MedSynapse-V: Bridging Visual Perception and Clinical Intuition via Latent Memory Evolution

Chunzheng Zhu[1], Yijun Wang[1], Jiaqi Zeng[1], Junyu Jiang[1], and Jianxin Lin[1⋆]

Hunan University, Changsha, China
linjianxin@hnu.edu.cn

Abstract. High-precision medical diagnosis relies not only on static imaging features but also on the implicit diagnostic memory experts instantly invoke during image interpretation. We pinpoint a fundamental cognitive misalignment in medical VLMs caused by discrete tokenization, leading to quantization loss, long-range information dissipation, and missing case-adaptive expertise. To bridge this gap, we propose MedSynapse-V, a framework for latent diagnostic memory evolution that simulates the experiential invocation of clinicians by dynamically synthesizing implicit diagnostic memories within the model's hidden stream. Specifically, it begins with a ***Meta Query for Prior Memorization*** mechanism, where learnable probes retrieve structured priors from an anatomical prior encoder to generate condensed implicit memories. To ensure clinical fidelity, we introduce ***Causal Counterfactual Refinement (CCR)*** which leverages reinforcement learning and counterfactual rewards derived from region-level feature masking to quantify the causal contribution of each memory, thereby pruning redundancies and aligning latent representations with diagnostic logic. This evolutionary process culminates in ***Intrinsic Memory Transition (IMT)***, a privileged-autonomous dual-branch paradigm that internalizes teacher-branch diagnostic patterns into the student-branch via full-vocabulary divergence alignment. Comprehensive empirical evaluations across multiple datasets demonstrate that MedSynapse-V, by transferring external expertise into endogenous parameters, *significantly outperforms existing state-of-the-art methods, particularly Chain-of-Thought (CoT) paradigms*, in diagnostic accuracy and multi-dataset generalization *without compromising the inference efficiency* of standard VLMs.

Keywords: VLMs · Implicit Diagnostic Memory · Latent Space Memory · Causal Counterfactual · Memory Distillation

1 Introduction

Seasoned diagnostic experts do not rely on stepwise logical reasoning when making clinical diagnoses; instead, they activate Implicit Diagnostic Memory,

⋆ Corresponding author.

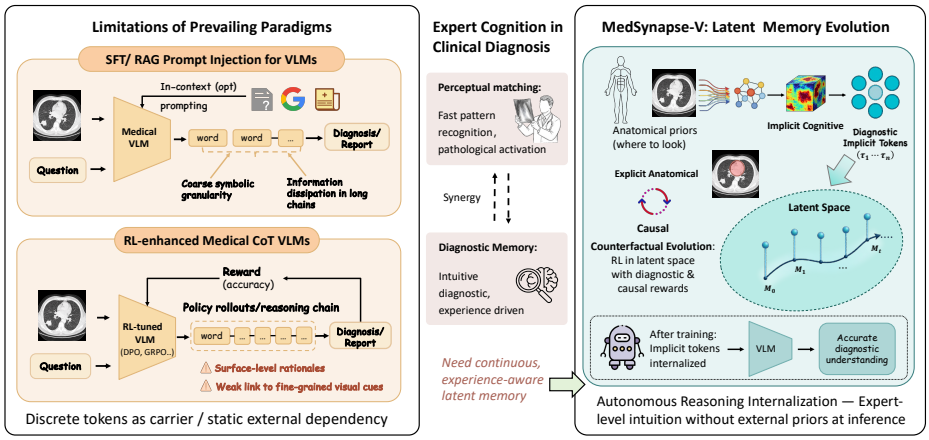

Fig. 1: Existing medical VLMs suffer from coarse symbolic granularity and long-range information dissipation in discrete reasoning. MedSynapse-V addresses this by evolving diagnostic implicit memory in latent space via anatomical prior condensation, causal counterfactual refinement, and autonomous latent memory internalization.

that enables near-instantaneous pattern recognition against accumulated case knowledge [3, 35, 48]. Although medical vision-language models (VLMs) have made substantial progress in diagnostic assistance [5, 23, 31, 41, 50], with reinforcement learning from verifiable rewards [19, 37, 44, 45] and chain-of-thought (CoT) [6,12,27,49,51,64] further advancing reasoning capabilities. However, their intrinsic reliance on discrete tokens engenders a profound *Cognitive Misalignment* with the inherently continuous nature of clinical expertise. As illustrated in Fig. 1, the limited granularity of a fixed vocabulary is inadequate for representing continuous pathological features such as gradual transitions in lesion density or textural heterogeneity, and the autoregressive decoding mechanism is prone to progressive attenuation of visual evidence over extended reasoning chains. Moreover, discrete symbols tend to encode generic linguistic priors rather than dynamic anatomical context, readily giving rise to "pseudo-logical" hallucinations that lack grounding in physical evidence.

An intuitive remedy is to supplement models with external diagnostic knowledge. Retrieval-augmented generation (RAG) prepends retrieved text fragments or similar cases to the input context [52,58,60,63], while soft-prompt and prefix-tuning methods concatenate learnable vectors to the input sequence to inject domain-specific cues [14,21,47]. However, both strategies inject information that remains *static* and *causally unverified*: it has undergone neither validation of causal relevance to the current diagnostic decision nor evolution into an intrinsic model capability through gradient-based optimization, persisting as a brittle external dependency prone to context saturation and information redundancy as the differential diagnosis space expands.

Recent latent computation paradigms [15, 24, 43, 53] offer a principled alternative by performing reasoning in continuous hidden state spaces, circumventing the expressiveness bottleneck of discrete symbols. However, their direct application to medical scenarios encounters two domain-specific obstacles. First, without structured anatomical priors, latent representations degenerate into abstract vectors decoupled from clinical semantics: they capture statistical regularities of the training distribution but fail to encode the structured spatial relationships (organ topology, lesion morphology, tissue boundaries) essential for diagnostic grounding. Second, without causal calibration, the coupling between latent representations and diagnostically critical visual features remains weak, as the model may produce correct answers by exploiting spurious correlations (*e.g.*, dataset-specific formatting cues) rather than attending to pathologically relevant regions, undermining reliability in clinical deployment.

These observations converge on a fundamental question: *Can a VLM progressively evolve its latent memory to simulate clinical intuition, enabling the rapid synthesis of case-adaptive diagnostic patterns, while ensuring this autonomous internal reasoning stream and its continuous refinement effectively steer the model toward clinically reliable decisions?*

This paper proposes MedSynapse-V, a framework for latent diagnostic memory evolution that addresses this question through three synergistic mechanisms operating in a progressive training paradigm. First, *Meta Query for Prior Memorization* (§3.2) deploys learnable meta-query probes to retrieve multi-scale spatially aware features from a frozen anatomical encoder pre-trained on large-scale segmentation tasks, condensing them into compact diagnostic implicit memory vectors that are injected into the VLM's hidden stream. This mechanism bridges the representation gap between the encoder's anatomical feature space and the VLM's generation space. Second, *Causal Counterfactual Refinement* (CCR; §3.3) performs reinforcement learning-driven memory optimization, introducing a novel causal counterfactual reward that quantifies the causal diagnostic contribution of each memory element through region-level feature masking interventions. By contrasting model behavior under original versus intervened memory conditions, CCR systematically prunes causally irrelevant components while reinforcing those with genuine diagnostic utility. Third, *Intrinsic Memory Transition* (IMT; §3.4) employs a privileged-autonomous dual-branch paradigm to distill the refined diagnostic patterns from a teacher branch (with the anatomical encoder) into a lightweight student branch via full-vocabulary Jensen-Shannon divergence alignment. At inference, the anatomical encoder is entirely removed, and the model generates diagnostic memory autonomously with computational overhead nearly identical to a standard VLM.

Comprehensive evaluations across seven medical multimodal benchmarks demonstrate that MedSynapse-V consistently outperforms a broad spectrum of state-of-the-art approaches, spanning medical-specific VLMs, RL-enhanced CoT paradigms, and general-purpose latent reasoning methods, in both diagnostic accuracy and cross-domain generalization, while introducing negligible additional inference cost compared with standard VLMs. Our main contributions are:

(1) We propose MedSynapse-V, the first framework that evolves diagnostic implicit memory in latent space for medical diagnosis, shifting from static external knowledge injection to progressive, autonomous memory internalization.
(2) We design Meta Query–based Prior Memorization coupled with Causal Counterfactual Refinement (CCR), which distills anatomical priors into compact latent memory and calibrates it via counterfactual interventions to retain only causally grounded diagnostic components.
(3) We introduce Intrinsic Memory Transition (IMT), a privileged–autonomous dual-branch distillation paradigm that internalizes encoder-dependent memory into autonomously generated intrinsic memory via full-vocabulary divergence alignment, eliminating all auxiliary modules at inference.
(4) In multimodal benchmarks, MedSynapse-V consistently surpasses mainstream CoT paradigms in diagnostic accuracy while maintaining inference efficiency on par with standard VLMs, validating latent memory evolution as a principled alternative to discrete token reasoning.

2 Related Works

Latent Computation. Our method is closely related to latent computation, a paradigm that leverages continuous latent states to intervene in or reshape the generation process of large language models [9, 13, 56, 59]. Existing works can be categorized into two paradigms: (I) achieving native latent reasoning at the architectural level, represented by Coconut [15], CODI [43], LatentR3 [62], and CoLaR [46], where the reasoning process is inherently completed in continuous space; (II) leveraging latent computation to regulate generation quality, represented by LaRS [54], LatentSeek [24], SoftCoT [53], and Coprocessor [30], which modulate output accuracy through latent space representations. The diagnostic implicit memory proposed in this paper can be viewed as an instantiation of the latter paradigm: condensing structured domain priors into continuous memory vectors to supply the diagnostic agent with critical incremental context. Unlike the aforementioned general purpose methods, MedSynapse-V further introduces causal counterfactual refinement and intrinsic memory transition mechanisms, *enabling implicit memory to undergo progressive evolution from external prior dependency to intrinsic model capability.*

Medical Alignment and Hallucination. The stringent safety requirements of medical diagnosis have driven model alignment strategies from generic preference imitation toward clinically constrained causal logic alignment [2,19,19,37,42,55]. Med-R1 and MedVLM-R1 [19,37], based on Group Relative Policy Optimization (GRPO) [42], enhance the robustness of reasoning through structured diagnostic rewards and logic backtracking mechanisms [36,39,40]. MMedPO [66] introduces contrastive optimization based on lesion region perturbation and clinical significance scoring, compelling the model to prioritize learning feature mappings with causal contributions to critical anatomical details during the alignment process [2, 55]. Several additional works [6, 22, 34, 49] explore explicitly anchoring the reasoning process in visual space to strengthen the utilization of spatial

evidence. The alignment signals of the aforementioned methods all operate on the model's *output layer* (the generated text sequence), whereas MedSynapse-V proposes a complementary perspective: CCR directly injects causal alignment signals into the model's *latent memory representations*, and IMT internalizes clinically aligned biases into the model's internal memory weights, enabling autonomous diagnostic capability at inference without external priors.

3 Methodology

3.1 Problem Formulation and Architecture Overview

Problem Formulation. Given an image $X \in \mathbb{R}^{H\times W\times 3}$ and a clinical query q, the objective is to generate an output sequence y containing diagnostic analysis and final conclusions. Let π_θ denote the VLM policy, $\mathcal{E}_{ana}$ the frozen pretrained anatomical encoder, and $\mathcal{P}_\phi$ the parameterized memory synthesis module. MedSynapse-V dynamically generates a set of diagnostic implicit memory $\mathcal{M} = \{m_1, \ldots, m_N\} \in \mathbb{R}^{N\times d_h}$ for injection into the VLM hidden stream, where d_h is the hidden state dimensionality. The overall policy is formalized as:

$$\pi_\theta(y \mid X, q) = \pi_\theta\big(y \;\big|\; X,\, q,\, \mathcal{P}_\phi\big(\mathcal{E}_{ana}(X)\big)\big)\,. \tag{1}$$

Unlike explicit CoT, which concatenates reasoning tokens to the input sequence, MedSynapse-V constructs a diagnostic experience invocation mechanism based on implicit memory within the representation space.

Architecture Overview. As illustrated in Figure 2, MedSynapse-V follows a three-stage progressive memory evolution training paradigm. *Stage I: Meta Query for Prior Memorization* (§3.2) extracts priors from the frozen anatomical encoder, condenses them into diagnostic implicit memory $\mathcal{M}$ through a learnable memory synthesis module and injects them into the VLM hidden stream, while simultaneously completing semantic alignment warmup. *Stage II: Causal Counterfactual Refinement* (CCR; §3.3) freezes the synthesis module and performs policy optimization within the $\mathcal{M}$ conditioned latent space based on GRPO, incorporating causal counterfactual rewards for memory refinement. *Stage III: Intrinsic Memory Transition* (IMT; §3.4) writes the refined diagnostic memory into the model's autonomous pathway through privileged autonomous dual branch full vocabulary divergence distillation, completely removing the anatomical encoder at inference. The entire process is represented as a progressive evolution chain of diagnostic memory: $\mathbf{F}_{ana} \xrightarrow{\textbf{Meta Query}} \mathcal{M} \xrightarrow{\textbf{CCR}} \mathcal{M}^\star \xrightarrow{\textbf{IMT}} \mathcal{M}_{auto}$, where $\mathbf{F}_{ana}$ denotes the anatomical encoder output features, $\mathcal{M}$ is the initial memory after semantic alignment, $\mathcal{M}^\star$ is the causally refined memory, and $\mathcal{M}_{auto}$ is the intrinsic memory generated by the autonomous module.

3.2 Meta Query for Prior Memorization

Clinical cognition research has shown that experienced physicians rapidly activate long accumulated anatomical knowledge during image interpretation, forming compressed contextualized expert memory to guide diagnosis [48]. Inspired

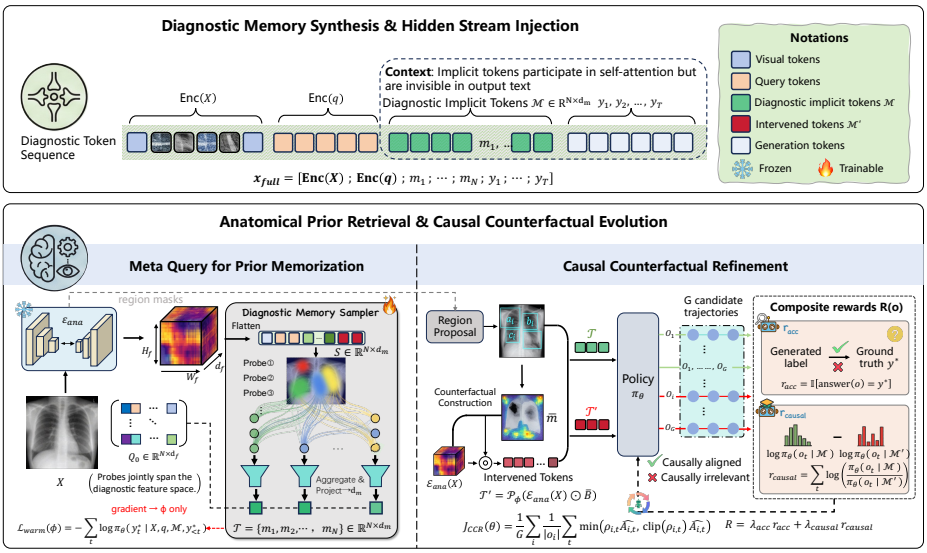

Fig. 2: Stages I and II of MedSynapse-V. The hook features from an encoder are condensed into diagnostic implicit memory via learnable meta-query probes and injected into the VLM hidden stream. The memory is then refined through RL with composite rewards, ensuring causal alignment between memory and clinical decision logic.

by this finding, this stage models this process as a structured prior retrieval and memorization mechanism: the anatomical encoder extracts spatially aware features, which are then condensed into diagnostic implicit memory through a learnable synthesis module and injected into the VLM hidden states.

Structured Prior Elicitation. Given an input image X, the frozen anatomical encoder $\mathcal{E}_{ana}$ outputs spatial features from the final layer: $\mathbf{F} = \mathcal{E}_{ana}(X) \in \mathbb{R}^{H_f \times W_f \times d_f}$, where $H_f \times W_f$ is the spatial resolution and d_f is the feature dimensionality. This feature map encodes structured spatial priors learned by the anatomical encoder from large scale medical image segmentation tasks, encompassing multi-granularity information such as lesion boundaries, organ topology, and tissue textures. It is flattened into a sequence $\mathbf{S} = \texttt{flat}(\mathbf{F}) \in \mathbb{R}^{M \times d_f}$, $M = H_f \times W_f$, forming the feature pool for candidate memory.

Diagnostic Memory Synthesis. To condense the high-dimensional feature pool into compact memory compatible with the VLM hidden state space, we design a lightweight *Diagnostic Memory Sampler* $\mathcal{P}_\phi$ that maintains N learnable *meta-query probes* $\mathbf{Q}_0 \in \mathbb{R}^{N \times d_f}$, each of which learns to attend to specific pathological semantic patterns (*e.g.*, boundary irregularity, density heterogeneity, or vascular-tissue spatial relationships). Using $\mathbf{Q}_0$ as queries and $\mathbf{S}$ as key-value pairs, $\mathcal{P}_\phi$ performs selective aggregation and dimensional alignment: $\mathcal{M} = \mathcal{P}_\phi(\mathbf{Q}_0, \mathbf{S}) \in \mathbb{R}^{N \times d_h}$, where d_h is the VLM hidden state dimensionality. The resulting N diagnostic implicit memory elements $\mathcal{M} = \{m_1, \ldots, m_N\}$ collectively span the feature subspace required for diagnosis. $\mathcal{P}_\phi$ extracts the

most diagnostically relevant compact representations from the encoder's high dimensional feature pool according to the task context, bridging them from the anatomical encoder's representation space to the VLM's hidden state space.

Diagnostic Memory Injection. The diagnostic implicit memory $\mathcal{M}$ is injected into the VLM generation sequence as continuous vectors, positioned after the question encoding and before the answer generation: $\mathbf{x}_{full} = \big[\texttt{Enc}(X);\ \texttt{Enc}(q);\ m_1;\ \ldots;\ m_N;\ y_1;\ \ldots;\ y_T\big]$, where $\texttt{Enc}(\cdot)$ denotes the encoding operation and $\{y_t\}_{t=1}^{T}$ are the answer tokens to be generated. Since $\mathcal{M}$ shares the d_h dimensional space with VLM hidden states, the self-attention mechanism enables the generation sequence to dynamically aggregate diagnostic evidence from $\mathcal{M}$, allowing the VLM to modulate its predictive distribution based on latent priors without additional adaptation layers or architectural modifications.

Semantic Alignment Warmup. Directly injecting outputs from a randomly initialized synthesis module into the VLM would cause semantic mismatch. To address this, a warmup stage is established before formal refinement: the VLM backbone θ and the anatomical encoder $\mathcal{E}_{ana}$ are frozen, and only ϕ is optimized with the standard next token prediction loss:

$$\mathcal{L}_{warm}(\phi) = -\sum_{t=1}^{T} \log \pi_\theta(y_t^\star \mid X, q, \mathcal{M}, y_{<t}^\star), \tag{2}$$

where $y^\star$ denotes the reference answer sequence. This stage establishes a well-conditioned initial semantic mapping between the anatomical encoder feature space and the VLM hidden state space, providing a semantically coherent and stable initialization upon which the subsequent RL stage can reliably build.

3.3 Causal Counterfactual Refinement

After warmup, $\mathcal{M}$ possesses basic semantic alignment capability, but directly applying standard SFT has two limitations [55]: binding to fixed reference trajectories restricts out-of-distribution generalization; the model may bypass $\mathcal{M}$ and directly map from (X, q) to answers, causing memory to degenerate into redundant placeholders. In this stage, we perform policy optimization within the $\mathcal{M}$-conditioned latent space, incorporating causal counterfactual intervention to guide memory refinement toward clinical alignment.

Conditioned Policy Modeling. First, we freeze $\mathcal{P}_\phi$ and the VLM backbone, optimizing only lightweight LoRA adapters (collectively denoted θ below). Based on GRPO [42], for each sample $(X, q, y^\star)$, G candidate trajectories $\{\mathbf{o}_1, \ldots, \mathbf{o}_G\}$ are sampled. $\mathcal{M}$ is explicitly incorporated into the conditioned context, and the policy gradient indirectly modulates the VLM's utilization pattern of $\mathcal{M}$ through the attention pathway. The policy model optimization objective is:

$$\mathcal{J}_{CCR}(\theta) = \frac{1}{G}\sum_{i=1}^{G} \frac{1}{|\mathbf{o}_i|} \sum_{t=1}^{|\mathbf{o}_i|} \min\Big(\rho_{i,t}\,\hat{A}_i,\ \texttt{clip}\big(\rho_{i,t},\ 1{-}\varepsilon,\ 1{+}\varepsilon\big)\hat{A}_i\Big), \tag{3}$$

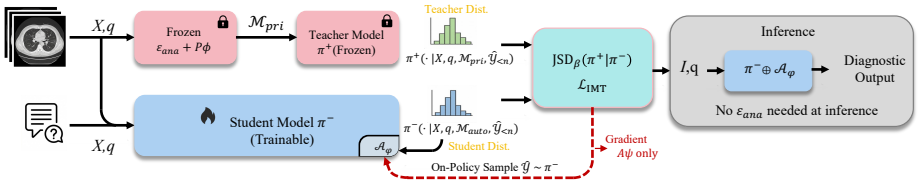

Fig. 3: Intrinsic Memory Transition (IMT) is achieved via Jensen–Shannon divergence alignment between the teacher (π^+, conditioned on encoder-derived $\mathcal{M}_{pri}$) and student (π^-, conditioned on $\mathcal{M}_{auto}$) branches. Gradients propagate solely to $\mathcal{A}_\psi$, enabling complete removal of the anatomical encoder at inference with negligible overhead.

where $\rho_{i,t} = \frac{\pi_\theta(\mathbf{o}_{i,t}|X,q,\mathcal{M},\mathbf{o}_{i,<t})}{\pi_{\theta_{old}}(\mathbf{o}_{i,t}|X,q,\mathcal{M},\mathbf{o}_{i,<t})}$ and $\hat{A}_i = \frac{R(\mathbf{o}_i)-\mu_G}{\sigma_G+\varepsilon_0}$, with μ_G and σ_G denoting the group reward mean and standard deviation, and ε_0 a stability constant. Note that both $\rho_{i,t}$ and $\hat{A}_i$ are conditioned on $\mathcal{M}$, allowing gradients to propagate through the attention pathway linking answer tokens with memory, thereby shaping how the VLM attends to the injected diagnostic cues during generation.

Interventional Reward Design. The composite reward $R(\mathbf{o}) = \lambda_{acc} \cdot r_{acc} + \lambda_{causal} \cdot r_{causal}$ consists of two components. The diagnostic accuracy reward is:

$$r_{acc}(\mathbf{o}) = \mathbb{I}\big[\texttt{answer}(\mathbf{o}) = y^\star\big]. \tag{4}$$

The causal counterfactual reward quantifies the causal contribution of $\mathcal{M}$ through intervention. The frozen anatomical encoder $\mathcal{E}_{ana}$ additionally provides highest-confidence region masks for diagnostically relevant areas; zeroing out the corresponding features yields the interventional memory $\mathcal{M}'$: $\mathcal{M}' = \mathcal{P}_\phi\big(\mathcal{E}_{ana}(X) \odot \overline{\mathbf{B}}\big)$, where $\overline{\mathbf{B}}$ is the inverted binary mask. The causal reward measures the performance discrepancy between the original and interventional conditions:

$$r_{causal}(\mathbf{o}) = \sum_{t=1}^{|\mathbf{o}|} \log \frac{\pi_\theta(\mathbf{o}_t \mid X, q, \mathcal{M}, \mathbf{o}_{<t})}{\pi_\theta(\mathbf{o}_t \mid X, q, \mathcal{M}', \mathbf{o}_{<t})}. \tag{5}$$

$r_{causal} > 0$ indicates that the original memory encodes information with causal contribution to diagnosis; $r_{causal} \leq 0$ indicates that the corresponding prior is causally irrelevant to the current decision. This design follows the principle of interventional effect estimation in causal inference, ensuring that the memory retained after refinement is causally consistent with clinical decision logic.

3.4 Intrinsic Memory Transition

After the causal RL, the model can generate high quality diagnostic outputs under the guidance of externally derived $\mathcal{M}$, but the persistent dependence on the encoder at inference introduces additional computational overhead. As shown in Fig. 3, IMT reformulates this problem as learning to autonomously generate equivalent latent space embeddings under the condition of removing the

auxiliary encoder, employing a privileged autonomous dual-branch paradigm to accomplish memory transition from extrinsic to intrinsic.

Privileged Branch and Autonomous Branch. The teacher branch (privileged) retains the complete pipeline with the anatomical encoder, generating reference memory $\mathcal{M}_{pri} = \mathcal{P}_\phi(\mathcal{E}_{ana}(X)) \in \mathbb{R}^{N\times d_h}$. The student branch (autonomous) removes the encoder and predicts equivalent memory solely through a lightweight Autonomous Memory Module $\mathcal{A}_\psi$ mounted on the VLM, using the VLM's own visual encoding features:

$$\mathcal{M}_{auto} = \mathcal{A}_\psi\big(\texttt{Enc}_{VLM}(X,q)\big) \in \mathbb{R}^{N\times d_h}. \tag{6}$$

Both branches share the VLM backbone parameters θ, ensuring that alignment is completed within the same function space.

Training Objective: Unified Full Vocabulary Divergence. The core signal of IMT is the full vocabulary divergence objective covering all generation positions. Given a trajectory $\hat{y} \sim \pi^-(\cdot \mid X, q, \mathcal{M}_{auto})$ sampled from the student branch, the trajectory averaged per position divergence is defined as:

$$\mathcal{L}_{IMT}(\psi) = \mathbb{E}_{(X,q,y^\star)}\left[\mathbb{E}_{\hat{y}\sim\pi^-}\left[\frac{1}{|\hat{y}|}\sum_{n=1}^{|\hat{y}|} D\big(\pi^+(\cdot \mid \hat{y}_{<n}) \,\big\|\, \pi^-(\cdot \mid \hat{y}_{<n})\big)\right]\right], \tag{7}$$

where $\pi^+(\cdot \mid \hat{y}_{<n}) \triangleq \pi_\theta(\cdot \mid X, q, \mathcal{M}_{pri}, \hat{y}_{<n})$ and $\pi^-(\cdot \mid \hat{y}_{<n}) \triangleq \pi_\theta(\cdot \mid X, q, \mathcal{M}_{auto}, \hat{y}_{<n})$ are the full vocabulary next-token distributions conditioned on privileged and autonomous memory, respectively. The divergence D adopts the generalized Jensen–Shannon divergence JSD_β [26]:

$$\mathrm{JSD}_\beta\big(\pi^+\|\pi^-\big) = \beta\, D_{KL}\big(\pi^+\|\bar{m}\big) + (1-\beta)\, D_{KL}\big(\pi^-\|\bar{m}\big), \tag{8}$$

where $\bar{m} = \beta\,\pi^+ + (1-\beta)\,\pi^-$. The teacher branch leverages privileged memory to expose the complete next token probability landscape to the student, driving $\mathcal{A}_\psi$ to learn to generate latent space embeddings behaviorally equivalent to privileged memory. Gradients backpropagate only through the student branch to $\mathcal{A}_\psi$, while the teacher branch serves as a fixed distributional target.

At inference, $\mathcal{A}_\psi$ directly generates $\mathcal{M}_{auto}$ from VLM visual encoding features and injects them into the hidden stream. The entire Meta Query pipeline and the anatomical encoder are completely removed, rendering the computational overhead of MedSynapse-V nearly identical to that of a standard VLM.

4 Experiments

4.1 Experimental Setup

Datasets **Training data:** We conduct comprehensive experiments on seven medical multimodal benchmarks spanning diverse task types and difficulty levels. **Stage I** (MQPM warmup) uses 50K image–text pairs from PubMedVision [5]

covering radiology and pathology. **Stage II** (CCR) constructs a mixed-modality RL set: 3K closed-ended VQA samples from OmniMedVQA [18] training split (8 modalities: CT, MRI, X-ray, dermoscopy, fundus, OCT, pathology, ultrasound) plus 1K open-ended samples from SLAKE [29] and PathVQA [16] training sets, totaling ∼4K samples. Region masks for r_{causal} are provided by MedSAM3 [28]. **Stage III** (IMT) reuses the Stage II data. **Evaluation benchmarks: (i)** Closed-ended VQA: VQA-RAD [20], SLAKE [29], PathVQA [16], PMC-VQA [61]; **(ii)** Clinical reasoning: MMMU Health & Medicine [57] (denoted MMMU*); **(iii)** Expert-level reasoning: MedXpertQA-MM [67] (Total score); **(iv)** Multi-granularity: GMAI-MMBench [7].

Baselines We compare against four categories of methods: **(1)** *General VLMs*: Qwen3-VL-8B [1] (our base model), InternVL3-8B [65]; **(2)** *Medical-specific VLMs*: RadFM [50], LLaVA-Med [23], GMAI-VL [25], HuatuoGPT-Vision [5], BiMediX2-8B [33], MedMO-8B [10]; **(3)** *RL-enhanced medical reasoning*: MedVLM-R1-2B [37], Med-R1-3B [19], MediX-R1-8B [32], MMedExpert-R1-7B [11]; **(4)** *Latent-space reasoning*: Coconut† [15], MCOUT-Multi† [38], IVT-LR† [4] (†: adapted with identical Qwen3-VL-8B backbone and training data). We additionally report MedSynapse-V-4B on the Qwen3-VL-4B backbone to assess scalability.

Implementation Details Our framework builds upon Qwen3-VL-8B-Instruct [1]. The frozen anatomical encoder $\mathcal{E}_{ana}$ employs MedSAM3 [28] pre-trained on large-scale multi-organ segmentation datasets. Stage I freezes both VLM and $\mathcal{E}_{ana}$, training only the Diagnostic Memory Sampler $\mathcal{P}_{\phi}$ with lr $= 2 \times 10^{-4}$ for 3 epochs. The diagnostic probe count is N=16; $\mathcal{P}_{\phi}$ is a 2-layer cross-attention Transformer with output dimension d_m=4096 (matching Qwen3-VL-8B). Images are processed at native dynamic resolution following Qwen3-VL's default configuration. Stage II freezes $\mathcal{P}_{\phi}$ and adapts VLM via LoRA [17] (rank=64, applied to all attention layers). GRPO generates G=4 candidate trajectories per sample, with clipping coefficient ε=0.2, reward weights λ_{acc}=1.0 and λ_{causal}=0.5, training for 200 steps with a rollout batch size of 32. Max generation length is 1024 tokens. Stage III introduces the Autonomous Memory Module $\mathcal{A}_{\psi}$ (2-layer MLP + LayerNorm, input from VLM's visual encoder features), with JSD coefficient β=0.5, lr $= 1 \times 10^{-4}$, 3 epochs. For each sample we draw one on-policy trajectory $\hat{y} \sim \pi^{-}$ per gradient step; the Stage II data is reused with identical preprocessing. Closed-ended VQA tasks report overall accuracy (%). For GMAI-MMBench and MedXpertQA-MM, we follow their respective official evaluation protocols. Inference efficiency is measured quantitatively as ms/sample and peak GPU memory (GB). More details are provided in the supplementary material.

4.2 Main Results

As shown in Table 1, MedSynapse-V (w/ $\mathcal{E}_{ana}$) achieves the highest average of **61.4**%, and the encoder-free MedSynapse-V (IMT) retains **59.6**%, surpassing all baselines. Compared to the strongest RL baseline MMedExpert-R1 (55.7%),

Table 1: Comprehensive comparison on seven medical benchmarks. Base model: Qwen3-VL-8B (unless otherwise noted). MMMU* = Health & Medicine track. †: general latent-space methods adapted to medical VQA with identical backbone and training data. Δ: absolute gap (pp) to MedSynapse-V (w/ $\mathcal{E}_{ana}$) average. All results are averaged over five independent runs. **Bold**: best; underline: second best.

Method	Size	VQA-RAD	SLAKE	PathVQA	PMC-VQA	MMMU*	MedXpert	GMAI	Average	Δ (pp)
General Vision-Language Models										
Qwen3-VL-8B [1]	8B	58.6	66.2	55.4	42.5	48.3	22.1	47.2	48.6	-12.8
InternVL3-8B [65]	8B	57.3	64.8	50.6	41.2	50.1	21.5	49.6	47.9	-13.5
Medical-Specific VLMs										
RadFM [50]	14B	50.6	34.6	38.7	25.9	27.0	17.3	28.5	31.8	-29.6
LLaVA-Med [23]	7B	51.4	48.6	56.8	30.1	36.9	19.7	31.2	39.2	-22.2
GMAI-VL [25]	7B	64.6	71.9	47.2	52.3	51.2	23.8	45.2	50.9	-10.5
HuatuoGPT-V [5]	7B	63.8	74.5	59.9	53.4	49.1	22.7	51.3	53.5	-7.9
BiMediX2 [33]	8B	62.4	68.3	52.7	42.8	48.6	22.2	34.6	47.4	-14.0
MedMO-8B [10]	8B	59.3	66.8	48.5	36.0	46.2	20.8	38.2	45.1	-16.3
RL-Enhanced Medical CoT / Latent Reasoning										
MedVLM-R1 [37]	2B	58.6	63.2	42.5	35.8	31.2	16.8	33.5	40.2	-21.2
Med-R1 [19]	3B	53.2	52.8	44.1	38.5	30.4	18.5	35.2	38.9	-22.5
MediX-R1 [32]	8B	56.4	65.8	44.2	56.2	53.5	24.9	48.2	49.9	-11.5
MMedExpert-R1 [11]	7B	65.2	72.8	58.1	56.8	57.3	27.5	52.1	55.7	-5.7
Coconut† [15]	8B	55.8	63.4	50.1	41.6	43.2	19.8	37.6	44.5	-16.9
MCOUT-Multi† [38]	8B	59.2	67.4	53.8	44.8	47.5	22.0	40.9	47.9	-13.5
IVT-LR† [4]	8B	62.3	70.1	56.2	47.8	50.4	23.5	43.1	50.5	-10.9
MedSynapse-V-4B (w/ $\mathcal{E}_{ana}$)	4B	67.8	75.2	60.5	53.1	54.5	24.2	47.3	54.7	-6.7
MedSynapse-V-4B (IMT)	4B	66.0	73.1	58.6	51.2	52.4	22.6	45.0	52.7	-8.7
MedSynapse-V (w/ $\mathcal{E}_{ana}$)	8B	**75.6**	**81.4**	**66.2**	**59.8**	**62.7**	**29.4**	**54.8**	**61.4**	–
MedSynapse-V (IMT)	8B	74.2	79.8	64.8	58.5	61.4	26.8	51.6	59.6	-1.8

MedSynapse-V (IMT) leads by +**3.9** pp without any auxiliary module at inference, with *the largest margins on visual-grounding benchmarks* (VQA-RAD +**9.0**, SLAKE +**7.0**, PathVQA +**6.7**), where discrete CoT tokens are prone to attenuating early visual evidence across long reasoning chains. On GMAI-MMBench spanning 38 modalities, MedSynapse-V scores 54.8%, confirming that the anatomical priors generalize beyond the training distribution.

RL baselines reveal a specialization dilemma. MediX-R1 benefits from multilingual pretraining and leads on PMC-VQA (56.2%), yet this breadth dilutes radiology-specific precision (VQA-RAD: 56.4%); MMedExpert-R1 achieves the most balanced profile by leveraging guideline-based reward. Small-scale models (MedVLM-R1 2B, Med-R1 3B) collapse on out-of-domain tasks (MedXpert below 19%), confirming that parameter capacity sets a hard ceiling RL alone cannot raise. n contrast, MedSynapse-V sidesteps this dilemma by injecting latent priors that benefit all task types uniformly, achieving the top performance on every benchmark without task-specific tuning.

Latent methods require domain priors. Among adapted latent baselines, the hierarchy Coconut (44.5%) < MCOUT-Multi (47.9%) < IVT-LR (50.5%) tracks optimization sophistication, yet even IVT-LR barely exceeds zero-shot Qwen3-VL-8B (48.6%). This inversion reveals that *latent compression without clinical grounding encodes statistical shortcuts* rather than diagnostic logic; the 10.9 pp gap to MedSynapse-V confirms that prior injection and causal calibration are prerequisites for effective latent reasoning in medicine.

Scaling efficiency. MedSynapse-V-4B (w/ $\mathcal{E}_{ana}$) reaches 54.7% with roughly half the parameters of 7B baselines; after encoder removal the IMT variant

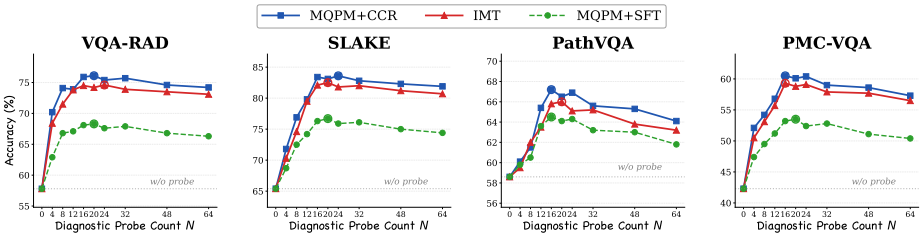

Fig. 4: Effect of diagnostic probe count N. Performance peaks around N=16 across benchmarks; further increasing N dilutes diagnostically relevant signals.

Table 2: Ablation study on the 8B backbone. All variants undergo the full three-stage pipeline including IMT (§3.4); results reflect inference *without* the anatomical encoder unless stated otherwise. MQPM: Meta Query for Prior Memorization (§3.2); CCR: Causal Counterfactual Refinement (§3.3); $\mathcal{M}$: diagnostic implicit memory. Best per group in **bold**.

Configuration	VQA-RAD	SLAKE	PathVQA	PMC-VQA	MMMU*	ms/token↓	Mem (GB)↓	Avg.
Qwen3-VL-8B (zero-shot, no $\mathcal{M}$)	58.6	66.2	55.4	42.5	48.3	126	16.2	54.2
(a) Progressive Training Stages								
MQPM → IMT (Stage I only, no RL)	59.5	67.4	57.1	45.6	47.2	104	16.5	55.4
MQPM → SFT → IMT (no RL)	63.2	71.2	60.4	49.8	51.3	104	16.5	59.2
Direct RL → IMT (skip MQPM)	57.4	64.3	54.8	43.1	44.7	105	16.5	52.9
MQPM → CCR → IMT (full)	**74.2**	**79.8**	**64.8**	**58.5**	**61.4**	102	16.5	**67.7**
(b) Reward Design								
r_{acc} only	70.1	75.8	61.2	54.3	56.6	102	16.5	63.6
$r_{acc} + r_{causal}$ (full)	**74.2**	**79.8**	**64.8**	**58.5**	**61.4**	102	16.5	**67.7**
(c) Encoder Retention vs. Removal								
w/ $\mathcal{E}_{ana}$ (no privileged branch)	75.6	81.4	66.2	59.8	62.7	168	22.8	69.1
w/o $\mathcal{E}_{ana}$ (IMT, default)	74.2	79.8	64.8	58.5	61.4	**102**	**16.5**	67.7
(d) Encoder Choice								
SAM-Med2D [8]	69.0	75.4	60.8	54.1	57.0	102	16.5	63.3
MedSAM3 [28] (default)	**74.2**	**79.8**	**64.8**	**58.5**	**61.4**	102	16.5	**67.7**
Random init. → IMT	56.4	64.0	55.1	41.2	43.3	102	16.5	52.0

still achieves 52.7% at 85 ms and 10.8 GB, surpassing MediX-R1-8B (49.9%). This efficiency stems from a structural advantage: *diagnostic expertise is distilled into 16 compact memory vectors consumed in a single forward pass*, rather than spread across 150+ verbose reasoning tokens.

4.3 Ablation Study

Table 2 reports comprehensive ablation study results. Specifically, ***(i)* Progressive training stages. *MQPM warmup is indispensable***: skipping it collapses Avg to 52.9, barely above zero-shot (54.2), because randomly initialized memory destabilizes early RL training. Replacing CCR with SFT reaches 59.2 but lags by 8.5 pp due to limited out-of-distribution generalization. The full pipeline (Avg **67.7**) confirms non-redundant contributions: MQPM grounds semantics, CCR refines via exploration, IMT compresses into an autonomous pathway. ***(ii)* Reward design. *r_{causal} is the dominant reward component*** (+**4.1** pp, 63.6 → 67.7). Without causal pressure the model bypasses $\mathcal{M}$ via di-

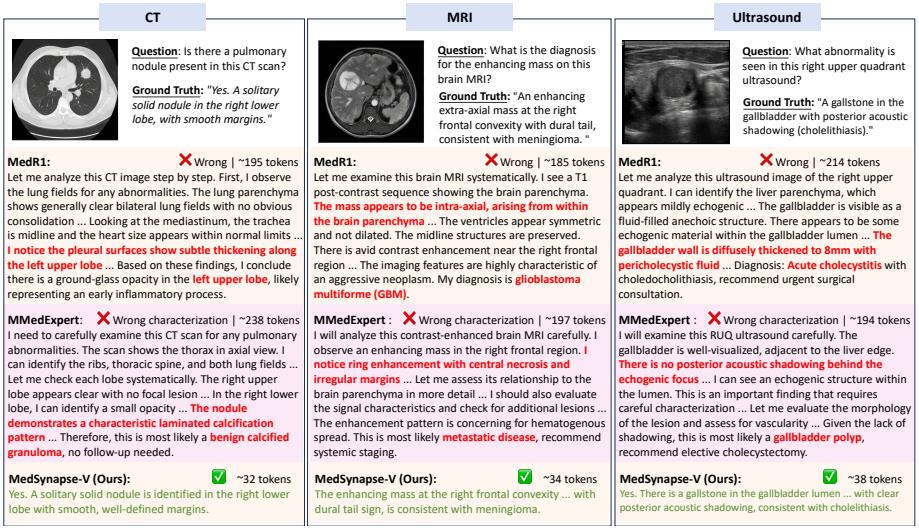

Fig. 5: Qualitative comparison across CT, MRI, and Ultrasound cases. MedSynapse-V produces concise, correct diagnoses, while Med-R1 and MMedExpert-R1 generate verbose CoT with hallucinated findings (red) leading to misdiagnoses.

rect shortcuts, treating memory as inert padding; the counterfactual intervention penalizes trajectories insensitive to diagnostic regions. The effect concentrates on radiology benchmarks and persists after IMT, indicating stronger memory utilization transfers more faithfully through distillation. ***(iii)* Encoder retention vs. removal.** ***IMT achieves near-lossless removal***: only 1.4 pp degradation (69.1 → 67.7) while latency drops 39% and memory decreases 6.3 GB. The gap is not uniform: core VQA metrics degrade minimally, whereas MedXpert and GMAI suffer more, suggesting ***complex reasoning depends more on encoder-derived priors*** than closed-ended recognition. ***(iv)* Anatomical encoder choice.** MedSAM3 outperforms SAM-Med2D by 4.4 pp (67.7 vs. 63.3), reflecting richer spatial representations from multi-organ segmentation pretraining. Random initialization yields only 52.0, confirming that gains originate from *what* the encoder knows, rather than *how* memory is aggregated. ***(v)* Probe count N.** As shown in Fig. 4, N=16 balances expressiveness against redundancy. The CCR to SFT gap widens with N (3.5 pp at N=4 vs. 7.2 pp at N=16), revealing that ***larger memory pools amplify bypass shortcuts*** and therefore benefit disproportionately from causal refinement.

4.4 In-Depth Case Analysis

As illustrated in Fig. 5, we compare MedSynapse-V with Med-R1 [19] and MMedExpert-R1 [11] across three distinct imaging modalities. Both baselines produce verbose CoT reasoning (∼185–238 tokens) yet arrive at incorrect diagnoses due to hallucinated observations erroneously propagating through the

chain. In the CT case, Med-R1 fabricates pleural thickening in the left upper lobe, while MMedExpert-R1 ***hallucinates a laminated calcification pattern*** and mischaracterizes the nodule as a benign granuloma. In the MRI case, Med-R1 misidentifies the extra-axial mass as intra-axial and concludes glioblastoma, whereas MMedExpert-R1 fabricates ring enhancement with central necrosis, both missing the classic meningioma presentation. In the ultrasound case, Med-R1 hallucinates gallbladder wall thickening to over-diagnose acute cholecystitis, while MMedExpert-R1 denies posterior acoustic shadowing and misdiagnoses a gallbladder polyp. In contrast, MedSynapse-V generates concise, correct answers (∼32–38 tokens) without explicit CoT, demonstrating that ***diagnostic implicit memory provides sufficient latent guidance while avoiding the hallucination cascades*** inherent in token-level CoT.

4.5 Efficiency, RL Dynamics, and Latent Space

Performance–efficiency trade-off. As shown in Fig. 6, MedSynapse-V (IMT) achieves 59.6% at 2.6 s/sample, comparable to zero-shot Qwen3-VL-8B (48.6%, 2.8 s) since both share the same backbone and the 16 memory vectors add negligible overhead. Full-scale 7–8B CoT methods (MediX-R1, MMedExpert-R1) require 5.8 s each due to 300–400 autoregressive reasoning tokens, while smaller CoT models (MedVLM-R1 2B, Med-R1 3B) offset verbosity with faster per-token speed yet remain 18–21 pp below MedSynapse-V. This confirms that ***compact latent memory provides diagnostic grounding without the token-generation overhead of full-scale CoT***.

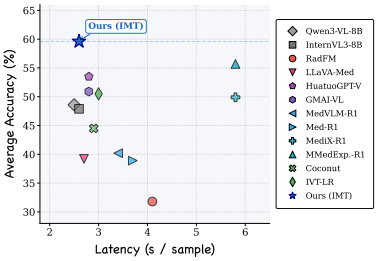

Fig. 6: Accuracy–latency trade-off across compared VLM categories.

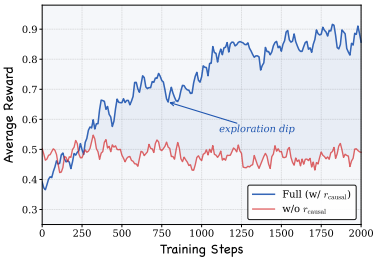

Fig. 7: The RL training reward dynamics with and without r_{causal}.

Training dynamics. Fig. 7 shows the full model ($w/\ r_{causal}$) improving steadily to ∼0.88 with a transient exploration dip near step 900, where the policy sacrifices reward to explore memory-reliant generation strategies, while the $w/o\ r_{causal}$ ablation plateaus at ∼0.48 throughout the training. This confirms that accuracy-only reward cannot distinguish memory-dependent from shortcut trajectories; without causal pressure the model bypasses $\mathcal{M}$ entirely, treating injected memory as inert padding.

Latent space structure. Fig. 8 visualizes the evolved memory $\mathcal{M}_{auto}$ via t-SNE across three granularities. At the cross-modality level (a), eight imaging types form compact clusters with clinically coherent proximity (*e.g.*, CT and MRI lie adjacent; dermoscopy and fundus form a nearby pair). Within individ-

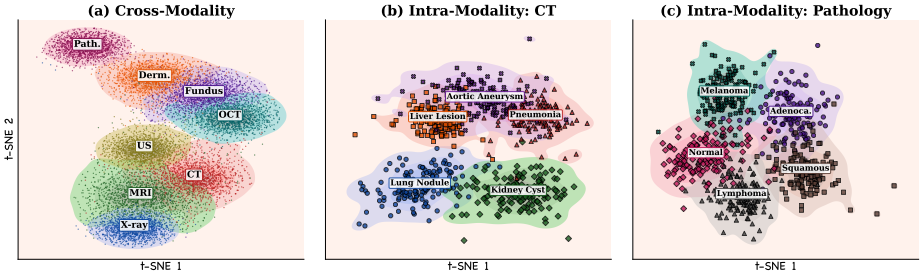

Fig. 8: t-SNE visualization of implicit memory $\mathcal{M}_{auto}$ after CCR. (a) Eight imaging modalities form well-separated clusters with clinically coherent proximity. (b, c) Within CT and Pathology, disease subtypes further segregate into distinct regions.

ual modalities (b, c), disease subtypes further segregate: CT memory separates lung nodules, liver lesions, kidney cysts, pneumonia, and aortic aneurysms, while pathology memory distinguishes adenocarcinoma, squamous cell carcinoma, normal tissue, lymphoma, and melanoma. This hierarchical organization confirms that r_{causal} ***reshapes the latent space into a diagnostically meaningful manifold*** rather than merely boosting task accuracy.

Why latent memory evolution works. Our ablations pinpoint two necessary conditions that general latent methods lack. *First, structured priors are indispensable*: replacing MedSAM3 with a random encoder collapses Avg from 67.7% to 52.0% (Table 2d). *Second, causal calibration activates the priors*: r_{causal} lifts accuracy by 4.1 pp (Table 2b) and reorganizes memory into the hierarchical diagnostic manifold shown in Fig. 8. Neither condition alone suffices, and their synergy is precisely what general latent methods lack.

5 Conclusion and Future Work

We propose MedSynapse-V, a medical vision-language model that performs clinical reasoning through compact latent tokens rather than explicit chain-of-thought generation. By combining causal counterfactual rewards with progressive memory evolution, our approach effectively internalizes diagnostic reasoning within a low-latency framework. Experiments across multiple medical benchmarks show that MedSynapse-V outperforms existing medical VLMs, general-purpose VLMs, and RL-based CoT methods in both accuracy and efficiency, confirming that latent cognitive processes guided by well-designed rewards can effectively replace verbose explicit reasoning in the medical domain.

Looking ahead, we aim to extend latent memory evolution to longitudinal analysis and multi-modal report generation by integrating heterogeneous clinical evidence sources. Our research will further investigate scaling implicit memory to accommodate broader differential diagnosis spaces with hundreds of competing hypotheses, validating the generalizability of latent cognitive architectures for complex clinical decision-making in high-stakes diagnostic environments.

Acknowledgments

This research was partially supported by the National Natural Science Foundation of China (Grants No. 62472157), and the Hunan Provincial Graduate Research Innovation Project (Grant No. CX20250587).

References

1. Bai, S., Cai, Y., Chen, R., Chen, K., Chen, X., Cheng, Z., Deng, L., Ding, W., Gao, C., Ge, C., et al.: Qwen3-vl technical report. arXiv preprint arXiv:2511.21631 (2025)
2. Bose, S., Rajendran, R.K., Debnath, B., Karydis, K., Roy-Chowdhury, A.K., Chakradhar, S.: Visual alignment of medical vision-language models for grounded radiology report generation. arXiv preprint arXiv:2512.16201 (2025)
3. Brunyé, T.T., Drew, T., Weaver, D.L., Elmore, J.G.: A review of eye tracking for understanding and improving diagnostic interpretation. Cognitive research: principles and implications **4**(1), 7 (2019)
4. Chen, C., Ma, Z., Li, Y., Hu, Y., Wei, Y., Li, W., Nie, L.: Reasoning in the dark: Interleaved vision-text reasoning in latent space. arXiv preprint arXiv:2510.12603 (2025)
5. Chen, J., Gui, C., Ouyang, R., Gao, A., Chen, S., Chen, G.H., Wang, X., Cai, Z., Ji, K., Wan, X., et al.: Towards injecting medical visual knowledge into multimodal llms at scale. In: Proceedings of the 2024 conference on empirical methods in natural language processing. pp. 7346–7370 (2024)
6. Chen, K., Rui, S., Jiang, Y., Wu, J., Zheng, Q., Song, C., Wang, X., Zhou, M., Liu, M.: Think twice to see more: Iterative visual reasoning in medical vlms. arXiv preprint arXiv:2510.10052 (2025)
7. Chen, P., Ye, J., Wang, G., Li, Y., Deng, Z., Li, W., Li, T., Duan, H., Huang, Z., Su, Y., et al.: Gmai-mmbench: A comprehensive multimodal evaluation benchmark towards general medical ai. Advances in Neural Information Processing Systems **37**, 94327–94427 (2024)
8. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
9. Deng, Y., Choi, Y., Shieber, S.: From explicit cot to implicit cot: Learning to internalize cot step by step. arXiv preprint arXiv:2405.14838 (2024)
10. Deria, A., Kumar, K., Dukre, A.M., Segal, E., Khan, S., Razzak, I.: Medmo: Grounding and understanding multimodal large language model for medical images. arXiv preprint arXiv:2602.06965 (2026)
11. Ding, M., Zhang, J., Wang, W., Zhong, H., Luo, X., Chen, W., Shen, L.: Mmedexpert-r1: Strengthening multimodal medical reasoning via domain-specific adaptation and clinical guideline reinforcement. arXiv preprint arXiv:2601.10949 (2026)
12. Gai, X., Zhou, C., Liu, J., Feng, Y., Wu, J., Liu, Z.: Medthink: Explaining medical visual question answering via multimodal decision-making rationale. arXiv preprint arXiv:2404.12372 (2024)
13. Geiping, J., McLeish, S., Jain, N., Kirchenbauer, J., Singh, S., Bartoldson, B., Kailkhura, B., Bhatele, A., Goldstein, T.: Scaling up test-time compute with latent reasoning: A recurrent depth approach. Advances in Neural Information Processing Systems **38**, 41340–41391 (2026)

14. Gu, T., Yang, K., Liu, D., Cai, W.: Lapa: Latent prompt assist model for medical visual question answering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4971–4980 (2024)
15. Hao, S., Sukhbaatar, S., Su, D., Li, X., Hu, Z., Weston, J., Tian, Y.: Training large language models to reason in a continuous latent space. arXiv preprint arXiv:2412.06769 (2024)
16. He, X., Zhang, Y., Mou, L., Xing, E., Xie, P.: Pathvqa: 30000+ questions for medical visual question answering. arXiv preprint arXiv:2003.10286 (2020)
17. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
18. Hu, Y., Li, T., Lu, Q., Shao, W., He, J., Qiao, Y., Luo, P.: Omnimedvqa: A new large-scale comprehensive evaluation benchmark for medical lvlm. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22170–22183 (2024)
19. Lai, Y., Zhong, J., Li, M., Zhao, S., Li, Y., Psounis, K., Yang, X.: Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. IEEE transactions on medical imaging (2026)
20. Lau, J.J., Gayen, S., Ben Abacha, A., Demner-Fushman, D.: A dataset of clinically generated visual questions and answers about radiology images. Scientific data **5**(1), 180251 (2018)
21. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.t., Rocktäschel, T., et al.: Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems **33**, 9459–9474 (2020)
22. Li, B., Yan, T., Pan, Y., Luo, J., Ji, R., Ding, J., Xu, Z., Liu, S., Dong, H., Lin, Z., et al.: Mmedagent: Learning to use medical tools with multi-modal agent. In: Findings of the Association for Computational Linguistics: EMNLP 2024. pp. 8745–8760 (2024)
23. Li, C., Wong, C., Zhang, S., Usuyama, N., Liu, H., Yang, J., Naumann, T., Poon, H., Gao, J.: Llava-med: Training a large language-and-vision assistant for biomedicine in one day. Advances in Neural Information Processing Systems **36**, 28541–28564 (2023)
24. Li, H., Li, C., Wu, T., Zhu, X., Wang, Y., Yu, Z., Jiang, E.H., Zhu, S.C., Jia, Z., Wu, Y.N., et al.: Seek in the dark: Reasoning via test-time instance-level policy gradient in latent space. arXiv preprint arXiv:2505.13308 (2025)
25. Li, T., Su, Y., Li, W., Fu, B., Chen, Z., Huang, Z., Wang, G., Ma, C., Chen, Y., Hu, M., et al.: Gmai-vl & gmai-vl-5.5 m: A large vision-language model and a comprehensive multimodal dataset towards general medical ai. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 23177–23185 (2026)
26. Lin, J.: Divergence measures based on the shannon entropy. IEEE Transactions on Information theory **37**(1), 145–151 (1991)
27. Lin, J., Zhu, C., Kneuertz, P.J., Bai, Y., Xue, Y.: When models learn to ask why: Adaptive causal reasoning for trustworthy medical vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5556–5568 (2026)
28. Liu, A., Xue, R., Cao, X.R., Shen, Y., Lu, Y., Li, X., Chen, Q., Chen, J.: Medsam3: Delving into segment anything with medical concepts. arXiv preprint arXiv:2511.19046 (2025)
29. Liu, B., Zhan, L.M., Xu, L., Ma, L., Yang, Y., Wu, X.M.: Slake: A semantically-labeled knowledge-enhanced dataset for medical visual question answering. In: 2021

IEEE 18th international symposium on biomedical imaging (ISBI). pp. 1650–1654. IEEE (2021)
30. Liu, L., Pfeiffer, J., Wu, J., Xie, J., Szlam, A.: Deliberation in latent space via differentiable cache augmentation. arXiv preprint arXiv:2412.17747 (2024)
31. Moor, M., Huang, Q., Wu, S., Yasunaga, M., Dalmia, Y., Leskovec, J., Zakka, C., Reis, E.P., Rajpurkar, P.: Med-flamingo: a multimodal medical few-shot learner. In: Machine learning for health (ML4H). pp. 353–367. PMLR (2023)
32. Mullappilly, S.S., Kurpath, M.I., Mohamed, O., Zidan, M., Khan, F., Khan, S., Anwer, R., Cholakkal, H.: Medix-r1: Open ended medical reinforcement learning. arXiv preprint arXiv:2602.23363 (2026)
33. Mullappilly, S.S., Kurpath, M.I., Pieri, S., Alseiari, S.Y., Cholakkal, S., Aldahmani, K., Khan, F., Anwer, R., Khan, S., Baldwin, T., et al.: Bimedix2: Bio-medical expert lmm for diverse medical modalities. arXiv preprint arXiv:2412.07769 (2024)
34. Nath, V., Li, W., Yang, D., Myronenko, A., Zheng, M., Lu, Y., Liu, Z., Yin, H., Law, Y.M., Tang, Y., et al.: Vila-m3: Enhancing vision-language models with medical expert knowledge. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 14788–14798 (2025)
35. Norman, G.: Dual processing and diagnostic errors. Advances in Health Sciences Education **14**(Suppl 1), 37–49 (2009)
36. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al.: Training language models to follow instructions with human feedback. Advances in neural information processing systems **35**, 27730–27744 (2022)
37. Pan, J., Liu, C., Wu, J., Liu, F., Zhu, J., Li, H.B., Chen, C., Ouyang, C., Rueckert, D.: Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 337–347. Springer (2025)
38. Pham, T.H., Ngo, C.: Multimodal chain of continuous thought for latent-space reasoning in vision-language models. arXiv preprint arXiv:2508.12587 (2025)
39. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. Advances in neural information processing systems **36**, 53728–53741 (2023)
40. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
41. Sellergren, A., Kazemzadeh, S., Jaroensri, T., Kiraly, A., Traverse, M., Kohlberger, T., Xu, S., Jamil, F., Hughes, C., Lau, C., et al.: Medgemma technical report. arXiv preprint arXiv:2507.05201 (2025)
42. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
43. Shen, Z., Yan, H., Zhang, L., Hu, Z., Du, Y., He, Y.: Codi: Compressing chain-of-thought into continuous space via self-distillation. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. pp. 677–693 (2025)
44. Su, Y., Li, T., Liu, J., Ma, C., Ning, J., Tang, C., Ju, S., Ye, J., Chen, P., Hu, M., et al.: Gmai-vl-r1: Harnessing reinforcement learning for multimodal medical reasoning. arXiv preprint arXiv:2504.01886 (2025)
45. Sun, H., Jiang, Y., Lou, W., Zhang, Y., Li, W., Wang, L., Liu, M., Liu, L., Wang, X.: Chiron-o1: Igniting multimodal large language models towards generalizable

medical reasoning via mentor-intern collaborative search. Advances in Neural Information Processing Systems **38**, 104180–104217 (2026)
46. Tan, W., Li, J., Ju, J., Luo, Z., Song, R., Luan, J.: Think silently, think fast: Dynamic latent compression of llm reasoning chains. Advances in Neural Information Processing Systems **38**, 4646–4668 (2026)
47. Van Sonsbeek, T., Derakhshani, M.M., Najdenkoska, I., Snoek, C.G., Worring, M.: Open-ended medical visual question answering through prefix tuning of language models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 726–736. Springer (2023)
48. Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., Reede, D.: Interpretive error in radiology. American Journal of Roentgenology **208**(4), 739–749 (2017)
49. Wang, Y., Liu, J., Gao, S., Feng, B., Tang, Z., Gai, X., Wu, J., Liu, Z.: V2t-cot: From vision to text chain-of-thought for medical reasoning and diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 658–668. Springer (2025)
50. Wu, C., Zhang, X., Zhang, Y., Hui, H., Wang, Y., Xie, W.: Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. Nature Communications **16**(1), 7866 (2025)
51. Wu, J., Deng, W., Li, X., Liu, S., Mi, T., Peng, Y., Xu, Z., Liu, Y., Cho, H., Choi, C.I., et al.: Medreason: Eliciting factual medical reasoning steps in llms via knowledge graphs. arXiv preprint arXiv:2504.00993 (2025)
52. Wu, J., Zhu, J., Qi, Y., Chen, J., Xu, M., Menolascina, F., Grau, V.: Medical graph rag: Towards safe medical large language model via graph retrieval-augmented generation. arXiv preprint arXiv:2408.04187 (2024)
53. Xu, Y., Guo, X., Zeng, Z., Miao, C.: Softcot: Soft chain-of-thought for efficient reasoning with llms. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 23336–23351 (2025)
54. Xu, Z., Wang, H., Bespalov, D., Wu, X., Stone, P., Qi, Y.: Lars: Latent reasoning skills for chain-of-thought reasoning. In: Findings of the Association for Computational Linguistics: EMNLP 2024. pp. 3624–3643 (2024)
55. Yu, H., Cheng, T., Cheng, Y., Feng, R.: Finemedlm-o1: Enhancing the medical reasoning ability of llm from supervised fine-tuning to test-time training. arXiv e-prints pp. arXiv–2501 (2025)
56. Yu, X., Xu, C., Zhang, G., Chen, Z., Zhang, Y., He, Y., Jiang, P.T., Zhang, J., Hu, X., Yan, S.: Vismem: Latent vision memory unlocks potential of vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 31544–31555 (2026)
57. Yue, X., Ni, Y., Zhang, K., Zheng, T., Liu, R., Zhang, G., Stevens, S., Jiang, D., Ren, W., Sun, Y., et al.: Mmmu: A massive multi-discipline multimodal understanding and reasoning benchmark for expert agi. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9556–9567 (2024)
58. Zakka, C., Shad, R., Chaurasia, A., Dalal, A.R., Kim, J.L., Moor, M., Fong, R., Phillips, C., Alexander, K., Ashley, E., et al.: Almanac—retrieval-augmented language models for clinical medicine. Nejm ai **1**(2), AIoa2300068 (2024)
59. Zhang, G., Fu, M., Yan, S.: Memgen: Weaving generative latent memory for self-evolving agents. arXiv preprint arXiv:2509.24704 (2025)
60. Zhang, W., Guo, J., Zhang, H., Zhang, P., Chen, J., Zhang, S., Zhang, Z., Yi, Y., Bu, H.: Patho-agenticrag: towards multimodal agentic retrieval-augmented generation for pathology vlms via reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 29921–29929 (2026)

61. Zhang, X., Wu, C., Zhao, Z., Lin, W., Zhang, Y., Wang, Y., Xie, W.: Pmc-vqa: Visual instruction tuning for medical visual question answering. arXiv preprint arXiv:2305.10415 (2023)
62. Zhang, Y., Xu, W., Zhao, X., Wang, W., Feng, F., He, X., Chua, T.S.: Reinforced latent reasoning for llm-based recommendation. arXiv preprint arXiv:2505.19092 (2025)
63. Zhao, X., Liu, S., Yang, S.Y., Miao, C.: Medrag: Enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot. In: Proceedings of the ACM on Web Conference 2025. pp. 4442–4457 (2025)
64. Zhu, C., Lin, Y., Chen, S., Wang, Y., Lin, J.: Medeyes: Learning dynamic visual focus for medical progressive diagnosis. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 13916–13924 (2026)
65. Zhu, J., Wang, W., Chen, Z., Liu, Z., Ye, S., Gu, L., Tian, H., Duan, Y., Su, W., Shao, J., et al.: Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. arXiv preprint arXiv:2504.10479 (2025)
66. Zhu, K., Xia, P., Li, Y., Zhu, H., Wang, S., Yao, H.: Mmedpo: Aligning medical vision-language models with clinical-aware multimodal preference optimization. arXiv preprint arXiv:2412.06141 (2024)
67. Zuo, Y., Qu, S., Li, Y., Chen, Z., Zhu, X., Hua, E., Zhang, K., Ding, N., Zhou, B.: Medxpertqa: Benchmarking expert-level medical reasoning and understanding. arXiv preprint arXiv:2501.18362 (2025)