# Toward Reliable Dermatological Medical Report Generation via Knowledge-Enhanced Structured Multi-modal Reasoning


1st Feng Tan
*ADSPLAB, School of ECE*
*Shenzhen Graduate School, Peking University*
Shenzhen, China
https://orcid.org/0009-0005-6185-5090

2nd Yuexian Zou
*ADSPLAB, School of ECE*
*School of ECE, Shenzhen Graduate School, Peking University*
Shenzhen, China
https://orcid.org/0000-0001-9999-6140



***Abstract*—Accurate and reliable dermatological medical report generation (DMRG) is essential for clinical decision support and improved healthcare accessibility. However, most existing models rely on direct neural mapping from images to text, leading to opaque reasoning and reports that often violate established medical knowledge. To address these limitations, we propose a Knowledge-Enhanced Structured Multi-Modal Reasoning (KE-SMMR) framework that replaces opaque black-box mappings with a structured collaborative paradigm between Multi-Modal Large Language Models (MLLMs) and Large Language Models (LLMs). This collaborative process is governed by three complementary constraints: (1) Perceptual Constraint for auditable and structured observations; (2) Dermatological Knowledge Constraint to restrict reasoning within medically valid boundaries; and (3) Inference Constraint to align multi-stage reasoning with clinical practice. Furthermore, we introduce Triple-Chain Reasoning (TCR), which aligns visual evidence with morphological, anatomical, and pathophysiological knowledge for multi-granularity verification. Extensive experiments demonstrate that our KE-SMMR framework significantly improves report accuracy and medical consistency, offering a knowledge-enhanced reasoning paradigm for reliable DMRG, particularly in data-scarce settings.**



***Index Terms*—dermatological medical report generation, explainable medical AI, knowledge-enhanced reasoning, vision-language models**


## I. Introduction

Skin diseases affect billions of people worldwide and constitute the fourth leading cause of non-fatal disease burden, imposing substantial pressure on global healthcare systems [1], [2]. Despite this prevalence, access to dermatological expertise remains highly uneven, particularly in low-resource and remote regions, where specialist shortages are severe [3]. In this context, automated dermatological diagnosis and medical report generation have emerged as critical research directions in medical image intelligence, with the potential to support clinical decision-making and improve healthcare accessibility [2], [4], [5]. Early research in dermatological artificial intelligence primarily focused on lesion-level image classification, achieving near-expert performance for specific conditions such as melanoma and basal cell carcinoma [4], [6]. However, these methods provide only coarse diagnostic labels and fail to meet clinical requirements for comprehensive, standardized medical reporting. Moreover, their heavy reliance on large-scale annotated datasets and limited generalization ability restrict practical deployment in real-world clinical scenarios. Recent advances in Multi-Modal Large Language Models (MLLMs) have enabled dermatological medical report generation (DMRG) by directly mapping skin lesion images to free-text descriptions [1], [5], [7]–[10]. While such approaches produce richer clinical narratives, they typically operate as black-box systems that entangle visual perception, diagnostic reasoning, and language generation within implicit representations. This paradigm not only demands extensive high-quality paired image–text data—often scarce in medical domains—but also introduces critical reliability issues [7]. In particular, prior-driven hallucinations frequently occur, where DMRG models generate medically incorrect statements based on dataset biases rather than observable visual evidence. The lack of explicit diagnostic structure and traceable reasoning of training DMRG models further limits medical consistency and undermines clinical trust [7], [11], [12]. As reliability and interpretability become critical concerns in medical AI, these limitations expose a fundamental gap in existing DMR systems [12]. Accurate reporting requires fine-grained interpretation of visual evidence, together with dermatological knowledge. Although recent studies have attempted to incorporate medical knowledge or reasoning prompts, most treat knowledge as an auxiliary generative signal rather than a strict constraint, leaving diagnostic reasoning insufficiently verifiable [13], [14].

Besides, some studies have shown that decomposing tasks through the collaboration of MLLMs and large language models (LLMs) can effectively improve zero-shot inference performance in medical visual question answering [15], [16]. However, current DMRG models focus more on fine-tuning models and rarely applies this framework. To bridge these gaps, we propose a Knowledge-Enhanced Structured Multi-modal Reasoning (KE-SMMR) framework for reliable DMRG. Our KE-SMMR framework introduces a collaborative task-allocation paradigm, specifically leveraging MLLMs and LLMs collaborative paradigm for knowledge-enhanced diagnostic inference.

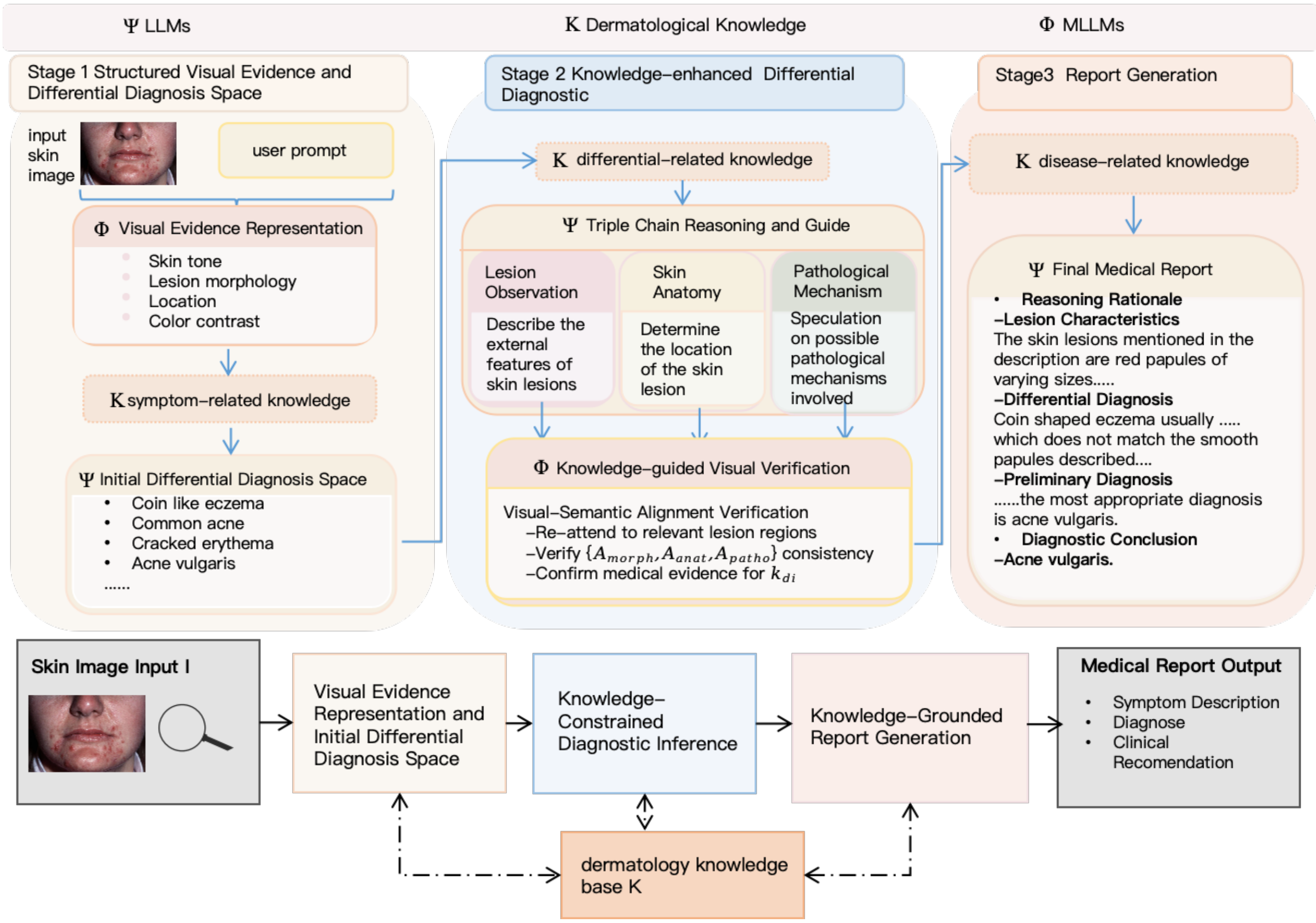


Fig. 1. Overview of our KE-SMMR framework for reliable dermatological medical report generation (DMRG).

Our core contributions are fourfold:

- **Structured Reasoning Paradigm:** We establish a framework called KE-SMMR. By implementing tri-level constraints (Perceptual, Dermatological Knowledge, and Inference) in our KE-SMMR framework, we significantly improve diagnose accuracy and enforce dermatological consistency.
- **Triple-Chain Reasoning (TCR):** We propose a clinically grounded reasoning paradigm that aligns visual evidence with morphological, anatomical, and pathophysiological knowledge, enabling multi-level diagnostic verification.
- **Curated Dermatological Knowledge Base:** We construct a comprehensive dermatological knowledge base covering over 250 skin diseases. Leveraged from authoritative medical sources, this dataset underwent rigorous cleaning and semantic segmentation to provide dermatological knowledge constraint for reasoning.
- **Extensive Experimental Validation:** Through extensive quantitative and qualitative experiments, our approach achieves superior accuracy and clinically consistent report generation.

## II. Related Work

### A. Dermatological Medical Report Generation

In DMRG, traditional deep learning has largely relied on end-to-end "black-box" generation and large-scale image-text paired data, without targeting reliability and interpretability [1], [5], [10]. Recent studies have highlighted the critical importance of reliability in DMRG. Some reasoning-centric models have integrated standardized Chain-of-Thought (CoT) narratives to mirror dermatological expert workflow [16]. For instance, SkinGPT-R1 utilizes a three-layer narrative structure to provide verifiable and transparent diagnostic explanations [17]. DermoGPT employs a Morphologically-Anchored Visual-Inference-Consistent (MAVIC) reinforcement learning reward, which penalizes logical disconnects between generated morphological descriptions and final diagnostic conclusions [18].

### B. Knowledge-Enhanced Reasoning in Medical Vision-Language Models

Specialized medical knowledge is essential for bridging the semantic gap between low-level visual patterns and high-level clinical language [11], [19]. Current research explores several structured knowledge-enhancement routes to guide model reasoning: (1)Ontology-Guided Reasoning: Framework utilize hierarchical clinical ontologies to decompose unstructured narratives into multi-aspect sub-captions, ensuring the model prioritizes diagnostically significant features during cross-modal alignment [11]. (2)Knowledge Graph (KG) Integration: Models integrate disease-specific KGs to unlock latent expert knowledge within Large Language Models (LLMs), using graph nodes to model complex inter-disease relationships and facilitate fine-grained feature extraction [20], [21]. (3)Retrieval-Augmented Reasoning: Models enhance context-aware reasoning by dynamically retrieving similar clinical evidence [22].

## III. Method

### A. Problem Definition

Given a dermatological image $I$ and a diagnostic query $Q$, our goal is to generate a reliable dermatological medical diagnostic report $Y$ that is grounded in visual evidence, consistent with dermatological knowledge, and supported by transparent diagnostic reasoning.

### B. Overview of the KE-SMMR Framework

As illustrated in Fig.1, our KE-SMMR framework is a three-stage pipeline. Dermatology medical knowledge base is noted as $K$. By integrating knowledge constraints alongside reasoning guidance from the LLM, KE-SMMR significantly elevates the zero-shot performance of the MLLM in DMRG without any fine-tuning. Specifically, we implement a structured task-allocation strategy: the MLLM $\Phi$ is used for visual observation grounded tasks, while the LLM $\Psi$ for high-level reasoning and knowledge integration. In Stage 1, we leverage $\Phi$ to extract structured visual evidence and derive initial diagnostic hypotheses, which serve as the grounded basis for further reasoning. The visual evidence perceived by $\Phi$ are the perceptual constraints $C_p$. The knowledge constraint $C_k$ is introduced via symptom-related knowledge and user prompts to delineate an initial differential diagnosis space $D$. This space serves as the medical logical boundary, ensuring subsequent reasoning remains grounded in relevant dermatological categories.

$$\mathcal{V}, \mathcal{D} = \Psi\left(\Phi(\mathcal{I}), \mathcal{Q} \mid \mathcal{K}_{\text{symptom}}\right) \tag{1}$$

Stage 2 centers on the interaction between knowledge $C_k$ and inference constraints $C_i$. Leveraging differential-related knowledge, $C_i$ is operationalized through a Triple Chain Reasoning(see more details in next section). In the final stage, we integrate disease-related knowledge, our KE-SMMR framework culminates in a structured medical report, where $C_i$ ensures that the rationale, lesion characteristics, and differential analysis are presented in a logically consistent, human-readable format, leading to a robust final diagnostic conclusion.

$$\mathcal{Y} = \Psi\left(\mathcal{V}, \mathcal{D}, \mathcal{Q}, \mathcal{K}_{\text{disease}}\right) \tag{2}$$

Across three stages, domain knowledge constraints $C_k$ focus on symptoms, differential diagnosis, and disease-specific information, respectively. Knowledge is retrieved from the dermatology knowledge base $K$ using queries generated by $\Psi$ tailored to each stage's context:

$$\mathcal{K}_{\text{stage}} = \text{Retrieve}(\mathcal{K}, \Psi(\text{context}_{\text{stage}})) \tag{3}$$

$C_i$ are reflected in the inference path from perception to differential diagnosis to report generation based on clinical diagnosis.

### C. Knowledge Base Construction

Dermatology medical knowledge base $K$ is a knowledge source to constrain, guide, and enhance the reasoning process. $K$ contains approximately 250 types of skin diseases, which we crawled from authoritative medical websites and have undergone cleaning, structuring, and manual screening to ensure the quality of source data. $K$ is standardized into a structured schema to support retrieval and reasoning, each entry is structured as: disease name, symptoms, morphology, anatomical location, pathophysiology and treatment. Specifically, we design instruct prompt to enable LLMs to structure and clean the knowledge corpus, and then conduct manual review. We ensure reliability by sourcing from established medical references. Given the visual query, we use semantic similarity (e.g., embedding-based retrieval) to retrieve top-k relevant disease entries as candidate constraints.

### D. Explicit Knowledge Alignment via Triple-Chain Reasoning

We propose Triple-Chain Reasoning (TCR), an explicit verification mechanism that utilizes knowledge as a constraint. In this paradigm, differential-related knowledge entities enhance the targeted extraction of clinical evidence, verifying a alignment between visual primitives and medical semantics. Specifically, TCR synchronizes visual evidence with the knowledge across three granularities: $\mathcal{A}_{\text{TCR}} = \{\mathcal{A}_{\text{morph}}, \mathcal{A}_{\text{anat}}, \mathcal{A}_{\text{patho}}\}$. Morphological chain $\mathcal{A}_{\text{morph}}$ indicating whether there is visual evidence consistent with dermatological descriptors observed by $\Phi$. Anatomical chain $\mathcal{A}_{\text{anat}}$ check if the anatomical position is aligned. Pathophysiological chain $\mathcal{A}_{\text{patho}}$ associate visual evidence with potential biological mechanisms to ensure that the extracted visual evidence is medically reasonable. In TCR, dermatological knowledge actively participates in a feedback loop that verifies visual evidence. Specifically, disease-related knowledge $K_{d_i}$ is used to enhance $\Phi$ to re-attend to relevant image regions and evaluate the alignment between disease-related visual attributes knowledge and observed visual evidence.

$$\mathcal{O}^{d_i}_{\text{verify}} = \Phi(\mathcal{I} \mid \mathcal{A}^{K_{d_i}}_{\text{TCR}}) \tag{4}$$

The verification results are further fed back to refine the differential diagnoses space.

$$\mathcal{D} = \Psi\left(\{\mathcal{O}^{d_i}_{\text{verify}}\}^{|\mathcal{D}|}_{i=1}, \mathcal{D}, \mathcal{Q}\right) \tag{5}$$

TABLE I
PERFORMANCE COMPARISON ON SD-198 [23] AND SKINCAP [24] DATASETS

| Category | Model | SD-198 | | | | SkinCap | | |
|---|---|---|---|---|---|---|---|---|
| | | BLEU-1↑ | ROUGE-L↑ | BERTScore↑ | Acc↑ | BLEU-1↑ | BERTScore↑ | ROUGE-L↑ |
| General MLLMs | Qwen2.5-VL-7B (2025) | 0.382 | 0.417 | 0.781 | 0.294 | 0.074 | 0.827 | 0.082 |
| | InternVL3-8B (2025) | 0.513 | 0.524 | 0.784 | 0.473 | 0.065 | 0.818 | 0.075 |
| Medical MLLMs | LingShu-7B (2025) | 0.504 | 0.524 | 0.850 | 0.456 | 0.104 | 0.853 | 0.109 |
| | HuatuoGPT-Vision-7B (2024) | 0.120 | 0.171 | 0.555 | 0.351 | 0.069 | 0.828 | 0.071 |
| | LLaVA-Med (2023) | 0.090 | 0.120 | 0.441 | – | 0.135 | 0.535 | 0.167 |
| Dermatology MLLMs | SkinVL-MM (2025) | 0.003 | 0.024 | 0.389 | 0.174 | 0.059 | 0.532 | 0.148 |
| | SkinVL-PubMM (2025) | 0.028 | 0.141 | 0.581 | 0.260 | 0.196 | 0.576 | **0.209** |
| | SkinVL-Pub (2025) | 0.029 | 0.180 | 0.623 | 0.323 | 0.152 | 0.542 | 0.176 |
| **ours** | **KE-SMMR** | **0.714** | **0.726** | **0.956** | **0.645** | **0.246** | **0.909** | 0.204 |

*Note:* Best results are shown in bold. MLLMs denotes Multi-modal Large Language Models. – indicates data not available.

TABLE II
ABLATION STUDY

| Setting | Differential Diagnose | DMRG | | |
|---|---|---|---|---|
| | Acc↑ | BLEU-1↑ | BERTScore↑ | ROUGE-L↑ |
| Zero-shot | 0.294 | 0.074 | 0.827 | 0.082 |
| Few-shot | 0.412 | 0.098 | 0.844 | 0.111 |
| SMMR | 0.583 | 0.239 | 0.900 | 0.203 |
| KE-SMMR | **0.645** | **0.246** | **0.909** | **0.204** |

## IV. EXPERIMENTS

### A. Experimental Setup

**Datasets.** We evaluate our KE-SMMR framework on two representative dermatological benchmarks, assessing its effectiveness in structured clinical reasoning and free-text medical report generation. (1) SkinCap [24] is a comprehensive dataset for DMRG. (2) SD-198 [23] is reformulated into a multiple-choice Visual Question Answering (VQA) task. This setup rigorously tests the baseline approaches' differential diagnostic capabilities by requiring it to identify the correct disease category among predefined candidates.

**Evaluation Metrics.** For the closed-set VQA task, we report Accuracy (Acc) to measure diagnostic precision. For the report generation task, we employ BLEU-1 and ROUGE-L to assess linguistic fluency and overlap, while BERTScore is utilized to evaluate semantic fidelity and medical consistency between the generated reports and ground truth.

**Baselines.** We construct a representative set of baseline approaches spanning three levels of domain expertise:

- General MLLMs: These models represent the advanced performance in general multi-modal understanding.
- Medical MLLMs: These models are pre-trained on broad medical data, allowing us to observe the performance gap between general medical knowledge and dermatology-specific reasoning.
- Dermatology MLLMs: As direct competitors, these baseline approaches represent dermatological models.

### B. Comparison with Baselines

Table I summarizes the quantitative comparison between our KE-SMMR framework and all baseline approaches. We acknowledge several concurrent or recent works such as [1], [4], [17] that also explore reliable DMRG. However, as their source data or pre-trained checkpoints are not publicly available at the time of this submission, a direct performance comparison is not feasible. Nevertheless, we provide an extensive comparison against 8 representative baselines across general, medical, and dermatological domains to ensure a rigorous evaluation. In differential diagnostic task, our method significantly outperforms baselines. This superiority mainly stems from alignment verification of knowledge for differential diseases. Crucially, the results reveal a diminished instruction-following capability and weak generalization in existing medical-general MLLMs. This underscores the limitations of monolithic models trained on narrow distributions. For DMRG task, medical models generally outperform general MLLMs due to their exposure to specialized corpora. Our framework achieves performance on par with the SkinVL series, which also demonstrates that explicit knowledge enhanced reasoning frameworks can be comparable to implicit parametric modeling.

### C. Ablation Study

**Ablation Settings.** To dissect the contribution of each component in our KE-SMMR framework, we define four incremental configurations based on the Qwen2.5-VL-7B-Instruct backbone:

1) **Zero-shot**. The Qwen2.5-VL-7B-Instruct model directly output, representing a standard end-to-end MLLMs approach.
2) **Few-shot**. Providing Qwen2.5-VL-7B-Instruct few examples of tasks and desired outputs within the query to establish patterns and constraints.
3) **Structured Multi-modal Reasoning (SMMR)**. MLLMs–LLMs collaborative reasoning without external knowledge constraints. Qwen2.5-vl-7B-Instruct is the MLLMs engine, and GPT-4o is the LLMs engine.

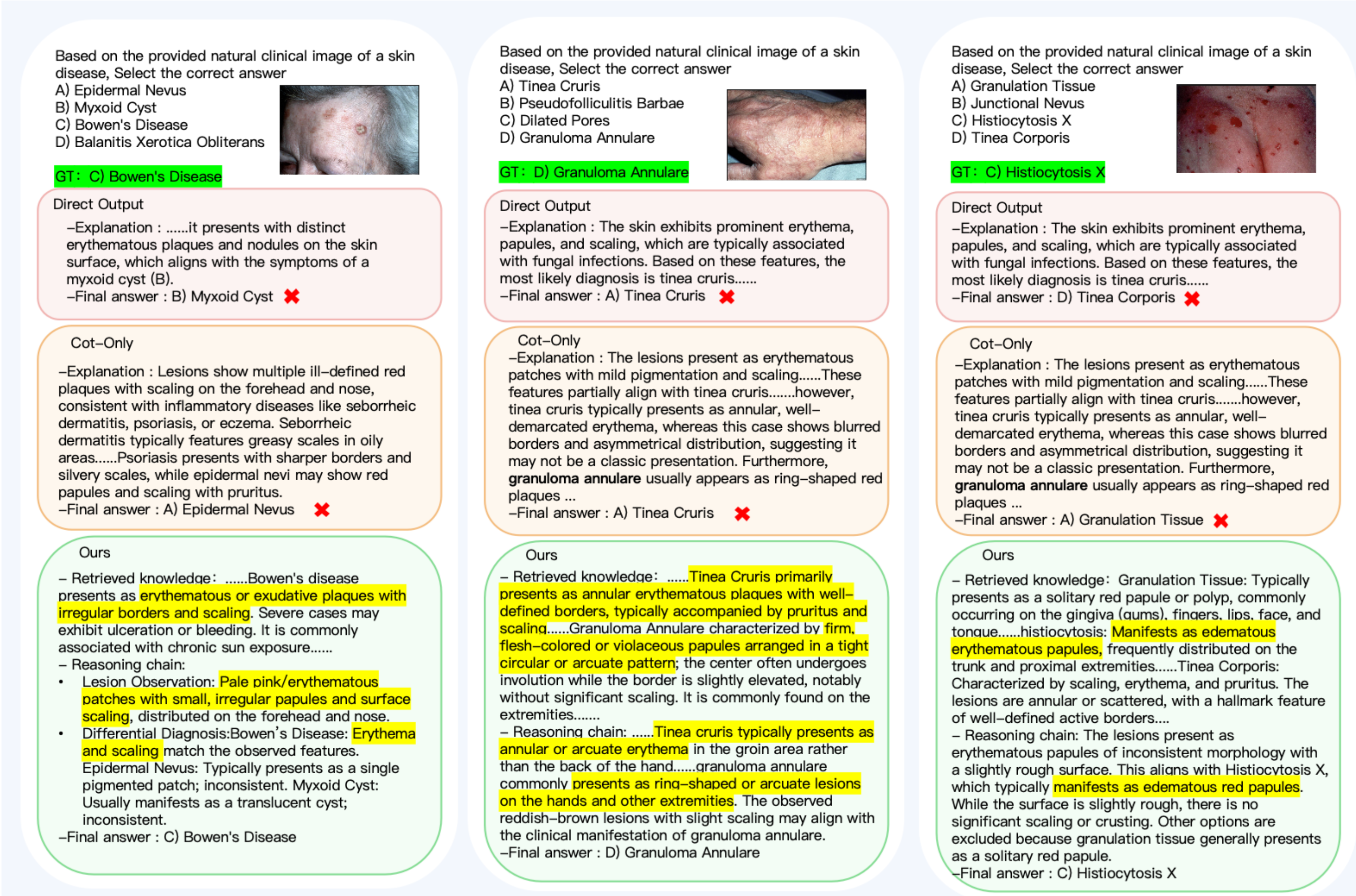


Fig. 2. Illustration of our KE-SMMR framework overcoming medical hallucinations and illogical reasoning via explicit knowledge constraints.

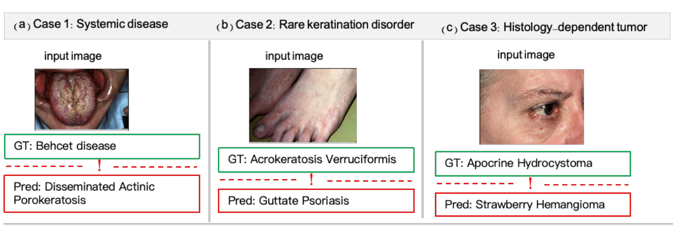


Fig. 3. Illustration of the methodological boundaries of our KE-SMMR framework: identifying diagnostic limitations across systemic, rare, and context-dependent dermatological conditions.

4) **Knowledge-Enhanced Reasoning (KE-SMMR)**. Base on Structured Multi-modal Reasoning, we add TCR, integrating Knowledge-Enhanced alignment to verify and calibrate the reasoning process with dermatological knowledge, which is our full KE-SMMR framework.

As reported in Table II. **Few-shot:** Providing few examples yields a significant performance gain. However, this is primarily attributed to instruction adherence, as prompt strictly constraints the required clinical templates where the solo backbone model often fails. However, when dealing with complex dermatological disease diagnosis reasoning tasks, few-shot prompts are not sufficient to obtain reliable responses. **SMMR:** Furthermore, we design a structured multi-modal reasoning framework, enforcing a "clinical pathway" from symptom identification to diagnosis. Integrating the structured framework further improves accuracy by **0.171** compared with **few-shot**, which indicates that the performance improvement of our framework is not only due to the better instruction adherence, proving the effectiveness of structured multi-modal reasoning framework. **KE-SMMR:** While the KE-SMMR framework achieves best results, its performance gain on DMRG task is notably more marginal compared to the substantial improvements observed in differential diagnosis. We attribute this performance ceiling to the "Sensitivity of Differential Space Construction": unlike the differential diagnosis task which benefits from a statically defined candidate space, report generation relies on a MLLMs' initial diagnosis. This dependency creates a bottleneck where any preliminary visual misalignment is propagated through the knowledge retrieval stage, thereby constraining the overall efficacy of explicit knowledge alignment in open-ended generation scenarios.

## D. Case Study

Figure.2 shows how our framework effectively overcomes the problems of medical hallucinations and imprecise logical reasoning by introducing explicit knowledge constraints. We have observed that the direct output of MLLMs often relies on surface visual similarity and broken logical reasoning chains. Structured reasoning improves interpretability, but there are still issues of knowledge illusion and incomplete verification. These results confirm that structured, knowledge-enhanced reasoning is essential for reliable DMRG. Despite the improvement brought by our KE-SMMR framework, we observe several representative failure cases that reflect intrinsic limitations, which also reflect broader challenges in dermatological visual diagnosis.

In Fig.3, systemic diseases (a) often manifest with heterogeneous and non-specific cutaneous symptoms, making reliable diagnosis fundamentally difficult when relying solely on visual cues. Rare diseases (b) suffer from long-tailed data distributions, where limited visual exemplars and overlapping appearances inevitably lead to diagnostic confusion. Certain dermatological conditions (c) cannot be conclusively determined from a single image, as accurate diagnosis may require additional clinical context such as patient history, laboratory tests, or histopathological evidence.

## V. Conclusion

In conclusion, we propose a KE-SMMR framework for DMRG that improves reliability by explicitly aligning visual evidence with dermatological knowledge. Through extensive quantitative and qualitative experiments, our KE-SMMR framework has demonstrated outstanding performance in DMRG. However, we acknowledge the limitations of our KE-SMMR framework, particularly regarding long-tail pathologies and fine-grained subtypes. Our future research will focus on further enhancing diagnostic accuracy and system robustness across a broader range of clinical dermatological scenarios.

## VI. Acknowledgements

The research was supported by the project of Chongqing Traditional Chinese Medicine Hospital's First Batch of Open Projects in Digital and Intelligent Traditional Chinese Medicine (YN2501).

## References

[1] J. Zhou, X. He, L. Sun, J. Xu, X. Chen, Y. Chu, L. Zhou, X. Liao, B. Zhang, S. Afvari *et al.*, "Pre-trained multimodal large language model enhances dermatological diagnosis using skingpt-4," *Nature Communications*, vol. 15, no. 1, p. 5649, 2024.

[2] S. N. Panda, V. Kukkala, and S. Iyer, "Ai-powered dermatological diagnosis: From interpretable models to clinical implementation a comprehensive framework for accessible and trustworthy skin disease detection," *arXiv preprint arXiv:2512.16235*, 2025.

[3] O. Uwishema, M. Ghezzawi, N. Charbel, S. Alawieh, S. Roy, M. Wojtara, C. M. Hakayuwa, I. K. Ja'afar, G. Nkurunziza, and M. Prasad, "Diagnostic performance of artificial intelligence for dermatological conditions: a systematic review focused on low-and middle-income countries to address resource constraints and improve access to specialist care," *International journal of emergency medicine*, vol. 18, no. 1, p. 172, 2025.

[4] L. Chakrabarti, B. Zhang, H. Tian, A. Vakanski, and M. Xian, "Automated skin cancer report generation via a knowledge-distilled vision-language model," *IEEE Access*, 2025.

[5] S. Yan, Z. Yu, C. Primiero, C. Vico-Alonso, Z. Wang, L. Yang, P. Tschandl, M. Hu, L. Ju, G. Tan *et al.*, "A multimodal vision foundation model for clinical dermatology," *Nature Medicine*, pp. 1–12, 2025.

[6] Y. Gulzar, S. Agarwal, S. Soomro, M. Kandpal, S. Turaev, C. W. Onn, S. Saini, and A. Bounsiar, "Next-generation approach to skin disorder prediction employing hybrid deep transfer learning," *Frontiers in Big Data*, vol. 8, p. 1503883, 2025.

[7] J. Yu, I. H. Cheong, Z. Kozlakidis, and H. Wang, "Advancements and challenges of artificial intelligence in dermatology: a review of applications and perspectives in china," *Frontiers in Digital Health*, vol. 7, p. 1544520, 2025.

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

[9] W. Xu, H. P. Chan, L. Li, M. Aljunied, R. Yuan, J. Wang, C. Xiao, G. Chen, C. Liu, Z. Li *et al.*, "Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning," *arXiv preprint arXiv:2506.07044*, 2025.

[10] W. Zeng, Y. Sun, C. Ma, W. Tan, and B. Yan, "Mm-skin: Enhancing dermatology vision-language model with an image-text dataset derived from textbooks," in *Proceedings of the 33rd ACM International Conference on Multimedia*, 2025, pp. 3769–3778.

[11] S. Yan, X. Li, M. Hu, Y. Jiang, Z. Yu, and Z. Ge, "Make: Multi-aspect knowledge-enhanced vision-language pretraining for zero-shot dermatological assessment," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2025, pp. 369–379.

[12] Q. Abbas, W. Jeong, and S. W. Lee, "Explainable ai in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges," in *Healthcare*, vol. 13, no. 17. MDPI, 2025, p. 2154.

[13] C. Y. Li, X. Liang, Z. Hu, and E. P. Xing, "Knowledge-driven encode, retrieve, paraphrase for medical image report generation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 6666–6673.

[14] S. Yang, X. Wu, S. Ge, S. K. Zhou, and L. Xiao, "Knowledge matters: Chest radiology report generation with general and specific knowledge," *Medical image analysis*, vol. 80, p. 102510, 2022.

[15] H. You, R. Sun, Z. Wang, L. Chen, G. Wang, H. Ayyubi, K.-W. Chang, and S.-F. Chang, "Idealgpt: Iteratively decomposing vision and language reasoning via large language models," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 11 289–11 303.

[16] L. Wei, W. Wang, X. Shen, Y. Xie, Z. Fan, X. Zhang, Z. Wei, and W. Chen, "Mc-cot: A modular collaborative cot framework for zero-shot medical-vqa with llm and mllm integration," *arXiv preprint arXiv:2410.04521*, 2024.

[17] Y. Shen, J. Qian, Z. Chen, Y. He, and J. Zhou, "Skingpt-r1: Adapter-only dual distillation for efficient dermatology reasoning," *arXiv preprint arXiv:2511.15242*, 2025.

[18] J. Ru, S. Yan, Y. Yin, Y. Zou, and Z. Ge, "Dermogpt: Open weights and open data for morphology-grounded dermatological reasoning mllms," *arXiv preprint arXiv:2601.01868*, 2026.

[19] S. Wu, B. Yang, Z. Ye, H. Wang, H. Zheng, and T. Zhang, "Improving medical report generation with adapter tuning and knowledge enhancement in vision-language foundation models," *arXiv preprint arXiv:2312.03970*, 2023.

[20] Y. Li, Z. Wang, Y. Liu, L. Wang, L. Liu, and L. Zhou, "Kargen: Knowledge-enhanced automated radiology report generation using large language models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 382–392.

[21] M. Jia, J. Duan, Y. Song, and J. Wang, "medikal: Integrating knowledge graphs as assistants of llms for enhanced clinical diagnosis on emrs," in *Proceedings of the 31st International Conference on Computational Linguistics*, 2025, pp. 9278–9298.

[22] X. Zhao, S. Liu, S.-Y. Yang, and C. Miao, "Medrag: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot," in *Proceedings of the ACM on Web Conference 2025*, 2025, pp. 4442–4457.

[23] X. Sun, J. Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *European conference on computer vision*. Springer, 2016, pp. 206–222.

[24] J. Zhou, L. Sun, Y. Xu, W. Liu, S. Afvari, Z. Han, J. Song, Y. Ji, X. He, and X. Gao, "Skincap: A multi-modal dermatology dataset annotated with rich medical captions," *arXiv preprint arXiv:2405.18004*, 2024.