MEDROUTE: DYNAMIC SPECIALIST ROUTING FOR MULTI-AGENT MEDICAL DIAGNOSIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Large Multimodal Models (LMMs) have shown promising diagnostic ability in medical reasoning, yet they are typically used as monolithic generalists or as agents assigned to static clinical roles. Existing multi-agent medical systems improve specialization through fixed expert panels, role-based discussion, or external knowledge augmentation, but their coordination is often weakly adaptive and not directly optimized for final diagnostic correctness. We propose `MedRoute`, a reinforcement-learning-based framework that formulates multi-agent medical diagnosis as a learned consultation policy. For each case, `MedRoute` constructs a compact, case-conditioned specialist pool and instantiates specialists with targeted clinical guidance. A General Practitioner (GP) then dynamically selects specialists conditioned on the input and accumulated diagnostic history, allowing each consultation to inform the next rather than producing individual expert opinions. To address the lack of step-wise supervision signals, we train the router with an outcome-level policy-gradient objective using group-relative advantage normalization, assigning credit to consultation trajectories based on final diagnostic correctness. A Moderator synthesizes the resulting diagnostic trajectory into the final prediction. Across three text-only and four image-text medical QA benchmarks, `MedRoute` consistently outperforms both single-model and multi-agent baselines, with gains of up to **+22.5%** on text-only and **+34.3%** on image-text over the single-model baseline. Our training code will be made publicly available.

1 INTRODUCTION

Large Multimodal Models (LMMs) have shown strong potential for medical reasoning, including clinical question answering, diagnostic image interpretation, report generation, and multimodal medical decision support (Singhal et al., 2023; Li et al., 2023a; Luo et al., 2023). By combining textual clinical context with visual inputs such as radiographs, CT scans, pathology slides, and other medical images, these models can produce clinically relevant diagnoses and explanations across a broad range of medical tasks (Luo et al., 2022; Singhal et al., 2023; Zhang et al., 2024a; Sellergren et al., 2026). However, medical diagnosis is rarely a single-pass reasoning problem (Tang et al., 2024; Kim et al., 2024; Zuo et al., 2025; Wang et al., 2025). The relevant expertise depends on the symptoms, anatomy, imaging modality, and evolving evidence of each case, and the correct diagnostic path may require consulting different specialists in different orders (Zhou et al., 2025; Chen et al., 2025). This creates a mismatch between how LMMs are commonly used, either as generalists or as agents assigned to fixed roles, and how clinical diagnosis is actually performed. A natural way to reduce this mismatch is to formulate medical reasoning as a multi-agent process. In clinical practice, a general practitioner may first assess the case, identify the likely area of concern, and then consult one or more specialists. Effective diagnosis requires more than access to specialist knowledge. It also requires a mechanism for deciding which specialist to consult next, how to condition that decision on previous findings, and when enough evidence has been gathered to produce a final answer.

Recent medical multi-agent systems have begun to explore this direction. MedAgents (Tang et al., 2024) uses role-playing LLM experts in a multi-round collaborative discussion framework for zero-shot medical reasoning. MDAgents (Kim et al., 2024) adapts the collaboration structure of LLM teams according to estimated medical task complexity. MAM (Zhou et al., 2025) decomposes multimodal medical diagnosis into predefined roles and incorporates external knowledge retrieval. These approaches show that role specialization and agent collaboration can improve medical reasoning.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

However, they largely treat coordination as a prompting, discussion, retrieval, or workflow design problem. They do not explicitly learn which sequence of specialists is most useful for solving a given case, nor do they condition each routing decision on the evolving diagnostic history or optimize ordering and stopping from downstream diagnostic outcomes.

In this work, we propose MedRoute, a reinforcement-learning-based framework that formulates multi-agent medical diagnosis as an outcome-optimized consultation routing problem. For each specific case, MedRoute defines a compact specialist pool and learns to route over it dynamically, rather than treating the generated expert set as a one-time panel for discussion. A General Practitioner router selects the next specialist conditioned on the input case and accumulated diagnostic history, enabling each intermediate opinion to influence subsequent consultations. To address the lack of step-wise supervision signals, we optimize specialist selection and ordering from final diagnostic outcomes using an outcome-level policy-gradient objective with group-relative advantage normalization (Shao et al., 2024; Guo et al., 2025). A Moderator agent synthesizes the resulting diagnostic trajectory into the final prediction output. An illustrative example is provided in Fig. 1 to showcase the advantage of MedRoute in a specific case. Our evaluation covers textbook-style medical QA and diverse imaging modalities (CT, X-ray, histopathology). Across three text-only and four image-text benchmarks, MedRoute exceeds the strongest single-model baseline by up to **+22.5%** on PubMedQA (text-only) and up to **+34.3%** on BT-MRI (image-text), and show significant improvements on other medical datasets.

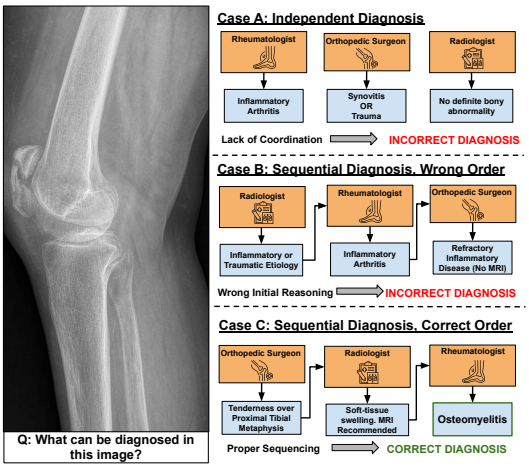

Figure 1: **History-conditioned specialist routing for medical diagnosis.** The same knee X-ray can lead to different outcomes depending on how specialist consultations are coordinated. Poorly ordered consultations may produce incorrect diagnoses, whereas MedRoute uses accumulated diagnostic history to select the next specialist and construct a coherent consultation trajectory, leading to the correct osteomyelitis prediction.

Our contributions can be summarized as follows:

- We introduce MedRoute, a novel and adaptive multi-agent framework that formulates multimodal medical diagnosis as an outcome-optimized consultation routing problem, where the agentic system can learn the most suitable sequence of specialist consultations based on the evolving diagnostic history.
- We design a router, namely the General Practitioner, which adaptively conditions each routing decision on the input case and accumulated diagnostic history, allowing specialist opinions to guide subsequent consultations rather than being generated independently.
- We train the router from final diagnostic outcomes using group-relative policy optimization and demonstrate substantial gains across seven medical QA benchmarks (three text-only, four image-text), outperforming both single-model and multi-agent state-of-the-art baselines.

2 Related Work

Medical Large Language and Multimodal Models. Large language and multimodal models have been widely studied for medical reasoning. Medical LLMs such as BioGPT (Luo et al., 2022) and MedAlpaca (Han et al., 2023) demonstrated the value of medical language modeling for biomedical and clinical tasks, while recent clinical and general-purpose LLMs have shown strong performance on medical knowledge and reasoning benchmarks (Singhal et al., 2023; Achiam et al., 2023). In vision-language settings, LLaVA-Med (Li et al., 2023a) introduced visual instruction tuning for biomedical images, BioMedGPT (Zhang et al., 2024a) developed a generalist biomedical vision-language foundation model across diverse medical tasks, and Meditron (Chen et al., 2023) trains large models achieving strong performance on medical reasoning benchmarks. Other systems focus on specialized imaging domains, including radiology-oriented models such as RadFM (Wu et al.,

2025) and RadVLM,(Deperrois et al., 2025), as well as medical vision-language assistants such as HuatuoGPT-Vision (Chen et al., 2024) and MedGemma (Sellergren et al., 2026). These models provide strong medical reasoning backbones, but they are typically used as single-model predictors or task-specific specialists. `MedRoute` is complementary to these advances: instead of proposing a new medical foundation model, it studies how existing LMM agents can be coordinated through an outcome-optimized specialist routing policy.

Multi-Agent Medical Diagnosis. LLM-based multi-agent systems have demonstrated significant potential in decomposing complex tasks through role-based collaboration, structured communication, and adaptive workflows (Kulkarni et al., 2020; Yao et al., 2022; Li et al., 2023b; Wang et al., 2024). Building on these general-purpose frameworks, recent work in the medical domain has integrated specialist roles, such as experts for individual analysis, consensus-based discussion, and report summarization, to enhance clinical reasoning (Tang et al., 2024; Zhou et al., 2025; Chen et al., 2025). For instance, systems like MDAgents (Kim et al., 2024) adapt collaboration structures based on estimated task complexity, while others leverage external retrieval and knowledge graphs to provide specialized medical context (Zuo et al., 2025; Wang et al., 2025). While these approaches show that role specialization and consensus mechanisms can improve diagnostic performance, they largely treat coordination as a static workflow design or rely on generic model orchestration. Existing systems typically rely on manually specified roles, predefined task templates, or fixed discussion protocols, lacking the ability to dynamically learn the most effective routing or synthesis strategy directly from diagnostic outcomes.

Learning to Route Specialist Agents. Routing has been widely studied in modular learning, mixture-of-experts models (Shazeer et al., 2017; Zhou et al., 2022), tool selection (Schick et al., 2023), and agent coordination (Wu et al., 2024; Fourney et al., 2024), where a controller selects among available modules or experts. In the context of LLMs, recent reinforcement learning methods have shown that outcome-level feedback can improve reasoning behavior even when intermediate supervision is unavailable (Ouyang et al., 2022). Group-relative policy optimization and related advantage-normalized policy-gradient methods have been used to train reasoning models from final rewards rather than dense step-wise labels (Shao et al., 2024; Guo et al., 2025). This setting is closely related to specialist consultation: medical QA datasets provide the final correct answer, but not the correct specialist sequence, consultation order, or stopping point. `MedRoute` applies this principle to multi-agent medical diagnosis. The router receives feedback only after the consultation trajectory is completed and a final answer is produced. By sampling multiple specialist trajectories for the same case and normalizing rewards within this group, `MedRoute` encourages the router to prefer consultation sequences that outperform alternative specialist orderings for the same input. This differs from static expert panels, retrieval-augmented workflows, and discussion-based aggregation by directly optimizing the specialist routing policy from final diagnostic outcomes.

3 METHOD

We propose a dynamic multi-agent framework (Fig. 2) for medical question answering in which a General Practitioner (GP) agent assembles and coordinates a small team of case-specific specialists. Given a medical case $x = (q, I)$, where q is the textual question and I is an optional medical image, the system first generates a set of relevant specialist roles. A learned router then dynamically selects which specialists to consult based on the case representation and the diagnostic history accumulated so far. Each selected specialist produces an intermediate opinion, and a Moderator agent synthesizes the full diagnostic record into the final answer $\hat{y}$.

3.1 CASE-CONDITIONED SPECIALIST POOL GENERATION

Medical questions often require different forms of expertise depending on symptoms, image findings, anatomy, and reasoning type. Before routing begins, `MedRoute` instantiates a custom candidate specialist pool for the current case. A frozen LMM-based specialist generator $G(\cdot)$ receives the input case and outputs a compact set of relevant specialist pool $\mathcal{S}_x = \{s_1, \ldots, s_{K_x}\} = G(q, I)$, where K_x is the number of roles generated for case x. For text-only cases, the generator receives only q. The generator is instructed to propose roles that would provide useful clinical perspectives, but not to answer the question itself. Each generated role is represented as $s_i = (n_i, \rho_i)$, where n_i is the specialist name and ρ_i is a short natural-language role description. Thus, a role specifies not only a broad specialty such as radiology, but also the particular perspective the agent should apply when interpreting the case. See Appendix Section B for examples and prompts used. These generated roles define the decision space for the proposed specialist router.

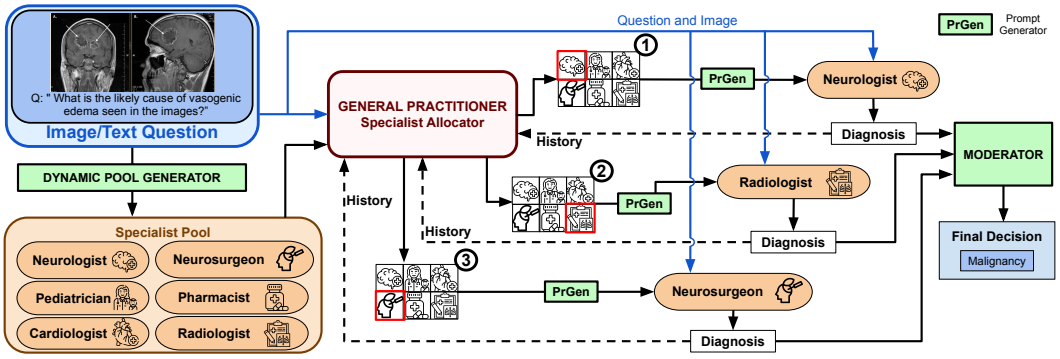

Figure 2: **Schematic illustration of our dynamic multi-agent framework.** The question and image are input to the General Practitioner (GP) Agent, a learned router for specialist allocation. From a case-conditioned specialist pool, the GP selects the first specialist based on the question (Neurologist in this example). The Neurologist's diagnosis is appended to the diagnostic history and used by the GP to route to the next specialist (Radiologist). The Radiologist's diagnosis similarly conditions the GP's selection of a third specialist (Neurosurgeon). The process repeats until the GP determines no further consultation is needed. Finally, all diagnoses from the selected specialists are passed to the Moderator agent, which synthesizes them and outputs the final decision.

3.2 General Practitioner

We implement the GP's routing process with a learned Specialist Allocator/Router, illustrated in Fig. 3. At each consultation step, the router observes the case, the specialist roles $\mathcal{S}_x$, and the accumulated diagnostic history, then selects the next agent from the available specialists and a decision role s_{dec}, which terminates specialist consultation once sufficient perspectives have been gathered.

Router inputs. Effective specialist allocation requires reasoning about both the clinical content of the case and the current state of the consultation. The router therefore receives three forms of context: a task embedding representing the original case, embeddings of the generated specialist roles, and an embedding of the diagnostic history accumulated so far. The task embedding is constructed from the available input modalities. The question q is encoded with a frozen text embedding model ϕ_T. When an image is present, I is additionally encoded with a frozen image embedding model ϕ_I. The available embeddings are projected to the routing transformer dimensionality using the learned task projection $TE(\cdot)$: for image-text cases, $z_x = TE([\phi_T(q); \phi_I(I)])$, while for text-only cases, only $\phi_T(q)$ is used. Each generated specialist role is embedded as $z_i^s = SRE(\phi_T(s_i))$, where $SRE(\cdot)$ projects specialist-role text into the router input space. $z_{\text{dec}}^s = SRE(\phi_T(s_{\text{dec}}))$ is the embedded decision-role token. At decision step t, the history record $H_t = [(a_1, d_1), \ldots, (a_{t-1}, d_{t-1})]$, where a_j is the selected specialist and d_j is its response, is embedded as $z_t^H = HE(\phi_T(H_t))$, where $HE(\cdot)$ is the history embedding projection.

Routing decision. At step t, the router chooses among generated specialists that have not yet been consulted, along with the decision role that terminates specialist consultation and passes the diagnostic record to the Moderator, defining the available action set as $\mathcal{A}_t = (\mathcal{S}_x \setminus \mathcal{C}_t) \cup \{s_{\text{dec}}\}$, where $\mathcal{C}_t$ is the set of previously consulted specialists. Previously selected specialists are masked to prevent redundant consultations.

To select the next action, the router introduces a specialist-selection token $z_t^{\text{sel}} = SP(z_x)$, where $SP(\cdot)$ is a learned projection from the task embedding. This token represents the GP's current query for the *next* specialist. The router then contextualizes the task, candidate role, history, and selection tokens with a routing transformer and scores each valid action by comparing its contextualized role token $\widetilde{z}_{t,i}^s$ to the contextualized selection token $\widetilde{z}_t^{\text{sel}}$, then defines the agent selection policy π_θ:

$$Z_t = [z_x;\ z_1^s;\ \ldots;\ z_{K_x}^s;\ z_{\text{dec}}^s;\ z_t^H;\ z_t^{\text{sel}}], \quad \widetilde{Z}_t = RT_\theta(Z_t), \quad \ell_{t,i} = R_\theta\left(\widetilde{z}_{t,i}^s - \widetilde{z}_t^{\text{sel}}\right),\ i \in \mathcal{A}_t,$$
$$\pi_\theta(a_t = i \mid x, H_t, \mathcal{S}_x) = \frac{\exp(\ell_{t,i})}{\sum_{j \in \mathcal{A}_t} \exp(\ell_{t,j})}. \tag{1}$$

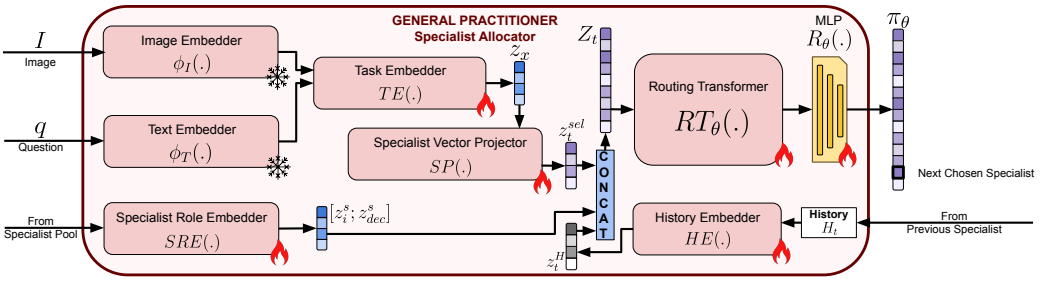

Figure 3: **Schematic illustration of the General Practitioner.** For each case, the router receives the question and, when available, an image. The available modality embeddings are fused by the task embedding projection $TE(\cdot)$ to produce the task token. Generated specialist roles are embedded by the specialist-role embedder $SRE(\cdot)$ to form candidate agent tokens. The task token is also projected by $SP(\cdot)$ to initialize a specialist-selection token, which represents the GP's query for the next action. At each routing step, the router combines the task token, candidate specialist tokens, diagnostic history embedding from $HE(\cdot)$, and the specialist-selection token, then contextualizes them with the routing transformer $RT_\theta(\cdot)$. The scoring MLP $R_\theta(\cdot)$ compares the contextualized selection token with each candidate token to produce routing logits over valid roles.

Here, $RT_\theta(\cdot)$ is the routing transformer, and $R_\theta(\cdot)$ is the scoring MLP. If no specialist has been consulted yet, the history token is omitted from Z_t. During training, the next agent is sampled from π_θ; during inference, the GP selects the highest-probability valid action.

3.3 Sequential Diagnostic Consultation

When the router selects an agent $a_t \in \mathcal{S}_x$, the corresponding role specification is used to instantiate a specialist LMM agent A_{a_t}. The specialist receives the original case and current diagnostic record, producing a response $d_t = A_{a_t}(q, I, H_t)$. This response is appended to the ordered diagnostic record, yielding $H_{t+1} = H_t \oplus [(a_t, d_t)]$, where $\oplus$ denotes record concatenation. This sequential consultation process allows later routing decisions to depend on earlier specialist findings. For example, an initial specialist may identify an abnormal image pattern, suggest a differential diagnosis, or reveal that an additional clinical perspective is needed. The GP can then condition its next routing decision on the updated record rather than committing to all specialist consultations upfront.

The process continues until the router selects s_{dec} or the maximum consultation budget $T_{\max}$ is reached. Let T denote the number of completed specialist consultations. The Moderator then synthesizes the original case and final diagnostic record $H_T = [(a_1, d_1), \ldots, (a_T, d_T)]$ into the final answer $\hat{y} = M(q, I, H_T)$. The Moderator is instructed to integrate specialist opinions, resolve disagreements when present, and output the answer in the required format.

3.4 Training the General Practitioner's Router

The usefulness of a routing decision is only observed after the consultation trajectory is completed and the Moderator produces a final answer. This creates a delayed credit assignment problem: an early specialist may be useful because it changes later routing decisions, even if its individual response is not directly the final answer. We therefore train the router with a policy-gradient objective using group-relative advantage normalization. For each training case x, we first generate its specialist set $\mathcal{S}_x$. We then sample G consultation trajectories from the current routing policy, where $\tau^{(g)} = [(a_1^{(g)}, d_1^{(g)}), \ldots, (a_{T_g}^{(g)}, d_{T_g}^{(g)})]$ and T_g is the number of specialists consulted before selecting s_{dec} or reaching $T_{\max}$. Each trajectory is passed to the Moderator to produce a final answer $\hat{y}^{(g)}$.

We compute an outcome reward using a reward-model judge LMM $RM(\cdot)$ that compares the predicted answer with the ground-truth answer y. The judge returns 1 when the prediction is semantically correct and 0 otherwise. To encourage concise consultation, we apply a length decay:

$$R^{(g)} = \gamma^{T_g} RM(\hat{y}^{(g)}, y), \qquad \gamma \in (0, 1]. \tag{2}$$

This reward preserves final-answer correctness as the primary signal while discouraging unnecessary specialist calls. Additionally, since cases differ substantially in difficulty, raw rewards can have high variance across inputs. We normalize rewards across trajectories sampled for the same case: $A^{(g)} = \frac{R^{(g)} - \mu_R}{\sigma_R + \epsilon}$, where μ_R and σ_R are the mean and standard deviation of $\{R^{(1)}, \ldots, R^{(G)}\}$, and ϵ

is a small constant for numerical stability. This group-relative advantage encourages the router to prefer trajectories that perform better than alternative routing choices for the same case.

The router is optimized with the stepwise policy-gradient loss:

$$\mathcal{L}_{\text{route}}(\theta) = -\frac{1}{G}\sum_{g=1}^{G} A^{(g)} \sum_{t=1}^{T_g} \log \pi_\theta \left(a_t^{(g)} \mid x, H_t^{(g)}, \mathcal{S}_x \right). \tag{3}$$

During training, only the input projection layers, routing transformer, and final scoring head belonging to the General Practitioner are trainable. The remaining agents and components are frozen.

3.5 Inference Workflow

At inference time, `MedRoute` receives a case $x = (q, I)$. The specialist generator first proposes a case-specific set of specialist roles $\mathcal{S}_x$. The available input modalities are embedded and projected to produce the task token z_x, and the generated roles are embedded to form the candidate specialist tokens. Starting from an empty diagnostic record H_1, the GP repeatedly applies the learned router to select either an unconsulted specialist or the decision role s_{dec}. Each selected specialist produces an intermediate diagnostic response, which is appended to the record and used to condition subsequent routing decisions. Once the router selects s_{dec} or the consultation budget $T_{\max}$ is reached, the Moderator receives the complete diagnostic record and returns the final answer $\hat{y}$. Prompts used for Specialist agents and the Moderator agent are shown in Appendix Section C.

4 Experimental Results and Discussion

Fig. 4 illustrates a complete inference run for a cardiology image-text case on `MedRoute` pipeline. Given the question and the axial CT plus ECG image, the case-conditioned generator $G(\cdot)$ instantiates a specialist pool $\mathcal{S}_x$, and the General Practitioner dynamically selects the specialists and routes through *Cardiologist*, *Thoracic Surgeon*, and *Hematologist*. Each diagnosis is appended to the diagnostic history, allowing later routing decisions to condition on earlier findings. The Moderator then synthesizes the three opinions into the final diagnosis, "*prominent enlargement of the right atrium and right ventricle*.".

Table 1: Evaluation dataset statistics.

Dataset	Modality	Test Size
MedQA (Jin et al., 2021)	Text only	1,273
MedMCQA (Pal et al., 2022)	Text only	2,816
PubMedQA (Jin et al., 2019)	Text only	1,000
PMC-VQA (Zhang et al., 2024b)	Image-text	2,000
PathVQA (He et al., 2020)	Image-text	6,719
DeepLesion (Yan et al., 2018)	Image-text	4,736
BT-MRI (Bhuvaji et al., 2025)	Image-text	394

Evaluation datasets. We evaluate on seven medical QA benchmarks chosen to span question formats and image modalities encountered in clinical reasoning (Table 1). The *text-only* suite covers three datasets: *MedQA* (Jin et al., 2021) (USMLE-style five-option clinical vignettes), *MedMCQA* (Pal et al., 2022) (four-option questions from Indian medical entrance exams across 21 medical subjects), and *PubMedQA* (Jin et al., 2019) (yes/no/maybe questions grounded in PubMed abstracts). The *image-text* suite spans four visual modalities: *PMC-VQA* (Zhang et al., 2024b) (PubMed Central figures, four-option MCQ), *PathVQA* (He et al., 2020) (histopathology, open-ended), *DeepLesion* (Yan et al., 2018) (CT, eight lesion sites), and *BT-MRI* (Bhuvaji et al., 2025) (brain MRI, tumor sub-types). DeepLesion and BT-MRI ship with only class-level labels, which we template into multiple-choice questions whose options are the dataset's full class set (Appendix D). For each dataset, the router is trained on 300 questions sampled uniformly at random from the train split, and evaluated on the full official test split for MedQA, PubMedQA, PMC-VQA, BT-MRI, and PathVQA; for MedMCQA and DeepLesion we restrict evaluation to instances with a unique correct label (2,816, and 4,736 instances respectively), as our correctness signal assumes a single gold option.

Backbone selection. All specialist agents and the Moderator are instantiated with the strongest single-model baseline per modality (Fig. 5): Qwen3-8B (Yang et al., 2025) for text-only and Qwen3.5-9B (Yang et al., 2025) for image-text. This isolates the contribution of routing from raw model capability and keeps the comparison against zero-shot baselines fair.

Dynamic specialist and prompt generator. The case-conditioned specialist generator $G(\cdot)$ in Section 3.1, which produces both the candidate specialist pool $\mathcal{S}_x$, and the dynamic per-specialist prompts are instantiated. $G(\cdot)$ is queried once per case and cached, and its outputs are restricted to (i) specialist titles and short role descriptions, and (ii) natural-language pointers indicating which clinical features, mechanisms, or differentials, each specialist should attend to; it never produces an answer or option label. All diagnostic reasoning, intermediate opinions, and the final decision are

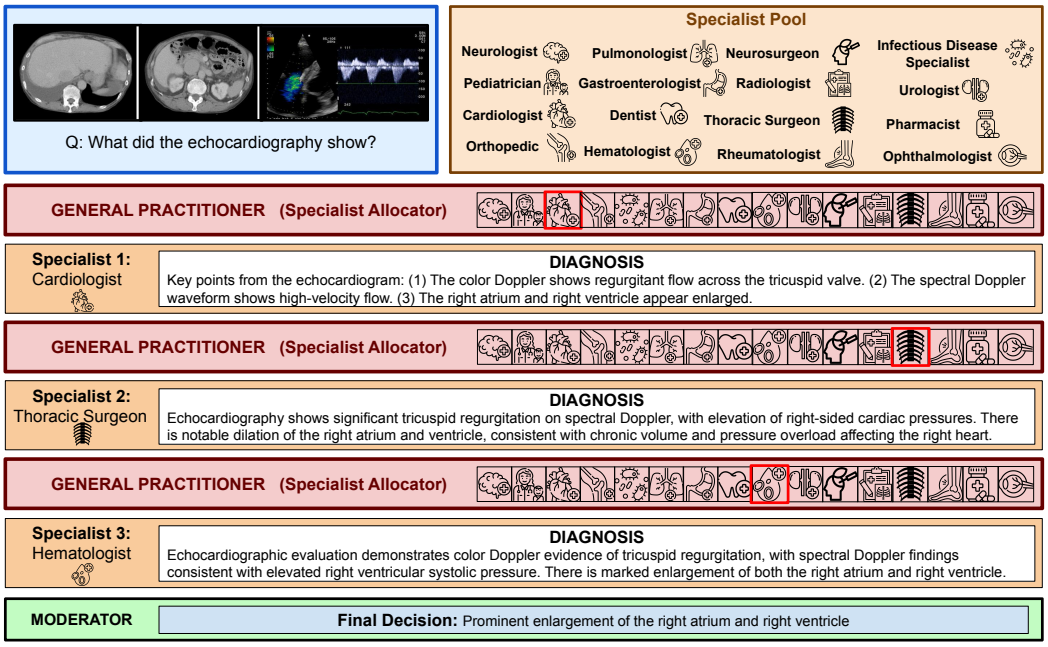

Figure 4: **Qualitative example.** Given a cardiology case with axial CT and an ECG, the case-conditioned generator $G(\cdot)$ proposes a specialist pool $\mathcal{S}_x$ and the *General Practitioner* routes dynamically: a Cardiologist first interprets the ECG; conditioned on that diagnosis, the *GP* next routes to a Thoracic Surgeon, whose diagnosis is consistent with the cardiologist's; the *GP* finally then routes to a Hematologist for an additional perspective. The three opinions, accumulated in the diagnostic record, are synthesized by the *Moderator* into the final answer "*prominent enlargement of the right atrium and right ventricle*." The example illustrates how earlier specialists findings, condition later routing decisions, and converge on a coherent diagnosis.

produced exclusively by the agent-level backbone, so $G(\cdot)$ acts as a lightweight scoping module that shapes the diagnosis while heavier per-rollout reasoning remains dependent on agent backbone.

Optimization and architecture. The router is implemented in PyTorch (Paszke et al., 2019): $RT_\theta(\cdot)$ is a pretrained GPT-2 small (Radford et al., 2019) (12 blocks), and $R_\theta(\cdot)$ is a two-layer MLP with ReLU (Agarap, 2018); all projections use a hidden dimension of 768. We use a frozen sentence-transformer (Reimers & Gurevych, 2019) as ϕ_T and, for image-text cases, frozen SigLIP (Zhai et al., 2023) (`siglip-so400m-patch14-384`, 1152-d) as ϕ_I, with both modalities L2-normalized before concatenation to balance their magnitudes. Training uses AdamW (Loshchilov & Hutter, 2019) ($\text{lr} = 3 \times 10^{-5}$) for one epoch over 300 sampled questions with 16 rollouts per question, rollout temperature $\tau = 0.7$, reward discount $\gamma = 0.98$, and entropy regularization (coefficient 0.05) to discourage routing collapse; the context length is 2,048 tokens. All LMM calls are served via vLLM (Kwon et al., 2023) on two NVIDIA A100 80GB GPUs (one for the agent backbone, one for $G(\cdot)$), with rollouts run in parallel.

Baselines. We compare `MedRoute` against three classes of baselines, general-purpose single-model LMMs, medical single-model LMMs and medical multi-agent frameworks.

(i) General-purpose and medical single-model LMMs establish the strongest non-routing baseline per modality: for the text-only suite we evaluate Qwen3-8B (Yang et al., 2025), Llama-3-/-3.1-8B (Grattafiori et al., 2024), Mistral-7B (Jiang et al., 2023), Phi-4-mini (Abdin et al., 2024b), MiMo-7B-Base (Xiaomi et al., 2025), Gemma-7B (Team et al., 2024), and DeepSeek-LLM-7B (Bi et al., 2024), alongside the medically fine-tuned MedAlpaca-7B (Han et al., 2023), Meditron-7B (Chen et al., 2023) and Medichat-Llama3-8B (Iyer, 2024). For image-text, we evaluate Qwen3.5-9B and Qwen3-VL-8B (Yang et al., 2025), Qwen2.5-VL (Bai et al., 2025), LLaVA-OneVision (Li et al.), and Phi-3.5-Vision (Abdin et al., 2024a), plus medical VLMs MedGemma-1.5-4B (Sellergren et al., 2026), HuatuoGPT-Vision-7B (Zhang et al., 2023a), and LLaVA-Med-v1.5 (Li et al., 2023a). Figure 5 reports their accuracies on MedQA, MedMCQA, and PMC-VQA, motivating our choice of Qwen3-8B and Qwen3.5-9B as the agent backbones.

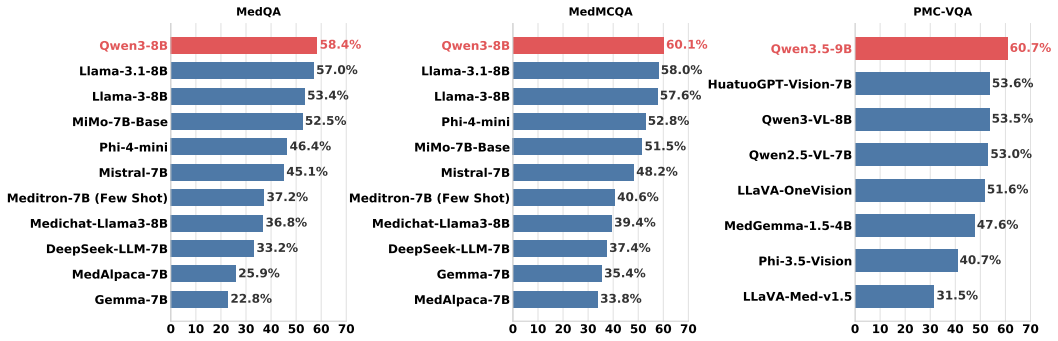

Figure 5: **Single-model baseline accuracies** on MedQA and MedMCQA (text-only LLMs) and PMC-VQA (image-text VLMs). Bars are sorted by accuracy within each panel. The strongest model per modality, *Qwen3-8B* for text-only and *Qwen3.5-9B* for image-text and selected as the agent backbone in `MedRoute`, which keeps the comparison against multi-agent frameworks fair.

(ii) Medical multi-agent frameworks: we adapt MAC (Chen et al., 2025), MedAgents (Tang et al., 2024), MDAgents (Kim et al., 2024) and MAM (Zhou et al., 2025) to use the same backbone as `MedRoute`, ensuring that any improvement is attributable to routing rather than backbone capability. MAC and MedAgents are text-only frameworks, while MDAgents and MAM operate in both text-only and image-text settings.

Text-Only Datasets. Table 2 shows that `MedRoute` attains the highest score on all three text-only datasets, improving over the Qwen3-8B backbone by **+14.1%** on MedQA, **+9.5%** on MedMCQA, and **+22.5%** on PubMedQA, and over the strongest multi-agent baseline by **+2.1%**, **+5.7%**, and **+7.9%** respectively. The gains align with how each framework traverses its specialist pool: MAC (Chen et al., 2025) and MedAgents (Tang et al., 2024) run a round-table over a one-shot panel with static role prompts. MDAgents (Kim et al., 2024) introduces a difficulty tier system and an advance structure for multi-agent deliberation, while MAM (Zhou et al., 2025) follows a fixed GP → Specialist → Director workflow. None of these learn the consultation policy from final-answer outcomes. `MedRoute` adds three independent sources of adaptivity, and Table 6 decomposes the +14.1% MedQA gain into **+5.7%** from the RL-trained router, **+3.5%** from dynamic per-specialist guidance, and **+4.9%** from case-conditioned pool generation.

Table 2: **Text-only diagnostic accuracy.** `MedRoute` attains the highest score on all three datasets; sharing the similar Qwen3-8B backbone.

Model	MedQA	MedMCQA	PubMedQA
Baseline (Qwen3-8B)	58.4	60.1	50.3
MAC (Chen et al., 2025)	60.6	60.7	55.6
MedAgents (Tang et al., 2024)	70.4	61.4	61.0
MDAgents (Kim et al., 2024)	65.6	63.1	48.4
MAM (Zhou et al., 2025)	67.6	63.9	64.9
`MedRoute`	**72.5**	**69.6**	**72.8**

Image-Text Datasets. Table 3 reports accuracy on five image-text datasets. `MedRoute` improves over the Qwen3.5-9B backbone on every dataset, improving by **+3.4%** on PMC-VQA, **+8.0%** on PathVQA, **+14.0%** on DeepLesion, and **+34.3%** on BT-MRI. `MedRoute` outperforms state-of-the-art models MDAgents and MAM on all four datasets. In `MedRoute`, the router conditions its decisions on a SigLIP image embedding and the question, the VLM-based generator $G(\cdot)$ proposes specialists matched to the imaging modality (e.g., Neuroradiologist for brain MRI, Thoracic Radiologist for chest X-ray), and each consulted specialist receives the original image. The largest gains appear on the narrow-modality, fine-grained discrimination tasks BT-MRI, where matching the panel to the imaging modality matters most. Additional ablations on the image embedding types and embedding models can be found in Appendix Sections A.1 and A.2 respectively.

Table 3: **Image-text diagnostic accuracy.** `MedRoute` attains the highest score on all five datasets; all methods share the Qwen3.5-9B agent backbone.

Model	PMC-VQA	PathVQA	DeepLesion	BT-MRI
Baseline (Qwen3.5-9B)	60.7	37.7	37.7	42.9
MDAgents (Kim et al., 2024)	59.1	40.4	51.1	69.9
MAM (Zhou et al., 2025)	61.9	44.3	50.7	71.8
`MedRoute`	**64.1**	**45.7**	**51.7**	**77.2**

Learned Routing Policy. Reinforcement learning is used because there is no labeled ground-truth routing between available specialists. In fact, the diagnosis of a previous doctor may impact the next relevant specialist, so even a set path for specific conditions may be suboptimal across patients. To isolate this behavior, we train a supervised variant of our router by assuming the first three specialists generated from the dynamic pool are the optimal order. Additionally, we evaluate a random-order variant where the specialist path is sampled completely at random. Table 4 shows that both supervised finetuning and random routing underperforms our RL variant, indicating that RL can discover unique, highly effective specialist paths.

Table 4: **Routing policy analysis.** Random routing and supervision with a predetermined order (SFT) underperforms RL exploration.

Routing Strategy	MedQA
Baseline (Qwen3-8B)	58.4
Random order	66.9$_{+8.5}$
SFT-route	70.1$_{+11.7}$
MedRoute	**72.5**$_{+14.1}$

Specialist Consultation Distribution. We analyze the number of consultation steps taken for each case. The distribution varies substantially across datasets (Fig. 6); the router consults three specialists for 95.2% of PathVQA cases, while consulting just one specialist for 86.8% of BT-MRI cases. Our flexible routing design neatly captures the variation in optimal number of specialists per case. PathVQA and PMC-VQA are very general and the datasets become more specific as we go down the plot, which exactly highlights the need of more specialists when the issues are more diverse. Furthermore, we verify that more consultations is not inherently better. In an experiment where we force 3 specialists in cases where 1 or 2 specialists were sampled, performance reduces by 1%-4% (Appendix A.4).

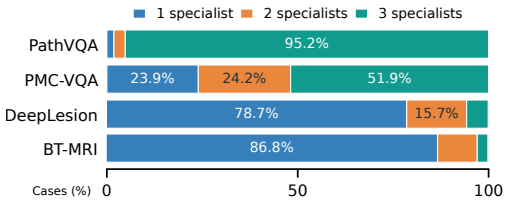

Figure 6: **Specialist consultations.** Distribution of specialist counts per case across each dataset. Optimal consultation amount differs per setting.

Cross-dataset Performance. To probe whether the specialist router overfits to dataset-specific routing or learns a transferable consultation policy, we train the router on one dataset and evaluate it directly on the other. Table 5 reports gains of **+12.6%** on MedMCQA→MedQA and **+7.3%** on MedQA→MedMCQA over the Qwen3-8B baseline, recovering nearly all of the same-dataset gain. The learned routing structure transfers across benchmarks, useful when per-dataset training data is impractical to collect.

Table 5: **Cross-dataset transfer.** Router trained on one dataset, evaluated directly on the other.

Model	MedQA	MedMCQA
Baseline (Qwen3-8B)	58.4	60.1
MedRoute (cross)	71.0	67.4
MedRoute (same)	72.5	69.6

Component-wise Analysis. Starting from the Qwen3-8B baseline, the RL-trained router with the fixed specialist pool and static prompts contributes **+5.7%** on MedQA; replacing the static role prompts with dynamic per-specialist prompts adds another **+3.5%**; replacing the fixed pool with case-conditioned pool generation adds a further **+4.9%**, for a total of **+14.1%** (Table 6).

Table 6: Text-only ablation on MedQA.

Configuration	Accuracy
Baseline (Qwen3-8B)	58.4
+ Routing (static prompts)	64.1$_{+5.7}$
+ Routing (dynamic prompts)	67.6$_{+9.2}$
+ Dynamic pool gen.	**72.5**$_{+14.1}$

5 Conclusion

We introduced MedRoute, a multi-agent framework for medical diagnosis that learns to coordinate specialist LMM agents through reinforcement learning. A *General Practitioner* router dynamically selects and orders specialist consultations conditioned on the case and accumulated history, and a *Moderator* synthesizes the diagnostic record into the final answer. Three sources of adaptivity drive the framework: case-conditioned specialist pool generation, dynamic per-specialist prompts, and the learned routing policy; for image-text cases, the router additionally consumes a SigLIP image embedding for visual grounding. Across seven medical QA benchmarks (three text-only, four image-text), MedRoute consistently outperforms both single-model and multi-agent baselines, with text-only gains of **+14.1%** on MedQA and **+9.5%** on MedMCQA, and image-text gains of **+34.3%** on BT-MRI and **+14.0%** on DeepLesion over the single-model baseline. Component ablations show each adaptivity source contributes independently, and cross-dataset transfer confirms generalization beyond the training distribution.

AI USE STATEMENT

In this work, we used generative AI tools for implementing methods. LLM-generated code was verified and tested for correctness by 3 authors. We have not used generative AI tools for generating datasets, formulating mathematical claims, or any of the rest of the required disclosure tasks. Additionally, we used generative AI tools for refining the manuscript. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the MedRoute architecture, training objective, and inference procedure in the main text, with implementation details, hyperparameters, prompts, dataset preprocessing, evaluation protocols, and additional ablations in the appendix. To facilitate independent reproduction and extension of this work, we plan to release our source code, training and evaluation configurations, generated specialist specifications, and trained router weights upon publication.

REFERENCES

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, et al. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*, 2024a.

Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J Hewett, Mojan Javaheripi, Piero Kauffmann, et al. Phi-4 technical report. *arXiv preprint arXiv:2412.08905*, 2024b.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

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

Sartaj Bhuvaji, Ankita Kadam, Prajakta Bhumkar, Sameer Dedge, and Swati Kanchan. Brain tumor classification (mri), 2025. URL `https://www.kaggle.com/dsv/12745533`.

Xiao Bi, Deli Chen, Guanting Chen, Shanhuang Chen, Damai Dai, Chengqi Deng, Honghui Ding, Kai Dong, Qiushi Du, Zhe Fu, et al. Deepseek llm: Scaling open-source language models with longtermism. *arXiv preprint arXiv:2401.02954*, 2024.

Junying Chen, Chi Gui, Ruyi Ouyang, Anningzhe Gao, Shunian Chen, Guiming Hardy Chen, Xidong Wang, Zhenyang Cai, Ke Ji, Xiang Wan, and Benyou Wang. Towards injecting medical visual knowledge into multimodal LLMs at scale. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 7346–7370, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

Xi Chen, Huahui Yi, Mingke You, WeiZhi Liu, Li Wang, Hairui Li, Xue Zhang, Yingman Guo, Lei Fan, Gang Chen, et al. Enhancing diagnostic capability with multi-agents conversational large language models. *NPJ digital medicine*, 8(1):159, 2025.

Zeming Chen, Alejandro Hernández Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, et al. Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*, 2023.

Nicolas Deperrois, Hidetoshi Matsuo, Samuel Ruipérez-Campillo, Moritz Vandenhirtz, Sonia Laguna, Alain Ryser, Koji Fujimoto, Mizuho Nishio, Thomas M Sutter, Julia E Vogt, et al. Radvlm: A multitask conversational vision-language model for radiology. *arXiv preprint arXiv:2502.03333*, 2025.

Adam Fourney, Gagan Bansal, Hussein Mozannar, Cheng Tan, Eduardo Salinas, Friederike Niedtner, Grace Proebsting, Griffin Bassman, Jack Gerrits, Jacob Alber, et al. Magentic-one: A generalist multi-agent system for solving complex tasks. *arXiv preprint arXiv:2411.04468*, 2024.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638, 2025.

Tianyu Han, Lisa C Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K Bressem. Medalpaca–an open-source collection of medical conversational ai models and training data. *arXiv preprint arXiv:2304.08247*, 2023.

Xuehai He, Yichen Zhang, Luntian Mou, Eric Xing, and Pengtao Xie. Pathvqa: 30000+ questions for medical visual question answering. *arXiv preprint arXiv:2003.10286*, 2020.

Sethu Iyer. Medichat-llama3-8b, 2024. URL `https://huggingface.co/sethuiyer/`. Hugging Face model page. Built upon the LLaMA-3 architecture.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023. URL `https://arxiv.org/abs/2310.06825`.

Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 11(14):6421, 2021.

Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William Cohen, and Xinghua Lu. Pubmedqa: A dataset for biomedical research question answering. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp. 2567–2577, 2019.

Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik S Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae W Park. Mdagents: An adaptive collaboration of llms for medical decision-making. *Advances in Neural Information Processing Systems*, 37: 79410–79452, 2024.

Parth Parag Kulkarni, Harsh Kasyap, and Somanath Tripathy. Dnet: An efficient privacy-preserving distributed learning framework for healthcare systems. In *International conference on distributed computing and internet technology*, pp. 145–159. Springer, 2020.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pp. 611–626, 2023.

Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, et al. Llava-onevision: Easy visual task transfer. *Transactions on Machine Learning Research*.

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, 2023a.

Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for" mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008, 2023b.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. Biogpt: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics*, 23(6):bbac409, 2022.

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

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

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.

Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. Medmcqa: A large-scale multi-subject multi-choice dataset for medical domain question answering. In *Conference on health, inference, and learning*, pp. 248–260. PMLR, 2022.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp. 3982–3992, 2019.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *Advances in neural information processing systems*, 36:68539–68551, 2023.

Andrew Sellergren, Chufan Gao, Fereshteh Mahvar, Timo Kohlberger, Fayaz Jamil, Madeleine Traverse, Alberto Tono, Bashir Sadjad, Lin Yang, Charles Lau, et al. Medgemma 1.5 technical report. *arXiv preprint arXiv:2604.05081*, 2026.

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

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. Large language models encode clinical knowledge. *Nature*, 620(7972):172–180, 2023.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. Medagents: Large language models as collaborators for zero-shot medical reasoning. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 599–621, 2024.

Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, et al. Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*, 2024.

Junlin Wang, Jue Wang, Ben Athiwaratkun, Ce Zhang, and James Zou. Mixture-of-agents enhances large language model capabilities. *arXiv preprint arXiv:2406.04692*, 2024.

Ziyue Wang, Junde Wu, Chang Han Low, and Yueming Jin. Medagent-pro: Towards multi-modal evidence-based medical diagnosis via reasoning agentic workflow. *arXiv e-prints*, pp. arXiv–2503, 2025.

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

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, et al. Autogen: Enabling next-gen llm applications via multi-agent conversations. In *First Conference on Language Modeling*, 2024.

LLM Xiaomi, Bingquan Xia, Bowen Shen, Dawei Zhu, Di Zhang, Gang Wang, Hailin Zhang, Huaqiu Liu, Jiebao Xiao, Jinhao Dong, et al. Mimo: Unlocking the reasoning potential of language model–from pretraining to posttraining. *arXiv preprint arXiv:2505.07608*, 2025.

Ke Yan, Xiaosong Wang, Le Lu, and Ronald M Summers. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging*, 5(3):036501–036501, 2018.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *The eleventh international conference on learning representations*, 2022.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 11975–11986, 2023.

Hongbo Zhang, Junying Chen, Feng Jiang, Fei Yu, Zhihong Chen, Guiming Chen, Jianquan Li, Xiangbo Wu, Zhang Zhiyi, Qingying Xiao, et al. Huatuogpt, towards taming language model to be a doctor. In *Findings of the association for computational linguistics: EMNLP 2023*, pp. 10859–10885, 2023a.

Kai Zhang, Rong Zhou, Eashan Adhikarla, Zhiling Yan, Yixin Liu, Jun Yu, Zhengliang Liu, Xun Chen, Brian D Davison, Hui Ren, et al. A generalist vision–language foundation model for diverse biomedical tasks. *Nature Medicine*, 30(11):3129–3141, 2024a.

Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. *arXiv preprint arXiv:2303.00915*, 2023b.

Xiaoman Zhang, Chaoyi Wu, Ziheng Zhao, Weixiong Lin, Ya Zhang, Yanfeng Wang, and Weidi Xie. Development of a large-scale medical visual question-answering dataset. *Communications Medicine*, 4(1):277, 2024b.

Yanqi Zhou, Tao Lei, Hanxiao Liu, Nan Du, Yanping Huang, Vincent Zhao, Andrew M Dai, Quoc V Le, James Laudon, et al. Mixture-of-experts with expert choice routing. *Advances in Neural Information Processing Systems*, 35:7103–7114, 2022.

Yucheng Zhou, Lingran Song, and Jianbing Shen. MAM: Modular multi-agent framework for multi-modal medical diagnosis via role-specialized collaboration. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 25319–25333, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-256-5.

Kaiwen Zuo, Yirui Jiang, Fan Mo, and Pietro Lio. Kg4diagnosis: A hierarchical multi-agent llm framework with knowledge graph enhancement for medical diagnosis. In *AAAI Bridge Program on AI for Medicine and Healthcare*, pp. 195–204. PMLR, 2025.

APPENDIX OVERVIEW

Section A: Additional Ablations

Section A.1 Captions vs. Image Embeddings

Section A.2: Choice of Image Embedding Model

Section A.3: Ablation on Number of Trajectories

Section A.4 Forced Consultation Extension

Section B: Dynamic Case-conditioned Specialist Pool Generation

Section C: Prompts used for Agent Calls

Section D: Dataset Details

Section D.1: DeepLesion

Section D.2: BT-MRI

Section D.3: PathVQA

Section D.4: Examples from Text Datasets

Section D.5: Examples from Vision Datasets

Section E: Additional Qualitative Examples

A ADDITIONAL ABLATIONS

In this section we include additional ablations for captions vs image embeddings, the choice of image embedding model, number of traces generated while training and forced consultation extension. First three experiments are performed on PMC-VQA for the image-text framework, while forced consultation extension is studied on all four image-text datasets.

A.1 CAPTIONS VS. IMAGE EMBEDDINGS

Extending our text-based router to PMC-VQA, we first feed each medical image to the agent VLMs to get an image caption. This caption-based variant reaches 63.2% (**+2.5%** over the Qwen3.5-9B baseline). Replacing the caption with a SigLIP (Zhai et al., 2023) image embedding fed to the router (Section 3.2) yields the final `MedRoute` system at 64.1% (**+3.4%**). Image embeddings preserve visual signal that captioning compresses away. The choice of vision encoder is studied in Appendix A.2.

Table 7: Image-text ablation on PMC-VQA.

Configuration	Accuracy
Baseline (Qwen3.5-9B)	60.7
Router with image captions	63.2$_{+2.5}$
Router with Image embeddings (SigLIP)	**64.1**$_{+3.4}$

A.2 CHOICE OF IMAGE EMBEDDING MODEL

The General Practitioner router in our framework includes an image embedder for image-text inputs (Section 3.2). As observed in Section A.1 above, replacing image captions with image embeddings improved performance. In our final framework, we use SigLIP(Zhai et al., 2023) encoder for obtaining the aforementioned image embedding. Here, we perform an ablation to assess different choices of image encoders. In Table 8, we see that our framework performs the best with SigLIP, leading us to the decision of our choice.

Table 8: Image embedding models ablation on PMC-VQA.

Image Embedding Model	Accuracy
CLIP(Radford et al., 2021)	63.1
BiomedCLIP(Zhang et al., 2023b)	63.2
DINOv2(Oquab et al., 2023)	63.6
SigLIP(Zhai et al., 2023)	**64.1**

A.3 Ablation on Number of Trajectories

As discussed in Section 3.4, we train the router with a policy-gradient objective using group-relative advantage normalization. For each training case, we sample G consultation trajectories from the current routing policy, and each trajectory is passed to the Moderator to make the final decision, which in turn is used to compute the reward. Increasing the number of trajectories G implies increasing computation time, and also improvement till a threshold after which it starts giving diminishing returns. We observe this trend in Table 9, where increasing the number of generated trajectories from 8 to 16 improves performance, but further increase to 32 leads to the model performing worse than the model trained with 8 trajectories.

Table 9: No. of trajectories ablation on PMC-VQA.

No. of Trajectories	Accuracy
8	63.2
16	**64.1**
32	62.4

A.4 Forced Consultation Extension

To examine whether the router stops consultations prematurely, we force cases assigned fewer than three specialists to continue until three have been consulted. As shown in Table 10, this reduces accuracy on all five datasets, by 1.0–4.0 percentage points. The largest decrease occurs on DeepLesion (-4.0 points), where 94.4% of cases were originally assigned fewer than three specialists. These results suggest that additional consultations do not necessarily improve the final answer and support the router's learned stopping decisions.

Table 10: **Effect of extending consultations.** Fraction of cases initially assigned fewer than three specialists and change in accuracy after forcing consultation with three specialists, relative to the original routing decisions.

Dataset	< 3 specialists (%)	Δ Acc. (pp)
BT-MRI	97.2	-2.4
DeepLesion	94.4	-4.0
PMC-VQA	48.1	-1.1
PathVQA	4.8	-1.0

B Dynamic Case-conditioned Specialist Pool Generation

As discussed in Section 3.1, we generate a pool of specialists emulating a real-world clinical setting. We dynamically custom candidate specialist pool for the current case at hand. We accomplish this with a frozen LMM-based specialist generator which is instructed to propose roles that would provide useful clinical perspectives, but not to answer the question itself.

The number of specialists k depends on the dataset. For general purpose QA/VQA datasets k is between 5-7. The prompt for generating specialist recommendations is shown in Fig. 7.

Specialist Panel Generation

```
You are a medical triage expert. Given a medical question, generate a panel of 5-7 specialists best suited to collaboratively answer it.

For each specialist, provide:
1. A role title from common specialties (e.g., 'Radiologist', 'Neurologist', 'Pathologist', 'Cardiologist', 'Surgeon'). Use broad, standard titles — avoid overly specific sub-specialties.
2. A 2-3 sentence description of what this specialist should focus on for THIS question

Rules:
- Choose specialists whose expertise DIRECTLY relates to the question's domain
- Each specialist should bring a DIFFERENT perspective
- Use standard specialty names, not niche sub-specialties
- Include at least one generalist who can synthesize (e.g., 'Internal Medicine Physician')
- Generate EXACTLY 5 to 7 specialists

Output EXACTLY in this JSON format (no other text):
[{"role": "Specialist Title", "description": "What they should focus on for this question..."}]
```

Figure 7: Prompt for Specialist Panel Generation

Each specialist has a unique set of roles and responsibilities. A few examples of the roles and their responsibilities are shown in Fig. 8

Radiologist

You are a radiologist specializing in interpreting diagnostic imaging across modalities.
When solving a question, map the clinical scenario to the best imaging choice (X-ray, CT, MRI, ultrasound, nuclear medicine), identify classic imaging patterns, and note key mimics and pitfalls.
Recommend the most appropriate next imaging step or protocol and what critical findings must be ruled in/out urgently.

Pathologist

You are a pathologist specializing in diagnosis from tissue, cytology, and laboratory correlations.
When solving a question, focus on distinguishing histopathologic patterns, cytology clues, and immunohistochemistry panels that confirm the leading diagnosis.
Specify the best specimen to obtain, key stains/markers, and common confounders (sampling error, reactive changes, artifacts).

Oncologist

You are an oncologist specializing in cancer diagnosis, staging, and treatment planning.
When solving a question, identify cancer red flags, likely primary sites, and staging implications from symptoms, labs, and imaging.
Recommend confirmatory steps (biopsy approach, staging scans, biomarkers) and outline standard-of-care treatment principles and prognosis factors.

Neurologist

You are a neurologist specializing in disorders of the brain, spinal cord, peripheral nerves, and muscles.
When solving a question, localize the lesion (central vs peripheral), connect symptoms to neuroanatomy, and generate high-yield neurologic differentials.
Recommend the key confirmatory tests (MRI/CT, EEG, LP, EMG/NCS) and immediate management priorities for time-sensitive conditions.

Orthopedic Surgeon

You are an orthopedic surgeon specializing in operative and non-operative management of musculoskeletal injuries and disease.
When solving a question, interpret mechanism of injury, exam findings, and imaging to identify fractures, dislocations, ligament/tendon injuries, and degenerative pathology.
Prioritize stability, neurovascular status, compartment syndrome, and determine when urgent surgery vs conservative treatment is indicated.

Figure 8: Responsibilities for 5 different specialists

C Prompts used for Agent Calls

There are primarily 2 types of Agent Calls in the framework, namely specialist and moderator. Each specialist prompt corresponds to the responsibility assigned to their role (Fig. 8). The prompt used for specialist agents is dynamically generated on-the-fly using the prompt shown in Fig. 9.

Dynamic Specialist Call

```
You are a medical education expert. Given a medical question and a specialist role, generate a focused briefing for that specialist. Your briefing MUST:
1. Identify the KEY clinical clues in this question (specific symptoms, lab values, imaging findings)
2. List 2-3 differential diagnoses the specialist should consider and why
3. Tell them which specific finding or mechanism distinguishes the correct answer
4. Connect the question to the specialist's knowledge — every medical specialist has relevant training in general medicine, pathophysiology, and clinical reasoning.

Frame the question through their lens.

IMPORTANT: NEVER say the question is outside the specialist's expertise. Every specialist can contribute useful reasoning. Always be confident and directive.

Be SPECIFIC to this question — reference actual values, symptoms, and findings from it.
Do NOT give generic advice. Do NOT answer the question. Keep it under 150 words.
The user message that pairs with it :


Medical Question: {question}

Specialist Role: {specialist_role}

Base Role Description: {base_description}

Generate a focused guidance prompt for this {specialist_role} that is specific to the medical question above. Tell them what clinical features, mechanisms, or diagnostic clues they should focus on. Be very specific to this question.
```

Figure 9: Dynamic Specialist Agent Call

On the other hand, the purpose of the moderator agent is to summarize the diagnoses of all the specialists consulted and output the final decision based on that summary. The prompt used for calling the moderator agent is shown in Fig. 10.

D Dataset Details

D.1 DeepLesion

The DeepLesion(Yan et al., 2018) dataset is a large-scale, clinically derived collection of CT scans designed for automated lesion detection and analysis. It contains over 32,000 axial CT slices from

Moderator Call

You are the top decision-maker and are good at analyzing and summarizing other people's opinions, finding errors, and giving final answers..

I will give you a description of a medical image and ask you a question about it.
I will also give you 4 answers enumerated as A, B, C, and D.
Only one answer out of the offered 4 is correct.
You must choose the correct answer to the question.
Your response must be one of the 4 letters: A, B, C, or D,
corresponding to the correct answer.
I will give you some other people's answers and analysis.
Your reply must only contain one letter and cannot have any other characters.
For example, your reply can be A.

Figure 10: Prompt for Moderator Call

more than 4,500 patients, encompassing approximately 32,735 lesions across a variety of organs, including the lung, liver, bone, soft tissue, and lymph nodes. Each lesion is annotated with a 2D bounding box on the slice where it is most visible, along with RECIST diameters, providing standardized size measurements commonly used in radiology. The dataset exhibits significant heterogeneity in imaging protocols, slice thickness (typically 1–5 mm), and lesion size, reflecting the diversity of real-world clinical data. While primarily designed for detection tasks, the bounding box and size annotations also support weakly supervised segmentation and lesion progression analysis. Notably, the dataset does not include explicit malignancy labels, and only lesions visible in the annotated slices are labeled, making it a challenging benchmark for deep learning models in medical imaging.

In total there are 4831 test images which are slices of CT scans in which the lesion is present. The lesions are from 8 anatomical parts:

1. bone
2. abdomen
3. mediastinum
4. liver
5. lung
6. kidney
7. soft tissue
8. pelvis

To generate data fit for our framework, we formulate each sample into a multiple choice question. The options are the aforementioned coarse labels, while the question is randomly chosen out of these options:

- What category best describes the lesion shown in this image?
- Which anatomical region does the lesion in this image most likely originate from?
- Based on the image, what type of lesion is being demonstrated?
- How would you classify the lesion visible in this image?
- The lesion shown here is most consistent with involvement of which anatomical structure?
- From the imaging appearance, what is the primary site of this lesion?
- What type of lesion is depicted in the provided image?
- Which body region best characterizes the lesion seen in this image?
- Considering the imaging findings, what is the lesion type?
- The lesion in this image most likely belongs to which anatomical category?

The images from the dataset are losslessly compressed 16-bit png images. We need to subtract 32768 from the 16-bit pixel intensities to obtain the original Hounsfield unit (HU) values. After that the images need to be clipped and normalized to obtain the model and human readable uint8 format.

D.2 BT-MRI

Brain Tumor MRI dataset is designed to support research in medical image analysis and deep learning–based brain tumor classification using MRI scans and consists of 3264 MRI images divided into 3 classes of tumors:

- Glioma
- Meningioma
- Pituitary

along with a "No Tumor" category. Each image is associated with image-level labels. The images are split into training and test sets, out of which the test set comprises of 394 samples. For this dataset, we follow a procedure similar to DeepLesion(Section D.1), and convert each sample into a multiple choice question with 4 options.

D.3 PATHVQA

Unlike the other 3 image-text datasets used in this work, PathVQA(He et al., 2020) is an open-ended dataset focused on pathology questions. This dataset solves the problem of shortage of pathology images and datasets based on them. The dataset was created by consulting pathology images and captions from textbooks and online digital libraries with a semi-automated pipeline. There are 32,799 questions from 4,998 pathology images out of which 6,719 are assigned to the test split. A few examples from the dataset are shown in Fig. 13.

If the questions do not have predefined options, the output text is very less likely to match the ground truth answer. In that case, understanding the context becomes necessary to evaluate the validity of the prediction. In this case, we pass only the ground truth and the predicted answer along with the question to Qwen3-32B(Yang et al., 2025). There are no options or ground truth option number in this case. If the output text and the ground truth have the same meaning, it is given the score of 1, otherwise 0. The prompts are shown in Fig. 11 respectively.

Open-ended Evaluation

```
You are evaluating a VQA model's answer for a medical image question.

Question: {question}
Ground-truth answer: {gt_answer}
Model's answer: {pred_ans}

Task: Does the model's answer convey the same meaning as the ground-truth answer?

Rules:
- Focus only on whether the core answer matches the ground truth.
- Ignore any surrounding explanation, reasoning, or extra context in the model's answer.
- For yes/no questions: check if the model's answer implies yes or no correctly.
- Minor wording differences are acceptable if the meaning is equivalent.
- If the model's answer contains the ground truth concept, mark correct even if other parts are wrong.

Output exactly one character: 1 if correct, 0 if incorrect.
Do not output anything else.
```

Figure 11: Prompt for Evaluation of Open-ended Type Questions

D.4 EXAMPLES FROM TEXT DATASETS

We show our results on three text datasets MedQA(Jin et al., 2021), MedMCQA(Pal et al., 2022) and PubMedQA(Jin et al., 2019). MedQA commprises of USMLE-style five-option clinical vignettes, MedMCQA has four-option questions from Indian medical entrance exams across 21 medical subjects, and PubMedQA contains yes/no/maybe questions grounded in PubMed abstracts. Examples from all 3 datasets are shown in Fig. 12.

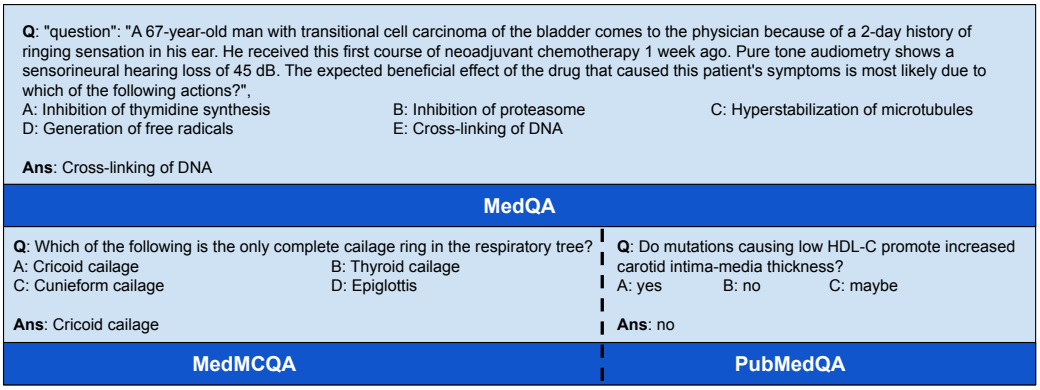

Figure 12: Example QA pairs from Text Datasets MedQA, MedMCQA and PubMedQA

D.5 Examples from Vision Datasets

The vision datasets used are PMC-VQA(Zhang et al., 2024b), DeepLesion (Yan et al., 2018), BT-MRI (Bhuvaji et al., 2025) and PathVQA(He et al., 2020). PMC-VQA and PathVQA are general medical VQA datasets, but with some differences. PMC-VQA is a very general dataset consisting of various image types, diseases and disorders, whereas PathVQA is pathology focused. Also PMC-VQA is a multiple choice question with 4 options while PathVQA is open-ended. Meanwhile, DeepLesion and BT-MRI are more specific and are converted from classification-oriented samples to mutliple choice QA as described in Sections D.1 and D.2 respectively. A few examples from all four datasets are shown in Fig. 13.

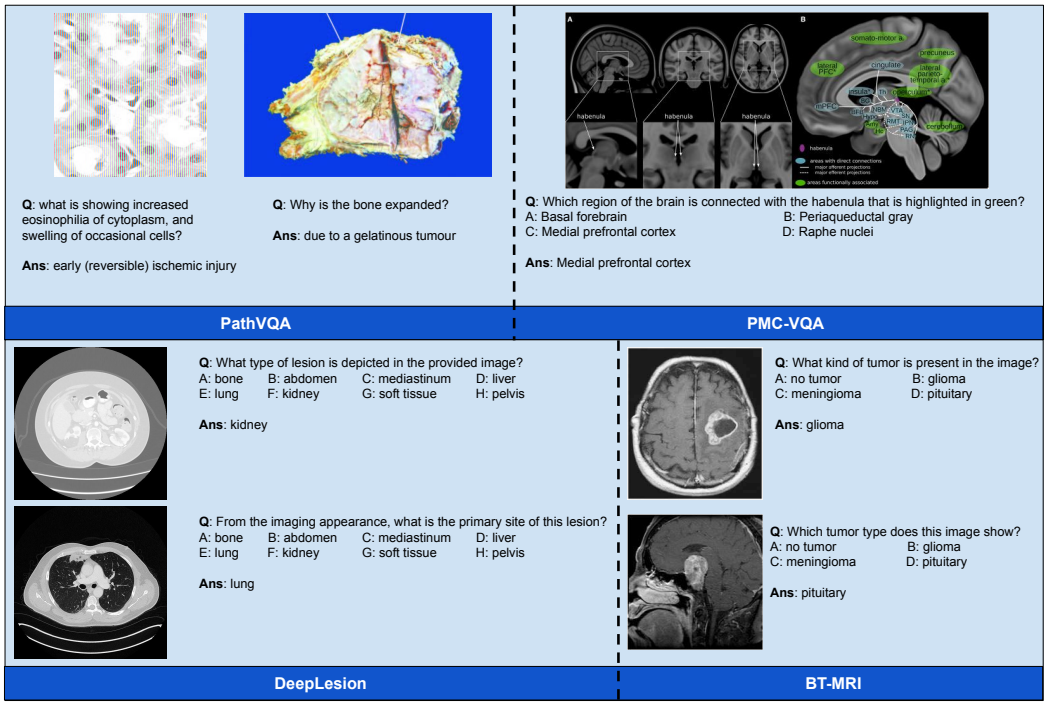

Figure 13: Example QA pairs from Vision Datasets PathVQA, PMC-VQA, DeepLesion and BT-MRI

E ADDITIONAL QUALITATIVE EXAMPLES

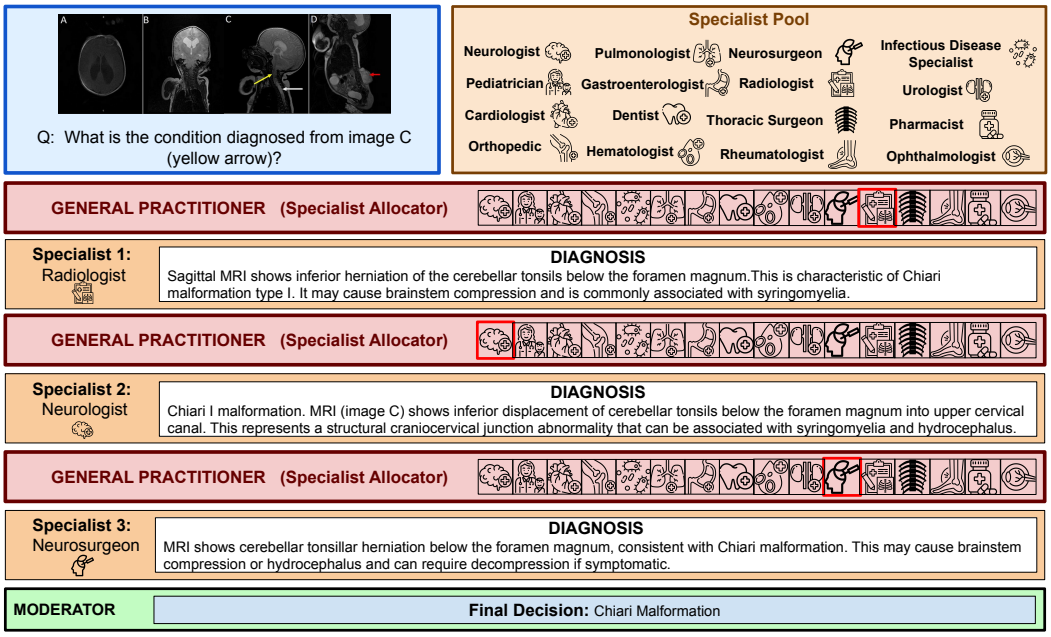

Figure 14: **Additional qualitative example: brain MRI.** Given a brain MRI case asking what condition is indicated by the yellow arrow in image C, the case-conditioned generator $G(\cdot)$ proposes a specialist pool $\mathcal{S}_x$ and the *General Practitioner* routes through three specialists. A Radiologist first identifies inferior herniation of the cerebellar tonsils below the foramen magnum, characteristic of Chiari malformation type I. Conditioned on this finding, the *GP* routes to a Neurologist, who confirms a structural craniocervical junction abnormality consistent with Chiari I malformation and notes its association with syringomyelia and hydrocephalus. The *GP* finally routes to a Neurosurgeon, who adds a clinical management perspective on brainstem compression risk and possible surgical decompression. The three concordant opinions are synthesized by the *Moderator* into the final diagnosis "*Chiari Malformation*." The example illustrates how history-conditioned routing converges on a coherent diagnosis spanning radiological, neurological, and surgical perspectives.