# PRISM-Med: A Perception, Routing, Inspection, and Self-Correction Multi-agent Collaboration System for Medical VQA

Jie Zhou
School of Software Engineering
East China Normal University
Shanghai, China
Email: 51285902116@stu.ecnu.edu.cn

***Abstract*—Medical Visual Question Answering (Med VQA) presents a unique challenge demanding rigorous precision and zero hallucinations. Unlike general VQA tasks, interpreting medical imagery requires rigorous multi-scale feature understanding and multi-step diagnostic reasoning. Existing approaches, predominantly relying on single-agent Multimodal Large Language Models, often suffer from "tunnel vision", focusing on salient features while missing subtle yet critical anomalies and lack robust self-verification mechanisms. Furthermore, the static inference strategies employed by current models lack the flexibility to adapt to varying question complexities, leading to inefficient resource scheduling. Inspired by how a prism decomposes light to reveal spectral details, we propose PRISM-Med (Perception, Routing, Inspection, Self-correction Multi-agent Collaboration System for Medical VQA), a modular framework. Our system features three key innovations: (1) a Two-Stage Diagnostic Architecture mimicking clinical "Triage-to-Consultation" with Organizer-based dynamic routing; (2) a Clinically Grounded Multi-Agent Framework with four specialized experts to mitigate tunnel vision; (3) a Dynamic Inference Optimization Strategy based on Proximal Policy Optimization (PPO) for adaptive generation configuration. PRISM-Med achieves state-of-the-art performance among open-source models on PMC-VQA, MedXQA, MMMU and OmniMedVQA benchmarks, outperforming baselines by approximately 3.68% on average. The code is available at https://gitee.com/zhou-jie021125/med.git.**

***Index Terms*—Medical VQA, Multi-Agent Collaboration, Adaptive Routing, Self-Reflection, Reinforcement Learning, PPO**

## I. Introduction

Situated at the convergence of high-stakes clinical reasoning and computer vision, Medical Visual Question Answering demands an artificial intelligence capable of navigating complex diagnostic deduction rather than merely recognizing pixel-level features. Although recent proprietary giants such as GPT-4o [1] and open-source contenders such as Qwen2.5-VL [3] have shown notable generalist capabilities, their deployment in the demanding environment of medical diagnostics exposes fundamental architectural limitations.

The primary deficiency stems from the inherent cognitive bias embedded within Single-Agent Models, which inadvertently mirror the radiological error known as "Satisfaction of Search" (SOS) [5]. In this phenomenon, a radiologist might prematurely terminate their scan after identifying a prominent anomaly, thereby overlooking subtle but critical peripheral evidence. Similarly, single-agent Multimodal Large Language Models (MLLMs) tend to fixate on salient visual features while discarding critical peripheral evidence. This tunnel vision severely impairs their diagnostic performance. Recent investigations into visual attention mechanisms confirm that holistic image encoding frequently fails to capture the fine-grained Regions of Interest (ROI) that are essential for a comprehensive diagnosis.

Beyond individual bias, the field currently faces a challenge due to the absence of a clinically grounded diagnostic structure, for although benchmarks like MedAgentBoard [9] have initiated the exploration of multi-agent collaboration, existing frameworks frequently depend on superficial role-playing prompts that fail to elicit the genuine, distinct domain expertise required for complex cases. Generic roles are insufficient. Instead of arbitrary personas, true diagnostic rigor demands an ensemble strictly corresponding to the dual-process cognitive models validated in clinical literature [14], [28]. This alignment compels the system to decompose image interpretation into scientifically distinct and specialized visual streams.

In the practical application of multi-agent frameworks, the profound inefficiency of Static Inference Configurations remains a critical bottleneck. Current systems indiscriminately apply identical hyperparameters like temperature and token limits to every query, ignoring the significant distinction between simple binary confirmations and complex differential diagnoses [16]. This rigidity inevitably leads to resource wastage. Even advanced attempts like MMMedAgent-RL [11], which utilize Reinforcement Learning (RL) to optimize agent routing, fail to address the continuous dynamics of the generation process itself.

To bridge this gap, we propose **PRISM-Med**, a Perception, Routing, Inspection, and Self-correction Multi-agent Collaboration System for Medical VQA. Our contributions are as follows:

- **Clinical Workflow-Inspired Architecture:** Mimicking

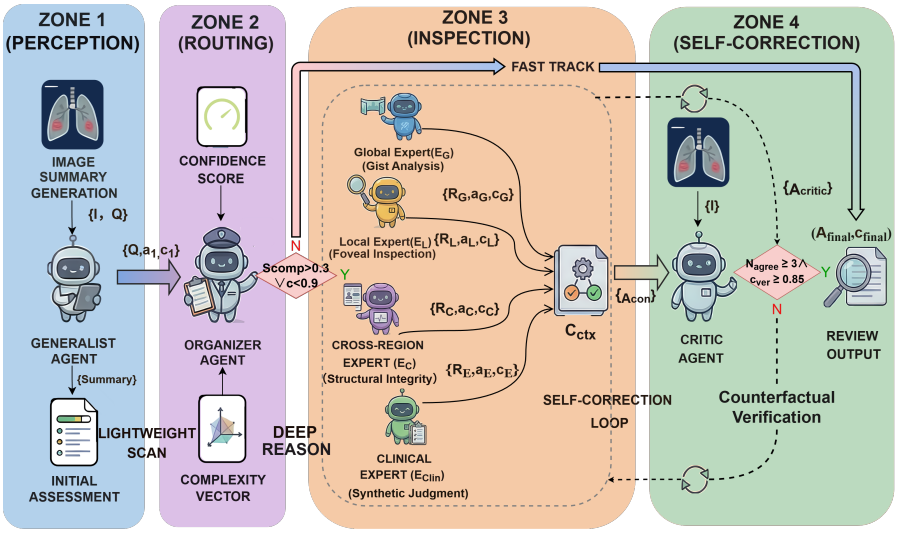


**Fig. 1.** Overview of the proposed PRISM-Med Framework. The workflow follows a "Perception-Routing-Inspection-Self-correction" loop. An Organizer Agent (Routing) dynamically directs queries between a fast Single Agent (Perception) and a deep-reasoning Four-Expert Ensemble (Inspection). A final Self-Correction module ensures answer reliability before generating output.

the clinical "Triage-to-Consultation" workflow, we introduce a Perception-Routing-Inspection-Self-correction pipeline that uses an Organizer gatekeeper to dispatch queries between lightweight scanning and deep reasoning.

- **Radiologically Grounded Ensemble:** To mitigate single agents' "tunnel vision," we implement our Four-Expert Ensemble (from cognitive search patterns) that decomposes image interpretation into global, local, comparative, and clinical visual streams.
- **Dynamic Inference Optimization:** Treating parameter tuning as a continuous (not static) control problem, we integrate Proximal Policy Optimization (PPO) [17] to adaptively adjust generation configurations (e.g., temperature, token limits, confidence) based on query complexity.

## II. Related Work

### A. Evolution of Medical MLLMs and Visual Grounding

Medical VQA has evolved from simple classification to complex reasoning. Early models adapted generalist encoders, but recent works highlight the necessity of fine-grained visual understanding. R-LLaVA [7] demonstrates that integrating visual ROI significantly reduces hallucinations compared to holistic processing, validating the need for focused inspection mechanisms. Similarly, VTQA [8] emphasizes the challenge of entity alignment in multi-hop reasoning. Despite these advances, Single-Agent Models still struggle to balance global context with local details, often requiring external verification mechanisms to ensure reliability [6].

### B. Multi-Agent Collaboration and Calibration

To overcome individual model limitations, research has shifted towards multi-agent systems. MedAgentBoard [9] provides a comprehensive benchmark, revealing that while collaboration generally improves performance, it is sensitive to prompt engineering and the capabilities of agents. To address confidence issues, AlignVQA [10] introduces a debate-driven framework that improves calibration, ensuring model confidence aligns with accuracy. Domain-specific multi-agent systems have also emerged in fields like agriculture [12] and cartoon understanding [13], employing specialized roles (e.g., Reflector, Critic). Inspired by these works, our proposed PRISM-Med advances this paradigm by defining expert roles not arbitrarily, but based on the Radiological Cognitive Model [14], ensuring diverse and clinically relevant perspectives.

### C. Reinforcement Learning in Adaptive Inference

Optimizing agent behavior via RL is a growing frontier. MMMedAgent-RL [11] utilizes RL to train a "triage" agent that dynamically assigns tasks to different specialists, optimizing the routing process. However, this approach largely treats the underlying inference process as static. In contrast, our proposed PRISM-Med leverages PPO to control the continuous inference parameters (e.g., temperature, token limits) of the experts themselves. This allows our multi-agent collaboration system to dynamically shift between deterministic fact-retrieval and creative diagnostic reasoning, offering a finer granularity of control compared to discrete role selection.

## III. Methodology

PRISM-Med is a dynamic multi-agent collaboration system (overall framework shown in Fig. 1) with four interdependent modules: *Perception*, *Routing*, *Inspection*, and *Self-correction*.

### A. Module 1: Perception (Initial Assessment)

To initiate the diagnostic process while maintaining the real-time responsiveness required by clinical environments, the system first synthesizes a text-based Image Summary via a vision encoder, establishing a rapid, albeit superficial,

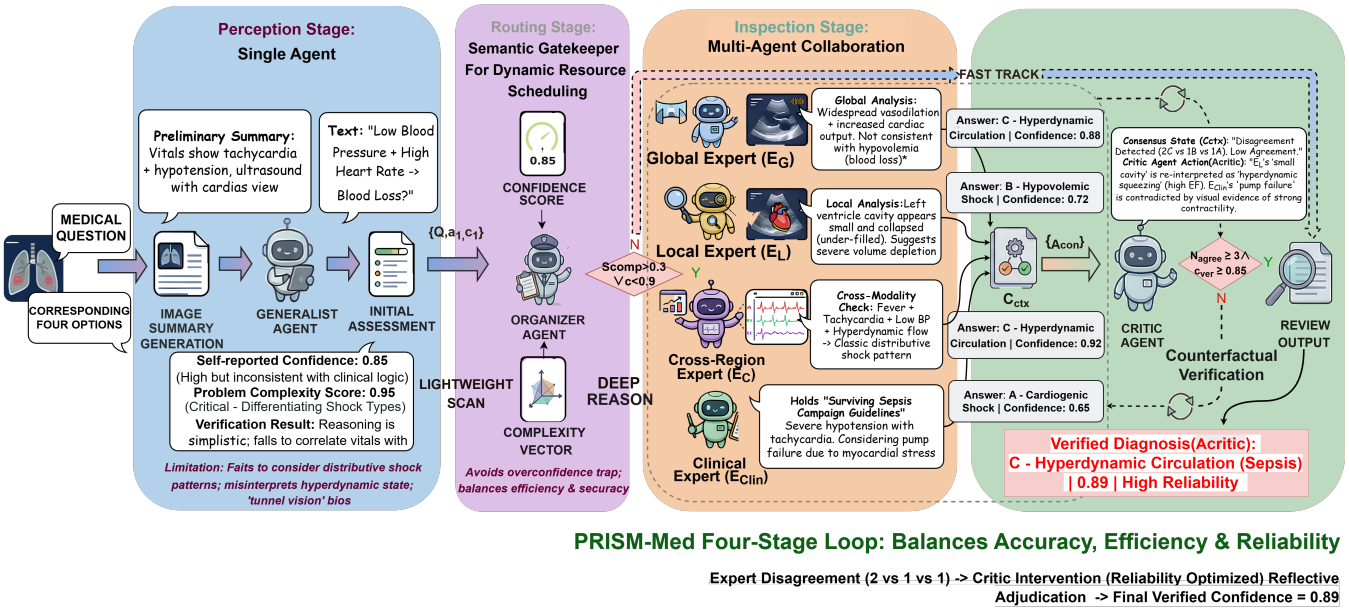


**Fig. 2.** Comprehensive Qualitative Case Study of the proposed PRISM-Med. It demonstrates the full decision loop resolving a complex sepsis case. In Zone 1 (Perception), the Single Agent incorrectly diagnoses "Blood Loss" by fixating on hypotension and overlooking conflicting vitals. The Organizer Agent in Zone 2 (Routing) detects this inconsistency and high complexity, routing the task to Zone 3 (Inspection) to decompose the problem with four specialized experts: the Global Expert ($E_G$) identifies widespread vasodilation (inconsistent with hypovolemia); the Local Expert ($E_L$) finds a hyperdynamic left ventricle with high ejection fraction, ruling out cardiogenic shock; the Cross-Region Expert ($E_C$) confirms a classic distributive shock pattern via cross-modality checks (fever + tachycardia + low BP + hyperdynamic flow); and the Clinical Expert ($E_{Clin}$) correlates these findings with sepsis guidelines. Finally, in Zone 4 (Self-Correction), the system integrates these insights to override the initial hallucination, delivering a verified high-confidence (0.89) diagnosis of "Hyperdynamic Circulation."

contextual baseline. Speed is prioritized here. Simultaneously, a "Generalist Agent" ($A_{gen}$) attempts to process the raw input tuple $(I, Q)$ to generate a preliminary hypothesis $a_1$ alongside a raw self-confidence score $c_1 \in [0, 1]$.

$$(a_1, c_1) \leftarrow A_{gen}(I, Q, \text{Summary}) \tag{1}$$

This stage efficiently filters out routine queries that do not require expensive computation; however, this efficiency often breeds complacency, creating a risk of "tunnel vision" that necessitates a stringent check.

### B. Module 2: Routing (Organizer-based Gating)

To intercept the potential hallucinations characteristic of the fast Generalist Agent before they propagate downstream, we deploy an LLM-based Organizer Agent that functions as a strict semantic gatekeeper. We define the rule-based Complexity Vector $S_{comp}$ by integrating medical VQA-specific complexity metrics (i.e., clinical reasoning steps, anatomical granularity, multimodal fusion demands) and universal LLM reasoning complexity rules (i.e., logical operators like "compare" or "difference", query length, medical entity count). The Organizer normalizes each dimension of $S_{comp}$ to $[0, 1]$ and calculates a comprehensive complexity score via weighted summation, then scrutinizes the tuple $(Q, a_1, c_1)$ against this score to judge if the query exceeds the Generalist's capability. Queries with a score above the threshold trigger the expert ensemble, while routine low-complexity queries are output directly. This mechanism effectively prevents the "overconfidence trap" common in single agents by matching task complexity with agent capabilities. Algorithm 1 details this mechanism.

### C. Module 3: Inspection (Role-Playing Ensemble)

Once a query is flagged as ambiguous or complex, it is immediately escalated to our Four-Expert Ensemble, where roles are not assigned arbitrarily but are rigorously mapped to the standard cognitive phases of a radiologist's search pattern.

- **Global Expert ($E_G$, "The Scout"):** Prioritizing the initial "Gist" phase, this expert rapidly parses the macro-scale anatomical layout to establish a foundational spatial context before any detailed inspection occurs.
- **Local Expert ($E_L$, "The Examiner"):** Complementing the global view, this expert partitions the image into fine-grained anatomical regions to execute a "Foveal Inspection," ensuring that subtle micro-anomalies like nodules are seldom overlooked.
- **Cross-Region Expert ($E_C$, "The Comparator"):** Focusing on structural integrity, this expert scrutinizes bilateral symmetry and temporal consistency across multiple

**Algorithm 1** Adaptive Routing Logic for the proposed PRISM-Med

**Input:** Question $Q$, Initial Confidence $c_1$, Answer $a_1$
**Output:** Decision $D \in \{\text{Stop}, \text{Proceed}\}$
Calculate Complexity Score $S_{comp}$ based on keywords
**if** $S_{comp} > 0.3$ (Complex Query) **then**
  **return** **Proceed** (to Inspection)
**else if** $c_1 < 0.9$ (Low Confidence) **then**
  **return** **Proceed** (to Inspection)
**else**
  Organizer evaluates if rationale for $a_1$ is sound
  **if** Organizer Approves **then**
    **return** **Stop** (Output $a_1$)
  **else**
    **return** **Proceed** (to Inspection)
  **end if**
**end if**

views to identify pathological deviations that disrupt normal physiological patterns.

- **Clinical Expert ($E_{Clin}$, "The Diagnostician"):** Bridging the pixel-semantic gap, this expert synthesizes the aggregated visual findings into coherent medical terminology, aligning abstract feature vectors with established clinical diagnostic guidelines.

These experts provide diverse perspectives to ensure coverage; however, a collection of opinions does not inherently constitute the truth, for a group can hallucinate together.

### *D. Module 4: Self-correction (Verification)*

To mitigate the risk of "groupthink" or collective error, the final module first aggregates the outputs of the Four-Expert Ensemble to generate a consensus answer $A_{\text{con}}$, which represents the majority opinion of the specialized experts weighted by their individual confidence scores. The system imposes a stringent Counterfactual Verification step driven by a skeptical Critic Agent ($A_{critic}$), which enforces rigorous validation to mitigate collective hallucinations. The Critic re-examines the raw medical image $I$ against the consensus answer $A_{\text{con}}$ to determine if pixel-level visual evidence actually supports the clinical features claimed in the answer.

$$c_{ver} \leftarrow A_{critic}(I, A_{\text{con}}) \tag{2}$$

Here, $c_{ver} \in [0, 1]$ denotes the verified confidence score output by the Critic Agent, quantifying the alignment between the consensus answer and the raw visual evidence. The final decision logic is systematic, accepting the consensus answer directly only if it satisfies a high-confidence agreement threshold that leaves no room for ambiguity:

$$\text{Condition}: \quad (N_{\text{agree}} \geq 3) \wedge (c_{\text{ver}} \geq 0.85) \tag{3}$$

where $N_{agree}$ is the number of experts that align with the consensus answer.Failure to meet this strict condition indicates either expert disagreement (fewer than 3 alignments) or weak visual evidence. In such cases, the system triggers a Reflective Synthesis step to re-evaluate the entire diagnostic context. This step aggregates the raw reasoning rationales $R_k$ from each expert $k \in \mathcal{E}$ (where $\mathcal{E}$ denotes the full set of four experts) to construct a comprehensive context vector $C_{\text{ctx}}$:

$$C_{\text{ctx}} \leftarrow \sum_{k \in \mathcal{E}} w_k \cdot R_k \tag{4}$$

where $w_k$ represents the confidence-weighted coefficients of each expert's rationale. Finally, instead of delegating the synthesis to a separate module, the Critic Agent assumes the role of the ultimate adjudicator. It produces the final calibrated answer $A_{\text{final}}$ by integrating the question $Q$, the comprehensive context vector $C_{\text{ctx}}$, and the weighted confidence scores of all experts. This design ensures that the final decision remains grounded in the rigorous skepticism characteristic of the Critic:

$$A_{\text{final}} \leftarrow A_{\text{critic}}(Q, C_{\text{ctx}}, \text{weighted\_conf}) \tag{5}$$

Algorithm 2 details this process for the proposed PRISM-Med.

Yet, to make this complex loop efficient, we cannot rely on static rules (e.g., fixed confidence thresholds or uniform expert weights); the multi-agent collaboration system must learn to adapt its own inference parameters dynamically.

**Algorithm 2** Self-Correction with Reflective Synthesis for the proposed PRISM-Med

**Input:** Image $I$, Question $Q$, Consensus Answer $A_{con}$, Expert Set $\mathcal{E} = \{(a_k, c_k, R_k)\}$
**Output:** Final Answer $A_{final}$
**Step 1: Visual Verification**
Critic Agent examines visual evidence for $A_{con}$
Calculate verified confidence $c_{ver} \leftarrow A_{critic}(I, A_{con})$
Calculate Expert Agreement Count $N_{agree}$
**Step 2: Decision Logic**
**if** $N_{agree} \geq 3$ **and** $c_{ver} \geq 0.85$ **then**
  **return** $A_{con}$ {High confidence, accept directly}
**else**
  **Execute Reflective Synthesis:**
  Construct context $C_{\text{ctx}} \leftarrow \sum_{k \in \mathcal{E}} w_k \cdot R_k$
  $A_{final} \leftarrow A_{critic}(Q, C_{ctx}, \text{weighted\_conf})$
  **return** $A_{final}$
**end if**

### *E. Optimization: RL Parameter Tuning*

We employ PPO to treat the inference process of the proposed PRISM-Med not as a fixed procedure, but as a learnable policy that dynamically adjusts generation parameters $\phi$ based on real-time feedback. As illustrated in Figure 3, the Policy Network acts as a dynamic controller, observing task-specific state features to output optimal configurations for creativity versus determinism.

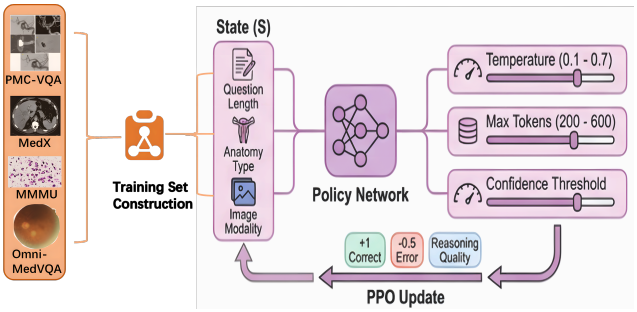


**Fig. 3.** Overview of RL-based Inference Optimization for the proposed PRISM-Med. The Policy Network encodes multi-modal state features to dynamically predict continuous inference hyperparameters.

State Space Encoding. The state $s_t$ is constructed as a dense vector concatenation:

$$s_t \leftarrow [E(Q_t) \oplus V_{comp} \oplus E_{modality}] \tag{6}$$

Here, $E(Q)$ is the BERT-based embeddings of the question, and $V_{comp}$ is the explicit complexity vector. Reward Function. The reward is defined as:

$$R = \alpha R_{acc} + \beta R_{agree} + \gamma R_{calib} \tag{7}$$

encouraging correctness ($R_{acc}$), expert consensus ($R_{agree}$), and penalizing overconfidence on wrong answers ($R_{calib}$). The PPO training procedure for the proposed PRISM-Med is detailed in Algorithm 3.

**Algorithm 3** RL-based Inference Parameter Optimization for the proposed PRISM-Med

**Initialize:** Policy network $\pi_\theta$, Value network $V_\phi$
**Input:** Training dataset $\mathcal{D} = \{(I_i, Q_i, A_i^{gt})\}_{i=1}^N$
**for** iteration $= 1, 2, \ldots, M$ **do**
  Initialize empty buffer $\mathcal{B}$
  **for** step $t = 1$ to $T$ **do**
    Construct State $s_t \leftarrow [E(Q_t) \oplus V_{comp} \oplus E_{modality}]$
    Sample Action $a_t \sim \pi_\theta(\cdot|s_t)$, where $a_t = [\text{Temp}, \text{Len}, \text{Con}]$
    Execute Expert Ensemble with parameters $a_t$
    Get predicted answer $\hat{A}_t$ and consensus score $R_{agree}$
    Calculate Penalty $R_{calib} = \mathbb{I}(\hat{A}_t \neq A_t^{\text{gt}}) \cdot \left(\text{Con} + \lambda_1 \frac{\text{Len}}{L_{max}} + \lambda_2 \frac{1}{\text{Temp}}\right)$
    Compute Reward $r_t = \alpha \mathbb{I}(\hat{A}_t = A_t^{\text{gt}}) + \beta R_{agree} - \gamma R_{calib}$
    Store transition $(s_t, a_t, r_t)$ in $\mathcal{B}$
  **end for**
  Compute Advantages using Generalized Advantage Estimation (GAE)
  Update $\theta$ via PPO objective $L^{CLIP}(\theta)$
  Update $\phi$ via Value loss $L^{VF}(\phi)$
**end for**
**Return:** Optimized Policy $\pi_{\theta^*}$

## IV. Experiments

### A. Setup

**Datasets.** We evaluate the proposed PRISM-Med on four diverse medical benchmarks to ensure a comprehensive assessment of its capabilities. **PMC-VQA** (PubMed Central Visual Question Answering) [19] is a large-scale radiological dataset derived from PubMed Central, focusing on visual instruction tuning. **MedXQA** (Medical Diagnostic Reasoning Visual Question Answering) [20] represents a complex diagnostic reasoning benchmark designed to test multi-step logical deduction in clinical scenarios. **MMMU** (Massive Multi-discipline Multimodal Understanding) (Medical) [21] is the medical subset of the Massive Multi-discipline Multimodal Understanding benchmark, evaluating expert-level reasoning across clinical medicine and basic sciences. Finally, **OmniMedVQA** (Omnidirectional Medical Visual Question Answering) [22] serves as a comprehensive multi-modal benchmark covering 12 distinct imaging modalities, providing a robust test for cross-domain generalization.

**Baselines.** We compare PRISM-Med against two categories of state-of-the-art models. The first category includes proprietary large-scale models such as GPT-4o [1] and Claude 3.5 Sonnet [2], which represent the state-of-the-art of general-purpose multimodal reasoning. The second category comprises leading open-source Medical MLLMs, specifically Qwen2.5-VL [3] and InternVL2.5 [4], which are specialized for visual-linguistic tasks and serve as strong direct competitors.

**Implementation Details.** Our framework utilizes Qwen2.5-VL-7B for the lightweight Perception stage to ensure low latency. The Inspection stage employs the more powerful Qwen2.5-VL-72B-Instruct to handle complex reasoning tasks. The RL optimization is performed using a three-layer Multi-Layer Perceptron (MLP) policy network, trained with a learning rate of $3e^{-4}$ and a clip range of 0.2.

### B. Main Results

Table I presents the comprehensive comparative evaluation against both proprietary giants and open-source contenders. Across the majority of benchmarks, the proposed PRISM-Med establishes a new state-of-the-art for open-source models. Specifically, on the instruction-heavy PMC-VQA and the reasoning-intensive MedXQA, our multi-agent collaboration system secures substantial margins of 5.0% and 2.8% respectively, validating the efficacy of the multi-expert ensemble in handling complex clinical semantics. While InternVL2.5 retains a slight edge on OmniMedVQA, the proposed PRISM-Med narrows the gap significantly without such targeted tuning. Overall, our method outperforms previous baselines by approximately **3.68% on average**.

### C. Ablation Studies

*1) Routing and Correction Efficiency Analysis:* A key efficiency feature of PRISM-Med is its dual-gate mechanism, consisting of front-end Early Stopping and tail-end Self-Correction, with its efficiency validated in Table II. Early

**Table I.** Performance Comparison on Four Medical VQA Benchmarks. The best/second-best open-source results are in **bold**/underlined.

| Models | PMC-VQA | MedXQA | MMMU | OmniMedVQA | Avg. |
|---|---|---|---|---|---|
| *Proprietary Models* | | | | | |
| GPT-4.1 | 55.2 | 45.2 | 75.2 | 75.5 | 62.78 |
| Claude Sonnet 4 | 54.4 | 43.3 | 74.6 | 65.5 | 59.45 |
| Gemini-2.5-Flash | 55.4 | 52.8 | 76.9 | 71.0 | 64.03 |
| *Open-source Models* | | | | | |
| MedVLM-R1-2B | 47.6 | 20.4 | 35.2 | 77.7 | 45.23 |
| LLaVA-Med-7B | 30.5 | 24.4 | 20.3 | 44.3 | 29.88 |
| HuatuoGPT-V-7B | 53.3 | 21.6 | 47.3 | 74.2 | 49.10 |
| Qwen2.5-VL-7B | 51.9 | 22.3 | 63.1 | 63.6 | 50.23 |
| Qwen2.5-VL-32B | 57.9 | 25.2 | 59.6 | 68.2 | 52.73 |
| InternVL2.5-38B | 57.2 | 24.4 | 61.6 | **79.9** | 55.78 |
| InternVL3-38B | 56.6 | 25.2 | 65.2 | 79.8 | 56.70 |
| Qwen2.5-VL-72B | 59.8 | 28.5 | 63.1 | 72.0 | 55.85 |
| **PRISM-Med (Ours)** | **64.8** | **31.3** | **66.7** | 78.7 | **60.38** |
| *vs. Open Source* | *+5.0%* | *+2.8%* | *+1.5%* | *-1.2%* | *+3.68%* |

Stopping filters out 44.9% of queries to cut computational load significantly, and combined with RL-optimized token limits, the design enables dynamic resource adaptation—achieving an average of just 145 tokens for the simpler PMC-VQA task and 302 for the complex reasoning task MedXQA. Self-Correction is activated for only 48.5% of queries, reserving costly verification for ambiguous cases alone. This dynamic routing strategy outperforms all ablation baselines, reducing average token consumption by **32.9%** and average latency by **31.7%** compared to the static baseline.

**Table II.** Routing and Correction Efficiency Analysis.

| Metric Group | PMC-VQA | MedXQA | MMMU | OmniMedVQA | Avg. |
|---|---|---|---|---|---|
| Early Stop Rate (%) | 62.0 | 34.5 | 38.2 | 45.0 | 44.9 |
| Self-Correction Activation Rate (%) | 38.5 | 62.1 | 51.4 | 42.0 | 48.5 |
| ***Efficiency: Token Consumption per Query*** | | | | | |
| Static Baseline (full inspection) | 260 | 415 | 270 | 235 | 295 |
| Only Early Stop | 182 | 356 | 215 | 189 | 236 |
| Only Self-Correction | 231 | 378 | 242 | 212 | 263 |
| **Ours (Dual-Gate)** | **145** | **302** | **188** | **158** | **198** |
| % Reduction vs Static§ | 44.2 | 27.2 | 30.4 | 32.8 | **32.9** |
| ***Efficiency: Inference Latency per Query (ms)*** | | | | | |
| Static Baseline (full inspection) | 325 | 520 | 340 | 290 | 369 |
| Only Early Stop | 248 | 455 | 272 | 228 | 301 |
| Only Self-Correction | 291 | 478 | 305 | 262 | 334 |
| **Ours (Dual-Gate)** | **185** | **385** | **235** | **190** | **249** |
| % Reduction vs Static§ | 43.1 | 26.0 | 30.9 | 34.5 | **31.7** |

*2) Detailed Component-wise Ablation on PMC-VQA:* To thoroughly analyze the contribution of expert diversity and the collaboration mechanism, we evaluate all possible permutations of expert combinations on the PMC-VQA dataset for the proposed PRISM-Med. Table III details the results. The addition of the Self-Correction module in the final step provides a notable boost (63.8% → 64.8%), proving its value in refining consensus.

**Table III.** Detailed ablation of expert permutations on PMC-VQA. We evaluate all 6 dual-expert and 4 tri-expert combinations.

| Configuration on PMC-VQA | Experts Involved | Accuracy (Acc) (%) |
|---|---|---|
| *Baseline* | Single Agent Only | 59.8 |
| *2-Expert Combinations* | $E_G + E_L$ | 60.1 |
| | $E_G + E_C$ | 60.5 |
| | $E_G + E_{Clin}$ | **61.8** |
| | $E_L + E_C$ | 59.5 |
| | $E_L + E_{Clin}$ | 61.2 |
| | $E_C + E_{Clin}$ | 61.5 |
| *3-Expert Combinations* | $E_G + E_L + E_C$ (w/o Clinical) | 62.0 |
| | $E_G + E_L + E_{Clin}$ (w/o Cross) | 63.1 |
| | $E_G + E_C + E_{Clin}$ (w/o Local) | **63.3** |
| | $E_L + E_C + E_{Clin}$ (w/o Global) | 62.8 |
| *Full Pipeline* | 4 Experts (w/o Verification) | 63.8 |
| | **Full PRISM-Med (All Steps)** | **64.8** |

## D. RL Policy Analysis

*1) Learned Hyperparameters Analysis:* As detailed in Table IV, the Policy Network of the proposed PRISM-Med adapts its strategy to the specific entropy of each dataset. For instance, on the reasoning-heavy MedXQA, the agent learns to elevate the Temperature ($T \approx 0.403$), extend the token limit ($L \approx 419$) and raise the confidence threshold, effectively promoting "creative" chain-of-thought exploration ($\tau \approx 0.827$). In contrast, for OmniMedVQA, which demands strict pattern matching, the policy converges towards high determinism ($T \approx 0.146$).

**Table IV.** Statistics of RL-generated inference parameters (Mean ± Std).

| Dataset | Temperature ($T$) | Max Tokens ($L$) | Confidence ($\tau$) |
|---|---|---|---|
| PMC-VQA | $0.189 \pm 0.016$ | $248 \pm 10$ | $0.753 \pm 0.009$ |
| MedXQA | $0.403 \pm 0.117$ | $419 \pm 72$ | $0.827 \pm 0.055$ |
| MMMU | $0.200 \pm 0.080$ | $255 \pm 48$ | $0.757 \pm 0.048$ |
| OmniMedVQA | $0.146 \pm 0.020$ | $222 \pm 15$ | $0.725 \pm 0.012$ |

*2) Impact of RL Optimization:* To quantify the gains from this adaptive strategy, Figure 4 visualizes the stepwise performance improvements of the proposed PRISM-Med. The architectural shift from a single agent (*Baseline*) to a multi-perspective ensemble (*w/o PPO*) yields a substantial accuracy gain across all datasets. The subsequent integration of PPO optimization (*Full*) further maximizes these gains. Notably, on OmniMedVQA, the system alone boosts accuracy by 3.5%, while PPO adds another 3.2%, demonstrating that clinical logic and adaptive optimization are synergistic.

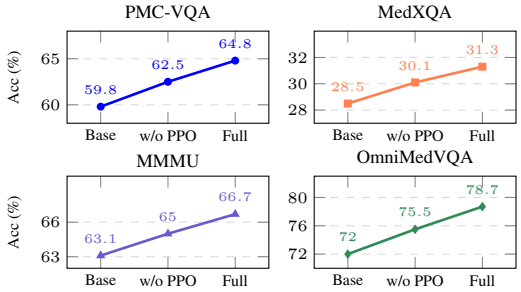


**Fig. 4.** Impact of RL Optimization. Stepwise accuracy (Acc) improvements on four datasets through multi-agent collaboration and PPO optimization.

## V. Conclusion

Just as a physical prism decomposes white light into its constituent spectral colors to reveal hidden details, we presented **PRISM-Med**—a "Perception-Routing-Inspection-Self-correction" Multi-agent Collaboration System for Medical VQA which disentangles complex medical queries into structured, interpretable reasoning streams. By emulating the rigorous workflow of medical boards and grounding expert roles in radiological search patterns, our system achieves superior accuracy and efficiency, with the integration of role-playing experts and RL-driven optimization significantly mitigating the limitations of single-agent models. In doing so, PRISM-Med provides the medical AI community with a new lens, one that shifts focus from black-box prediction to transparent, clinically aligned problem-solving.

## REFERENCES


[1] "GPT-4o System Card," arXiv preprint arXiv:2410.21276, 2024.
[2] Anthropic. "Claude 3.5 Sonnet: Technical Report," 2024.
[3] S. Bai, et al., "Qwen2.5-VL Technical Report," arXiv preprint arXiv:2502.13923, 2025.
[4] Z. Chen, et al., "InternVL2.5: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024.
[5] K. S. Berbaum, et al., "Satisfaction of search in diagnostic radiology," Investigative Radiology, vol. 25, no. 2, pp. 133–140, 1990.
[6] F. Liu, et al., "Visual Hallucination in Multi-modal Large Language Models," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024.
[7] X. Chen, et al., "R-LLaVA: Improving Med-VQA Understanding Through Visual Region of Interest," in Proc. ICLR Workshop on Foundation Models for Computer Vision, 2025.
[8] K. Chen and X. Wu, "VTQA: Visual Text Question Answering via Entity Alignment and Cross-Media Reasoning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024.
[9] Y. Zhu, et al., "MedAgentBoard: Benchmarking Multi-Agent Collaboration with Conventional Methods for Diverse Medical Tasks," arXiv preprint arXiv:2505.12371, 2025.
[10] A. Pandey, et al., "AlignVQA: Debate-Driven Multi-Agent Calibration for Vision Language Models," TCS Res. Tech. Rep., 2025.
[11] P. Xia, et al., "MMMedAgent-RL: Optimizing Multi-Agent Collaboration for Multimodal Medical Reasoning," arXiv preprint arXiv:2506.00555, 2025.
[12] Y. Ke, et al., "Dynamic Orchestration of Multi-Agent System for Real-World Multi-Image Agricultural VQA," arXiv preprint arXiv:2509.24350, 2025.
[13] T. Wu and T. Markchom, "Understanding Multi-Agent Reasoning with Large Language Models for Cartoon VQA," arXiv preprint arXiv:2601.03073, 2026.
[14] H. L. Kundel and C. F. Nodine, "Visual search, error patterns, and complex decision making in radiology," in Eye Movements and Higher Psychological Functions, 1978, pp. 311–336.
[15] C. F. Nodine and H. L. Kundel, "The nature of visual expertise in radiology," J. Radiol. Technol., vol. 58, no. 4, pp. 303–310, 1987.
[16] Y. Zhou, et al., "Large Language Models are Human-Level Prompt Engineers," in Proc. Int. Conf. Learn. Representations (ICLR), 2023.
[17] J. Schulman, et al., "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.
[18] R. Rafailov, et al., "Direct Preference Optimization: Your Language Model is Secretly a Reward Model," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2023.
[19] X. Zhang, et al., "PMC-VQA: Visual Instruction Tuning for Medical Visual Question Answering," in Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv. (MICCAI), 2023.
[20] J. Wang, et al., "MedXQA: A Large-scale Multi-modal Benchmark for Medical Diagnostics," in Proc. Joint Int. Conf. Comput. Linguistics, Lang. Resour. Eval. (LREC-COLING), 2024.
[21] W. Yue, et al., "MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert-level AI," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024.
[22] Y. Chen, et al., "OmniMedVQA: A Comprehensive Benchmark for Medical Multi-modal Learning," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024.
[23] C. Li, et al., "LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2023.
[24] T. Tu, et al., "Towards Generalist Biomedical AI," NEJM AI, vol. 1, no. 3, pp. 185–194, 2024.
[25] M. Li, et al., "MMedAgent: Multimodal Medical Agent System," in Proc. Conf. Empir. Methods Nat. Lang. Process. (EMNLP), 2024.
[26] J. Li, et al., "Agent Hospital: A Simulation Platform for Medical Agents," in Proc. ACM Int. Conf. Multimedia (ACM MM), 2024.
[27] A. Zeng, et al., "Socratic Models: Composing Zero-Shot Multimodal Reasoning with Language," in Proc. Int. Conf. Learn. Representations (ICLR), 2023.
[28] P. Croskerry, "Clinical cognition and diagnostic error: applications of a dual process model of reasoning," Adv. Health Sci. Educ., vol. 14, no. 1, pp. 27–35, 2009.