Nikita Surnachev, Yaroslav Aksenov, Ian Maksimov, Nikita Balagansky, and Daniil Gavrilov. Learn your reference model for real good alignment. *arXiv preprint arXiv:2404.09656*, 2024.

Shailja Gupta, Rajesh Ranjan, and Surya Narayan Singh. A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. *arXiv preprint arXiv:2410.12837*, 2024.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022.

Yutao Hu, Tianbin Li, Quanfeng Lu, Wenqi Shao, Junjun He, Yu Qiao, and Ping Luo. 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.

Ying Jin, Zhuoran Zhou, Haoquan Fang, and Jenq-Neng Hwang. Daug: Diffusion-based channel augmentation for radiology image retrieval and classification. *arXiv preprint arXiv:2412.04828*, 2024.

Alistair EW Johnson, Tom J Pollard, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Yifan Peng, Zhiyong Lu, Roger G Mark, Seth J Berkowitz, and Steven Horng. Mimic-cxr-jpg, a large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*, 2019.

Christoph F Kurz, Tatiana Merzhevich, Bjoern M Eskofier, Jakob Nikolas Kather, and Benjamin Gmeiner. Benchmarking vision-language models for diagnostics in emergency and critical care settings. *npj Digital Medicine*, 8(1):423, 2025.

Wei Lan, Wenyi Chen, Qingfeng Chen, Shirui Pan, Huiyu Zhou, and Yi Pan. A survey of hallucination in large visual language models. *arXiv preprint arXiv:2410.15359*, 2024.

Jason J Lau, Soumya Gayen, Asma Ben Abacha, and Dina Demner-Fushman. A dataset of clinically generated visual questions and answers about radiology images. *Scientific data*, 5(1):1–10, 2018.

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.

Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pp. 74–81, 2004.

Tianwei Lin, Wenqiao Zhang, Sijing Li, Yuqian Yuan, Binhe Yu, Haoyuan Li, Wanggui He, Hao Jiang, Mengze Li, Xiaohui Song, et al. Healthgpt: A medical large vision-language model for unifying comprehension and generation via heterogeneous knowledge adaptation. *arXiv preprint arXiv:2502.09838*, 2025.

Bo Liu, Li-Ming Zhan, Li Xu, Lin Ma, Yan Yang, and Xiao-Ming Wu. Slake: A semantically-labeled knowledge-enhanced dataset for medical visual question answering. In *2021 IEEE 18th international symposium on biomedical imaging (ISBI)*, pp. 1650–1654. IEEE, 2021.

Bo Liu, Ke Zou, Liming Zhan, Zexin Lu, Xiaoyu Dong, Yidi Chen, Chengqiang Xie, Jiannong Cao, Xiao-Ming Wu, and Huazhu Fu. Gemex: A large-scale, groundable, and explainable medical vqa benchmark for chest x-ray diagnosis. *arXiv preprint arXiv:2411.16778*, 2024.

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.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, 2002.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

Guohao Sun, Can Qin, Huazhu Fu, Linwei Wang, and Zhiqiang Tao. Stllava-med: Self-training large language and vision assistant for medical. *arXiv e-prints*, pp. arXiv–2406, 2024.

Jinhong Wang, Tajamul Ashraf, Zongyan Han, Jorma Laaksonen, and Rao Mohammad Anwer. Mira: A novel framework for fusing modalities in medical rag. *arXiv preprint arXiv:2507.07902*, 2025a.

Pengyu Wang, Shuchang Ye, Usman Naseem, and Jinman Kim. Mrgagents: A multi-agent framework for improved medical report generation with med-lvlms. *arXiv preprint arXiv:2505.18530*, 2025b.

Xiyao Wang, Jiuhai Chen, Zhaoyang Wang, Yuhang Zhou, Yiyang Zhou, Huaxiu Yao, Tianyi Zhou, Tom Goldstein, Parminder Bhatia, Furong Huang, et al. Enhancing visual-language modality alignment in large vision language models via self-improvement. *arXiv preprint arXiv:2405.15973*, 2024.

Jürgen Weese and Cristian Lorenz. Four challenges in medical image analysis from an industrial perspective, 2016.

Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C Cattin. Diffusion models for medical anomaly detection. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 35–45. Springer, 2022.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. Medklip: Medical knowledge enhanced language-image pre-training in radiology. *arXiv preprint arXiv:2301.02228*, 2023.

Peng Xia, Ze Chen, Juanxi Tian, Yangrui Gong, Ruibo Hou, Yue Xu, Zhenbang Wu, Zhiyuan Fan, Yiyang Zhou, Kangyu Zhu, et al. Cares: A comprehensive benchmark of trustworthiness in medical vision language models. *Advances in Neural Information Processing Systems*, 37:140334–140365, 2024a.

Peng Xia, Kangyu Zhu, Haoran Li, Tianze Wang, Weijia Shi, Sheng Wang, Linjun Zhang, James Zou, and Huaxiu Yao. Mmed-rag: Versatile multimodal rag system for medical vision language models. *arXiv preprint arXiv:2410.13085*, 2024b.

Peng Xia, Kangyu Zhu, Haoran Li, Hongtu Zhu, Yun Li, Gang Li, Linjun Zhang, and Huaxiu Yao. Rule: Reliable multimodal rag for factuality in medical vision language models. *arXiv preprint arXiv:2407.05131*, 2024c.

Shusheng Xu, Wei Fu, Jiaxuan Gao, Wenjie Ye, Weilin Liu, Zhiyu Mei, Guangju Wang, Chao Yu, and Yi Wu. Is dpo superior to ppo for llm alignment? a comprehensive study. *arXiv preprint arXiv:2404.10719*, 2024.

Yan Yang, Xiaoxing You, Ke Zhang, Zhenqi Fu, Xianyun Wang, Jiajun Ding, Jiamei Sun, Zhou Yu, Qingming Huang, Weidong Han, et al. Spatio-temporal and retrieval-augmented modelling for chest x-ray report generation. *IEEE Transactions on Medical Imaging*, 2025.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. Self-rewarding language models. *arXiv preprint arXiv:2401.10020*, 3, 2024.

Ryad Zemouri, Noureddine Zerhouni, and Daniel Racoceanu. Deep learning in the biomedical applications: Recent and future status. *Applied Sciences*, 9(8):1526, 2019.

Yiyang Zhou, Chenhang Cui, Rafael Rafailov, Chelsea Finn, and Huaxiu Yao. Aligning modalities in vision large language models via preference fine-tuning. *arXiv preprint arXiv:2402.11411*, 2024.

Kangyu Zhu, Peng Xia, Yun Li, Hongtu Zhu, Sheng Wang, and Huaxiu Yao. Mmedpo: Aligning medical vision-language models with clinical-aware multimodal preference optimization. *arXiv preprint arXiv:2412.06141*, 2024.