COVERAGE-AWARE REASONING WITH MEDICAL TOKENS FOR DIAGNOSIS PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Large language models (LLMs) offer promising potential for next-visit diagnosis prediction, owing to their ability to integrate longitudinal clinical evidence and reason over it in natural language. However, reinforcement learning for LLM reasoning commonly rewards each trajectory according to the correctness of its final answer. In next-visit diagnosis prediction, multiple diagnoses can be simultaneously valid, but independently rewarding one diagnosis per trajectory does not distinguish repeated hits from coverage of different diagnoses. The policy can therefore concentrate on a few correct diagnoses, leaving others uncovered. Meanwhile, LLM tokenizers can split ICD codes into several generic tokens with limited clinical meaning, requiring multiple decoding steps to predict each diagnosis and hindering reasoning over a large disease vocabulary. To address both challenges, we propose CARing, a framework that represents diagnoses with compositional Semantic IDs (SIDs) and optimizes reasoning trajectories for multi-label coverage. Concretely, we first encode ontology-enriched disease semantics into compact SIDs through residual quantization, and ground the resulting SID tokens in natural language and longitudinal EHR contexts through multi-task alignment and reasoning-enriched training to unlock transferable LLM reasoning. CARing further improves unordered multi-label prediction through a coverage reward for reinforcement learning and multi-positive supervision. At inference time, the model supports both efficient direct constrained decoding and multi-chain reasoning with rank fusion. On MIMIC-III and MIMIC-IV, CARing exceeds all EHR-trained baselines in weighted F_1 and attains the highest top-k recall at every reported cutoff, including R@30 of 46.04% and 46.52% in reasoning mode.

1 INTRODUCTION

Next-visit diagnosis prediction aims to infer diagnoses at a patient's next encounter from their previous visits. Existing EHR predictors learn from longitudinal patient histories (Choi et al., 2016a; Ma et al., 2018; Rasmy et al., 2021; Chen et al., 2024; Ke et al., 2026). Large language models (LLMs) offer a way to incorporate explicit reasoning into this task by synthesizing historical evidence and generating intermediate clinical hypotheses in natural language (Jiang et al., 2025; Cao et al., 2026). Reinforcement learning is commonly used to improve LLM reasoning by rewarding the correctness of each trajectory's final answer (Shao et al., 2024; Liu et al., 2025). However, next-visit diagnosis prediction is an *unordered multi-label prediction* problem over a large diagnosis vocabulary. Multiple diagnoses can be simultaneously correct, and their order within a visit has no intrinsic meaning. When each trajectory predicts one diagnosis and is rewarded only for its own correctness, a group that repeats one correct diagnosis earns the same total reward as an equally sized group covering different correct diagnoses. The policy can therefore concentrate on a few correct diagnoses and leave others uncovered (Yue et al., 2025; GX-Chen et al., 2026), as illustrated in Figure 1.

Effective reasoning also requires a diagnosis representation that an LLM can understand and generate efficiently. Although ICD codes organize diagnoses hierarchically, their letters and numbers do not describe the diseases in ordinary language. General-purpose LLM tokenizers can split a single code into several character-level tokens rather than treat it as a disease-level unit, and such fragmented codes are hard to ground, reason over, and generate across a large diagnosis vocabulary (Soroush et al., 2024). Medical tokenizers such as MedTok and MedRQ compress code descriptions, relations, and ontologies into reusable discrete codebooks (Su et al., 2025; Lv et al., 2026), but

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.

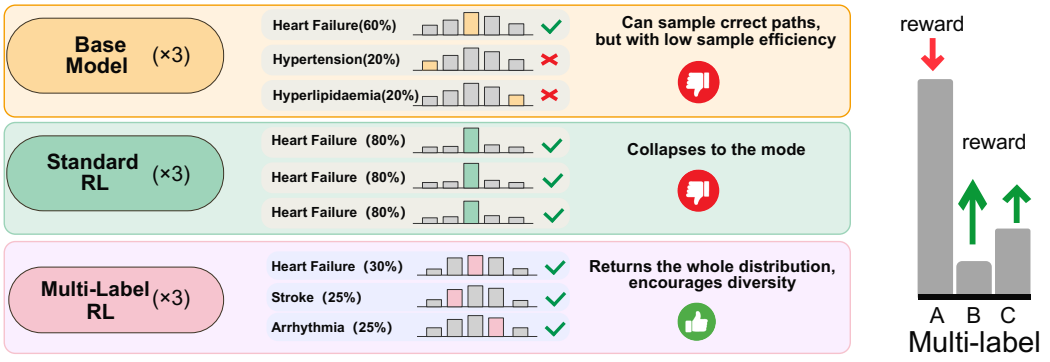

Figure 1: Motivation for multi-label reinforcement learning in next-visit diagnosis prediction. Standard RL can collapse onto one correct diagnosis, leaving others uncovered. Our coverage reward discounts repeated predictions and encourages coverage of distinct correct diagnoses.

their downstream predictors use the resulting representations as model inputs rather than as tokens for reasoning and prediction. MERA maps each disease to an atomic special token and handles the unordered diagnoses within a visit (Ma et al., 2025), but its vocabulary and associated parameters grow linearly with the number of diagnoses, and natural language reasoning stays outside its prediction path. Semantic IDs (SIDs) in recommendation suggest a compositional alternative: items reuse codebook tokens (Rajput et al., 2023; Zheng et al., 2024), and recent systems let an LLM reason in natural language before generating an identifier (Liu et al., 2026; Team et al., 2026). Inspired by this approach, we use medical SIDs as a shared token language linking disease concepts to longitudinal patient histories. Unlike recommendation, where the target is usually a single item, a future visit can contain multiple correct diagnoses, so the model must recover complementary labels.

We propose CARing (**C**overage-**A**ware **R**easoning with Compositional Medical Tokens for Multi-Label Next-Visit Diagnosis Prediction), a framework that brings LLM reasoning to diagnosis prediction through medical SIDs and trains reasoning trajectories to cover multiple correct diagnoses.

Our technical contributions are as follows:

- **Adapting LLM reasoning to diagnosis prediction through medical SIDs.** We use medical SIDs to connect clinical language, patient histories, and diagnosis outputs within one autoregressive model. Disease semantics and ICD ontology paths are compressed through residual quantization into shared codebook tokens. Six alignment tasks and narratives generated by a teacher LLM ground these tokens in clinical context. Reasoning activation then trains the model to generate a clinical rationale before predicting a diagnosis SID.
- **Optimizing reasoning trajectories for diagnosis coverage.** A coverage reward divides each correct prediction's credit by its frequency within a group of sampled trajectories, so the total group reward equals the number of distinct correct diagnoses. An auxiliary multi-positive objective supervises only the SID answer tokens, supplying learning signals for correct diagnoses that policy sampling may miss. These objectives encourage trajectories to recover complementary diagnoses.
- **Direct prediction and explicit reasoning within a common inference framework.** The direct mode ranks diagnoses through constrained decoding without generating a rationale. The reasoning mode samples multiple chains of thought before prediction, ranks diagnoses by constrained decoding after each chain, and combines the rankings with reciprocal rank fusion. Both modes restrict predictions to valid medical SIDs.

We evaluate CARing on MIMIC-III and MIMIC-IV (Johnson et al., 2016; 2023). CARing outperforms all baselines trained on EHR data in weighted F_1 and achieves the highest top-k recall among all evaluated methods. The reasoning mode reaches R@30 of 46.04% and 46.52% on the two datasets, respectively. Ablation studies show that semantic enrichment improves direct prediction and supports the complementary roles of the coverage reward and multi-positive supervision.

2 Related Work

We briefly summarize related work and refer readers to Appendix B for a detailed discussion.

Next-visit diagnosis prediction is commonly addressed by EHR models that encode visit sequences with recurrent or attention architectures (Choi et al., 2016a;b; Gao et al., 2020; Yang et al., 2023b) and incorporate medical ontologies or patient graphs (Ma et al., 2018; Lu et al., 2021; Chen et al., 2024; Tan et al., 2025). These models output scores over a fixed diagnosis vocabulary rather than generating diagnoses through explicit reasoning. Bringing LLMs to this task requires diagnosis tokens that the model can read and generate. Medical tokenizers learn discrete codes for diseases but mainly feed them to downstream predictors (Su et al., 2025; Lv et al., 2026), while MERA adds one vocabulary token per disease, so the vocabulary grows with the number of diagnoses (Ma et al., 2025). Semantic IDs in recommendation instead compose each item from shared codebook tokens, and recent systems let an LLM reason in text before generating the identifier (Liu et al., 2026; Team et al., 2026). Explicit reasoning has also improved clinical prediction (Jiang et al., 2025; Cao et al., 2026), but mainly for binary outcomes or small label spaces. Our work represents diseases with compositional medical SIDs, so that one LLM reads patient histories, reasons in natural language, and generates diagnoses from a large vocabulary in which a visit has several correct labels.

Large reasoning models generate intermediate steps before the final answer (Wei et al., 2022) and gain accuracy from sampling multiple chains at test time (Wang et al., 2023; Snell et al., 2025). Reinforcement learning with verifiable rewards (RLVR) trains such reasoning from outcome feedback (Shao et al., 2024; Guo et al., 2025), but rewarding each trajectory only for its own correctness tends to concentrate the policy on a few answers. RLVR raises pass@1 while base models reach higher pass@k at large k (Yue et al., 2025), and weak KL regularization with equal rewards tends to specify a unimodal target distribution (GX-Chen et al., 2026). Recent methods optimize diversity together with correctness by scoring the whole rollout group, favoring rarely sampled valid outputs (Anschel et al., 2025) or each trajectory's marginal contribution to group diversity (Li et al., 2026). In diagnosis prediction, recall depends on how many correct diagnoses of a visit the model covers. Our coverage reward divides each correct SID's credit by its frequency in the group, so the group reward counts distinct correct diagnoses, and a multi-positive loss supplies signal for correct diagnoses that no trajectory samples.

3 Method

Our framework first constructs ontology-enriched medical SIDs from shared tokens, then grounds these tokens in clinical text and longitudinal patient histories, and finally optimizes reasoning trajectories for multi-label diagnosis coverage. Figure 2 illustrates the overall pipeline.

3.1 Task Formulation

Let $\mathcal{D}$ denote the diagnosis vocabulary. A patient's longitudinal record up to visit t is $H_t = (V_1, \ldots, V_t)$, where $V_j \subseteq \mathcal{D}$ is the set of diagnoses at visit j. The prediction target is the unordered diagnosis set at the next visit, $Y_t = V_{t+1} = \{d_1, \ldots, d_m\} \subseteq \mathcal{D}$.

To avoid imposing an order on the target diagnoses, we expand each visit-level example (H_t, Y_t) into case-level instances $\{(H_t, d_i)\}_{i=1}^{m}$. Each instance pairs the same history with one target. We use them for SID–language alignment and reasoning activation, with each trajectory predicting one diagnosis SID. Reinforcement learning and visit-level evaluation retain the complete target set Y_t.

3.2 Ontology-Enriched Compositional Medical Tokenization

Ontology-enriched disease semantics. For each disease $d \in \mathcal{D}$, we pair its canonical title $n(d)$ with its root-to-leaf ICD ontology path $p(d)$, represented schematically as $u_d = \texttt{Disease: } n(d) \oplus \texttt{ ICD ontology path: } p(d)$. We encode the textual metadata u_d into a continuous semantic representation $\mathbf{x}_d$, which is then fed into a residual quantization module.

Residual quantization. Following residual-quantized VAEs (RQ-VAEs) (Zeghidour et al., 2022), an MLP encoder maps $\mathbf{x}_d$ to a latent vector $\mathbf{z}_d = E(\mathbf{x}_d) \in \mathbb{R}^d$. Given L codebooks $\mathcal{C}_l = \{\mathbf{e}_k^{(l)}\}_{k=1}^{K}$, residual quantization starts with $\mathbf{r}_d^{(0)} = \mathbf{z}_d$. At each level, it selects $c_d^{(l)} = \arg\min_k \|\mathbf{r}_d^{(l)} - \mathbf{e}_k^{(l)}\|_2^2$

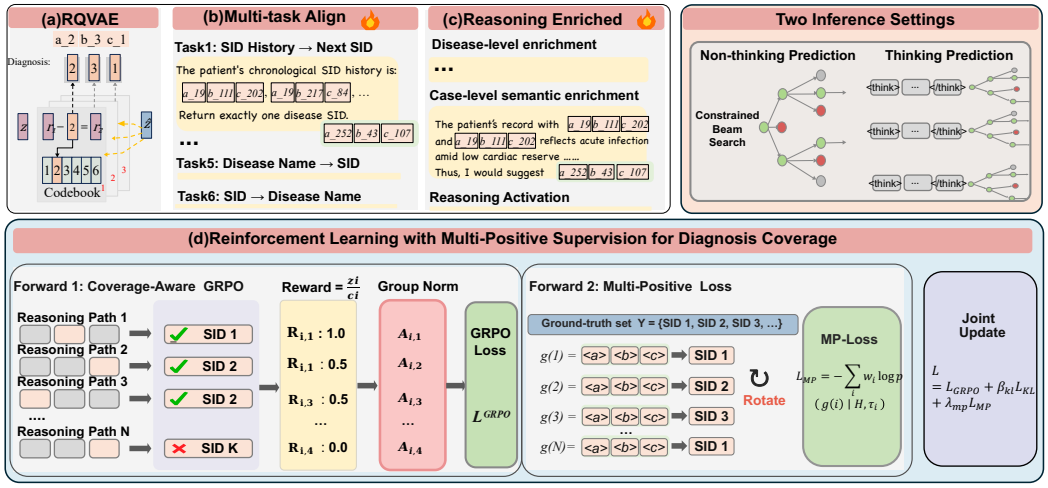

Figure 2: Overview of CARing. The framework constructs compositional medical SIDs, aligns them with clinical language and reasoning, combines a coverage reward with multi-positive answer supervision, and supports direct and multi-chain reasoning inference.

and updates the residual as $\mathbf{r}_d^{(l+1)} = \mathbf{r}_d^{(l)} - \mathbf{e}^{(l)}_{c_d^{(l)}}$. The quantized latent $\mathbf{z}_{q,d} = \sum_{l=0}^{L-1} \mathbf{e}^{(l)}_{c_d^{(l)}}$ is decoded back to the original embedding space. Each selected codebook entry corresponds to a medical token. Ordered by quantization level, these tokens form the SID sequence $q(d) = (s_d^1, \ldots, s_d^L)$ for disease d. These components are jointly trained with $\mathcal{L}_{\text{tok}} = \|\mathbf{x}_d - D(\mathbf{z}_{q,d})\|_2^2 + \frac{1}{L}\sum_{l=0}^{L-1}\big(\|\text{sg}[\mathbf{r}_d^{(l)}] - \mathbf{e}^{(l)}_{c_d^{(l)}}\|_2^2 + \beta\|\mathbf{r}_d^{(l)} - \text{sg}[\mathbf{e}^{(l)}_{c_d^{(l)}}]\|_2^2\big)$, where $\text{sg}[\cdot]$ is stop-gradient and β controls the commitment term. The resulting SID is a compact sequence of discrete tokens that preserves disease semantics for autoregressive modeling.

3.3 Multi-Task SID–Clinical Context Alignment

Although the RQ-VAE preserves disease-level similarity, its component tokens are newly added to the LLM vocabulary and initially have no lexical meaning. To enable the LLM to interpret, reason over, and generate SID tokens in clinical contexts, we conduct full-parameter autoregressive training on six complementary tasks, where SID tokens and natural language tokens co-occur across diverse medical scenarios. Four *next-visit prediction* tasks cover the combinations required by the downstream model: SID history → SID (T1), textual disease history → SID (T2), SID history → disease name (T3), and textual disease history → disease name (T4). Two *concept translation* tasks establish bidirectional grounding: disease name → SID (T5) and SID → disease name (T6). T1–T4 jointly align textual and SID histories with both language and SID outputs, with T1 matching the final prediction interface; T5–T6 teach the SID vocabulary to encode recognizable disease semantics. All tasks are jointly optimized under the standard next-token prediction objective. Due to space limitations, detailed input and output formats for all six tasks are provided in Appendix C.1.

3.4 Reasoning-Enriched Training

The six alignment tasks ground SIDs through concise translation and prediction templates, but expose them to limited linguistic and clinical contexts. To address this issue, we leverage a large teacher language model with rich world knowledge. We augment SID–language alignment with two offline teacher-generated corpora and general-domain reasoning data, then train the aligned model to follow an explicit reason-then-predict pathway.

Disease-level semantic enrichment. For each disease, the teacher generates a clinical narrative in which its SID denotes the concept across manifestations, investigations, comorbidities, and ontology relations. To ground the description, each prompt includes metadata from the PyHealth ICD-9-CM ontology (Yang et al., 2023a) and the 2015 Clinical Classifications Software (CCS) and Chronic Condition Indicator (CCI) released by the AHRQ Healthcare Cost and Utilization Project (Agency

for Healthcare Research and Quality, 2015b;a). The resulting text interleaves SID tokens with natural language. Training on the entire narrative with next-token prediction exposes each SID to clinical language beyond the fixed T1–T6 templates.

Case-level semantic enrichment. For each case-level training instance, the teacher first receives the historical SID visits and a reference target used only as private guidance, and generates a reasoning path r. A second teacher call integrates the history, r, and the target into a coherent narrative.

We concatenate these two enriched corpora, the T1–T6 datasets, and general-domain reasoning data for continued full-parameter alignment training. The general data are intended to mitigate degradation of the backbone's language and reasoning abilities.

Reasoning activation. Rich alignment does not by itself ensure that the model will externalize reasoning before predicting a SID. We reuse the teacher-generated reasoning path r from case-level semantic enrichment to construct supervised samples for each history–diagnosis pair (H, d). Starting from the enriched alignment checkpoint, we apply supervised fine-tuning to history-only prompts, supervising r followed by the target SID $q(d)$ while masking the prompt tokens. This fine-tuning targets instruction following and the reason-then-SID response format.

Representative training sequences are provided in Appendix C.2, and their objectives are defined in Appendix C.2.1. The teacher prompts for data enrichment are provided in Appendix C.3.

3.5 Reinforcement Learning with Multi-Positive Supervision

Reasoning activation uses case-level instances with one target SID, whereas a future visit generally contains multiple unordered correct diagnoses. An exact match (EM) reward assigns the same credit to every correct prediction, even when several trajectories repeat one diagnosis. It therefore cannot distinguish repeated hits from coverage of different diagnoses. We address this mismatch with GRPO (Shao et al., 2024) at the visit level, combining a coverage reward over sampled trajectories with supervision over the full positive set.

Sampling by visit. For each history H, we sample N trajectories from the old policy. Each trajectory consists of a reasoning sequence τ_i followed by a three-token SID prediction y_i. Rewards are computed against the complete next-visit SID set $Y = \{q(d) : d \in V_{t+1}\}$.

Coverage reward. Let $z_i = \mathbf{1}[y_i \in Y]$ indicate a correct prediction and let $c_i = \sum_{j=1}^{N} \mathbf{1}[y_j = y_i]$ be its frequency in the rollout group. We assign

$$r_i = \frac{z_i}{c_i}. \tag{1}$$

Each distinct correct SID predicted by the group contributes one unit to the total reward, regardless of how often it appears (Appendix D). When rewards vary within the group, GRPO assigns larger advantages to less-repeated correct SIDs than to repeated hits, encouraging trajectories to cover complementary correct diagnoses.

Policy optimization. After computing the shaped rewards, GRPO compares trajectories generated from the same history without training a value model. Let μ_H and σ_H be the mean and standard deviation of the N rewards. Each rollout receives the normalized outcome advantage

$$\mu_H = \frac{1}{N}\sum_{j=1}^{N} r_j, \qquad \sigma_H = \mathrm{Std}(\{r_j\}_{j=1}^{N}), \qquad \widehat{A}_i = \frac{r_i - \mu_H}{\sigma_H + \epsilon}. \tag{2}$$

Here $\epsilon > 0$ stabilizes the normalization. We minimize the token-averaged GRPO objective

$$\mathcal{L}_{\mathrm{GRPO}} = -\frac{1}{\sum_i T_i}\sum_{i=1}^{N}\sum_{t=1}^{T_i} \min\left[\rho_{i,t}(\theta)\widehat{A}_i,\ \mathrm{clip}(\rho_{i,t}(\theta), 1-\varepsilon, 1+\varepsilon)\widehat{A}_i\right]. \tag{3}$$

Here $o_i = (a_{i,1}, \ldots, a_{i,T_i})$ contains the valid response tokens and $h_{i,t} = (H, a_{i,<t})$ is the context of token t. The token-level importance ratio is $\rho_{i,t}(\theta) = \pi_\theta(a_{i,t} \mid h_{i,t})/\pi_{\theta_{\mathrm{old}}}(a_{i,t} \mid h_{i,t})$, and $\varepsilon > 0$

is the clipping width. The same $\widehat{A}_i$ applies to the reasoning and SID answer tokens. Minimizing this objective increases the likelihood of reasoning and answer tokens in above-average trajectories and decreases it for below-average trajectories, subject to clipping. Further optimization details are given in Appendix E.4.

Multi-positive supervision. Coverage reward gives no policy-gradient signal for correct diagnoses absent from the rollout group. We therefore supervise alternative positives at the answer level. Let $\mathcal{I}$ index rollouts with an identifiable three-token terminal SID. For $i \in \mathcal{I}$, we retain its reasoning prefix τ_i, replace the answer with $g(i) \in Y$, and optimize only the L SID tokens:

$$\mathcal{L}_{\mathrm{MP}} = -\frac{1}{N} \sum_{i \in \mathcal{I}} w_i \, \frac{1}{L} \sum_{l=1}^{L} \log \pi_\theta \left(s^l_{g(i)} \mid H, \tau_i, s^{<l}_{g(i)} \right). \tag{4}$$

We assign $g(i)$ round-robin over a fixed enumeration of Y, with a step-dependent starting offset. For $S = \{g(i) : i \in \mathcal{I}\}$ and $n_g = |\{i \in \mathcal{I} : g(i) = g\}|$, we set $w_i = N/(|S| n_{g(i)})$, giving each selected SID equal total weight. The loss is zero when $\mathcal{I}$ is empty.

Joint optimization. The actor is optimized with

$$\mathcal{L} = \mathcal{L}_{\mathrm{GRPO}} + \beta_{\mathrm{KL}} \, \widehat{D}_{\mathrm{KL}}(\pi_\theta \| \pi_{\mathrm{ref}}) + \lambda_{\mathrm{MP}} \mathcal{L}_{\mathrm{MP}}, \tag{5}$$

where β_{KL} and λ_{MP} weight the KL and multi-positive terms, and π_{ref} is the fixed reasoning activation policy; see Appendix E.4 for gradients.

3.6 Two Inference Settings

Because next-visit diagnoses form an unordered set, we rank individual diagnosis SIDs rather than generate the full set as one autoregressive sequence. Both settings use catalog-constrained beam search to obtain valid, deduplicated candidates (Appendix E.5).

Direct inference (Non-Thinking). We use the checkpoint after reasoning activation. With explicit thinking disabled, the model conditions on the longitudinal SID history and ranks diagnoses by their constrained decoding scores.

Multi-chain reasoning (Thinking). We use the checkpoint after joint optimization with the coverage reward and multi-positive loss. From the same history, we sample R reasoning chains $\{\tau_r\}_{r=1}^{R}$. Constrained decoding then produces a diagnosis ranking $\mathcal{R}_r$ conditioned on each chain. We rank the union of these candidates using reciprocal rank fusion (RRF) (Cormack et al., 2009):

$$S_{\mathrm{RRF}}(d \mid H) = \sum_{r:\, d \in \mathcal{R}_r} \frac{1}{\kappa + \mathrm{rank}_{\mathcal{R}_r}(d)}, \tag{6}$$

where $\kappa > 0$ controls the weight assigned to top-ranked candidates.

4 Experiments

In this section, we aim to answer the following research questions:

- **RQ1:** How does CARing compare with ML and LLM baselines?
- **RQ2:** Do ICD ontology paths improve SID quality and diagnosis recall?
- **RQ3:** How do different design choices affect the performance of CARing?
- **RQ4:** How do the coverage reward and multi-positive loss affect diagnosis coverage and diversity?

Datasets. We conduct experiments on two real-world electronic health record datasets, MIMIC-III (Johnson et al., 2016) and MIMIC-IV (Johnson et al., 2023), following the preprocessing protocol of prior diagnosis-prediction studies (Ma et al., 2018; Chen et al., 2024). Due to the larger scale of MIMIC-IV, we sample 10,000 patients for our experiments. Detailed preprocessing procedures and dataset statistics are provided in Appendix E.1.

Table 1: Overall next-visit diagnosis prediction performance on MIMIC-III and MIMIC-IV. All values are percentages. Best and second-best results are shown in bold and underlined, respectively.

Model	MIMIC-III					MIMIC-IV				
	w-F_1	R@10	R@20	R@30	R@40	w-F_1	R@10	R@20	R@30	R@40
RETAIN	21.25	24.13	34.30	40.12	45.91	19.54	25.42	33.82	38.91	42.78
KAME	18.58	22.55	31.43	37.67	42.54	10.59	17.97	25.01	29.56	33.15
StageNet	18.13	21.36	30.37	36.64	41.10	11.46	18.41	25.29	30.20	33.82
CGL	21.79	25.51	35.28	41.52	46.75	20.08	24.07	33.96	40.04	44.31
TRANS	19.10	23.32	32.26	38.39	43.14	15.84	20.55	28.71	34.37	38.67
BoxLM	22.48	25.33	35.87	41.82	46.79	21.07	24.15	34.79	41.67	44.53
HieraMed	21.81	24.84	34.65	41.20	46.32	19.10	24.18	33.78	40.87	44.73
DeepSeek non-thinking	<u>30.60</u>	25.92	36.04	40.89	43.60	**30.22**	27.81	35.66	38.90	40.20
DeepSeek thinking	**30.84**	<u>27.26</u>	<u>37.04</u>	42.57	46.34	<u>27.16</u>	<u>29.67</u>	36.88	39.71	42.45
CARing-Non-think	25.09	26.29	<u>37.04</u>	<u>43.85</u>	**50.67**	24.58	28.24	<u>37.85</u>	<u>43.57</u>	<u>48.82</u>
CARing-Think	27.58	**29.15**	**39.69**	**46.04**	<u>50.58</u>	26.62	**31.49**	**41.23**	**46.52**	**50.19**

Evaluation Metrics. We adopt weighted F_1 score (w-F_1) (Bai et al., 2018; Lu et al., 2021) and top-k recall (R@k) (Choi et al., 2016a; Lu et al., 2021) for next-visit diagnosis prediction. w-F_1 is a weighted sum of the F_1 score for each diagnosis. R@k is the ratio of ground-truth diagnoses retrieved among the top-k predictions to the total number of ground-truth diagnoses.

Baselines. We compare with four groups of baselines: (1) *neural EHR predictors*, including RETAIN (Choi et al., 2016b), KAME (Ma et al., 2018), StageNet (Gao et al., 2020), CGL (Lu et al., 2021), and TRANS (Chen et al., 2024); (2) a *PLM-based predictor*, BoxLM (Tan et al., 2025); (3) an *LLM-assisted predictor*, HieraMed (Ke et al., 2026), which embeds LLM-generated medical-code descriptions; and (4) a *zero-shot LLM*, DeepSeek-V4-Flash (DeepSeek-AI, 2026), evaluated without task-specific parameter updates. Please refer to Appendix E.2 for more details.

Implementation Details. We adopt Qwen3-1.7B (Yang et al., 2025) as the backbone and perform full-parameter fine-tuning throughout all stages. DeepSeek-V4-Flash serves as the teacher model for corpus enrichment. The effective batch sizes for multi-task SID–language alignment, reasoning activation, and reinforcement learning are 1024, 1024, and 256, with learning rates of 3×10^{-4}, 1×10^{-5}, and 1×10^{-6}, respectively. For reinforcement learning, we sample 16 rollouts per input and set the multi-positive loss weight to $\lambda_{\text{MP}} = 5 \times 10^{-3}$. All experiments are conducted on four NVIDIA A800 GPUs. Further details are provided in Appendix E.3.

4.1 Main Results (RQ1)

Performance across both datasets. Table 1 shows that CARing has the highest w-F_1 among EHR-trained models on both datasets and the highest recall at every reported cutoff across all methods. The thinking setting leads at seven of eight recall cutoffs; direct inference is 0.09 points higher on MIMIC-III R@40. DeepSeek-V4-Flash has the highest w-F_1 overall on both datasets, and its thinking mode exceeds CARing-Think by 0.54 points on MIMIC-IV. Compared with the direct setting, the thinking model scores 2.49 and 2.04 points higher in w-F_1 on MIMIC-III and MIMIC-IV.

Generalization on MIMIC-IV. The MIMIC-IV cohort contains 10,000 sampled patients, and many diagnoses in its test set do not appear in the EHR training split. All seven EHR-trained baselines lose 1.41–7.99 points in w-F_1 relative to MIMIC-III; nearly all their recall scores also decline. CARing-Think loses 0.96 w-F_1 points (27.58 to 26.62), but R@10, R@20, and R@30 increase by 2.34, 1.54, and 0.48 points, respectively, while R@40 decreases by 0.39 points. It remains 5.55 w-F_1 points above the strongest EHR-trained baseline, BoxLM, on MIMIC-IV. The smaller w-F_1 decline and mixed recall changes are consistent with semantic enrichment and pretraining supporting prediction under sparse label coverage.

4.2 ICD Ontology Paths in Medical SIDs (RQ2)

Table 2: SID quality and MIMIC-III recall with disease names alone or names plus ICD ontology paths. Entropy is in bits; its maximum is 8 for a 256-entry codebook.

Metric	Layer	Name only	Name + ICD path
R@30 (%)		38.97	**39.73**
Codebook utilization (%)	1	100.00	100.00
	2	98.83	**100.00**
	3	97.27	**100.00**
Token distribution entropy	1	7.943	**7.951**
	2	7.542	**7.919**
	3	7.016	**7.736**

Effectiveness of ontology enrichment. In Table 2, the name plus path variant scores 39.73 R@30 versus 38.97 for name alone after multi-task alignment. It uses all three codebooks fully. The SID entropy is also higher in all layers, indicating a more even distribution across codebook tokens. Appendix A examines the semantic structure of these SIDs through randomly sampled prefix cases and a full-vocabulary comparison of the two variants.

4.3 Ablation Study (RQ3)

Effect of Reasoning-Enriched Training for CARing-Non-think. S1 denotes the checkpoint after multi-task SID–clinical context alignment, before semantic enrichment. Semantic enrichment (S2) raises w-F_1 by 2.04 points and R@30 by 1.21 points, showing that SID-grounded narratives add clinical semantics beyond fixed templates. General corpora (S3) yield small EHR gains (0.37 w-F_1 and 0.39 R@30), but IFEval and GSM8K in Appendix E.6 show that they partially recover the instruction-following and reasoning abilities attenuated during alignment. Reasoning activation (S4) gives the largest gain over S3 (+4.45 w-F_1; +5.38 R@10), showing that activation improves direct ranking without test-time reasoning.

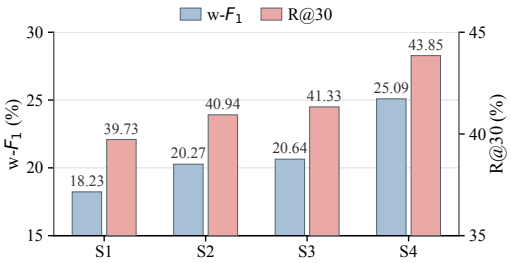

Figure 3: Effect of the four stages of reasoning-enriched training on MIMIC-III under non-thinking direct inference.

Effect of the coverage reward for CARing-Think. We compare four settings under the same thinking-mode inference protocol to separate the effects of visit-level RL and answer-level supervision. As shown in Figure 4, the reasoning-activation checkpoint (S4) attains 23.57 w-F_1 and 39.47 R@30. With EM reward (S5), every correct rollout receives equal credit, regardless of how often its SID appears in the group. Its scores fall to 18.88 and 34.67, respectively. This decline is consistent with repeated predictions of a few correct diagnoses, a pattern examined in RQ4. The comparison with S6 isolates how credit is distributed among correct rollouts: both rewards recognize a correct SID, but only the coverage reward discounts repeated predictions of that SID. With this reward, S6 raises w-F_1 to 25.41 and R@30 to 40.27, exceeding both S4 and S5. Adding the multi-positive loss (S7) further raises them to 27.58 and 46.04, gains of 2.17 and 5.77 points over S6. The larger gain in recall is consistent with the auxiliary loss supplying a learning signal for target diagnoses that the sampled trajectories may miss. Together, these results support the complementary roles of rewarding distinct sampled positives and supervising the full set of target diagnoses.

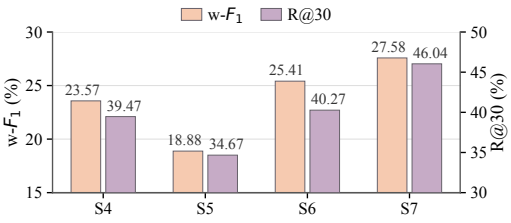

Figure 4: Training ablation on MIMIC-III in thinking mode.

4.4 Coverage and Diversity in Training (RQ4)

For each visit, we monitor 16 sampled trajectories. Unique SIDs count all distinct predictions, while unique correct SIDs count distinct next-visit diagnoses. We also measure mean inverse frequency $1/c_i$ over correct predictions, where c_i counts repetitions of the predicted SID within the group; larger values indicate less repetition.

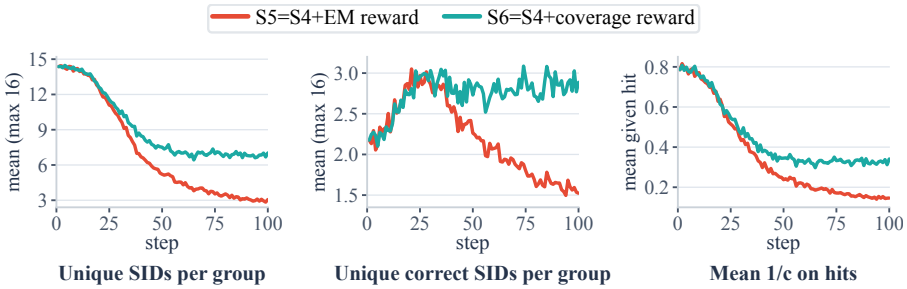

Figure 5: MIMIC-III trajectories for S5 (EM reward) and S6 (coverage reward), both initialized from S4.

Effect of the coverage reward. Under EM reward, unique correct SIDs rise from 2.18 at step 1 to 2.86 at step 30, then fall to 1.52 by step 100 (Figure 5). The coverage reward retains 2.89 unique correct SIDs at step 100 and produces 7.03 unique SIDs per group, versus 3.06 under EM. S6's higher mean inverse frequency indicates less repetition among correct predictions, consistent with S6's higher held-out scores in Figure 4.

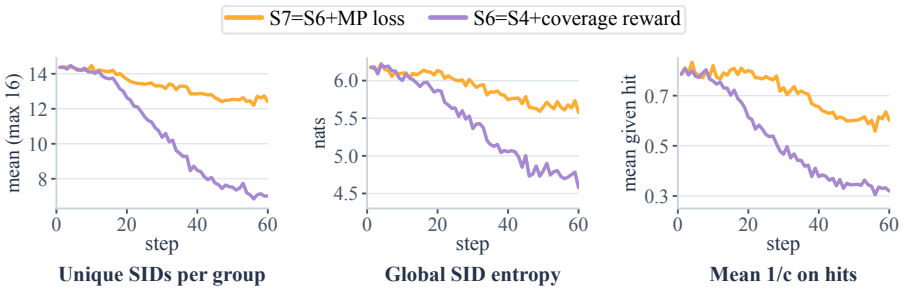

Figure 6: MIMIC-III trajectories for S6 and S7. The missing S7 value at step 42 is linearly interpolated between steps 41 and 43.

Effect of the multi-positive loss. The coverage reward can distinguish correct SIDs only after sampling. At step 60, S7 produces 12.40 unique SIDs per group, versus 7.02 for S6 (Figure 6). Batch-level SID entropy, $-\sum_s p(s)\log p(s)$, is 5.58 versus 4.58, and mean inverse frequency among correct predictions is 0.60 versus 0.32. The additional supervision thus reduces repetition while retaining more candidate diagnoses; S7 also improves held-out w-F_1 and R@30 over S6 (Figure 4).

5 Conclusion

Predicting diagnoses at the next visit requires a model to recover several valid diagnoses from a large vocabulary. Rewarding each reasoning trajectory only for correct predictions can favor repeated answers and leave other diagnoses uncovered. We introduced CARing, which represents diagnoses with compositional medical SIDs grounded in clinical language and patient histories. The model learns to reason before predicting a SID, while a coverage reward and supervision over multiple correct diagnoses encourage complementary answers across trajectories. It supports direct ranking and multiple reasoning chains. On MIMIC-III and MIMIC-IV, CARing reaches R@30 of 46.04% and 46.52%, respectively, has the highest recall at every reported cutoff, and exceeds all EHR-trained baselines in weighted F_1.

AI USE STATEMENT

We primarily utilized LLMs for paper polishing and programming assistance. Specifically, for paper writing, we employed LLMs (*e.g.,* GPT-5) to refine sentence fluency as well as provide suggestions on precise and appropriate expression. All the modifications generated by LLMs were carefully reviewed by the authors to ensure that the final content faithfully reflects the authors' intended ideas. For code generation, LLMs (*e.g.,* GitHub Copilot) were mainly used to generate some code snippets, which were subsequently adapted and integrated into the implementation framework by the authors. In both paper polishing and programming assistance, the authors were in full control, with LLMs serving solely as tools to improve efficiency.

REPRODUCIBILITY STATEMENT

For reproducibility, we describe the method and training objectives in Section 3 and Appendices C and D. Data preprocessing, implementation details, and experimental settings are provided in Appendix E. The code, training logs, and test logs are available at https://anonymous.4open.science/r/CARing-645B/README.md and in the supplementary materials.

REFERENCES

Agency for Healthcare Research and Quality. Chronic condition indicator (CCI) for ICD-9-CM. Healthcare Cost and Utilization Project (HCUP), 2015a. URL https://hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp.

Agency for Healthcare Research and Quality. Clinical classifications software (CCS) for ICD-9-CM. Healthcare Cost and Utilization Project (HCUP), 2015b. URL https://hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.

Oron Anschel, Alon Shoshan, Adam Botach, Shunit Haviv Hakimi, Asaf Gendler, Emanuel Ben Baruch, Nadav Bhonker, Igor Kviatkovsky, Manoj Aggarwal, and Gerard Medioni. Group-aware reinforcement learning for output diversity in large language models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 32394–32415, 2025. doi: 10.18653/v1/2025.emnlp-main.1649.

Tian Bai, Shanshan Zhang, Brian L Egleston, and Slobodan Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 43–51, 2018.

Ofir Ben Shoham and Nadav Rappoport. CPLLM: Clinical prediction with large language models. *PLOS Digital Health*, 3(12):e0000680, 2024.

Yushi Cao, Yiming Chen, Hongchao Jiang, Hung-yi Lee, and Robby T Tan. Remedi: Reasoner for medical clinical prediction. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 20232–20241, 2026.

Jiayuan Chen, Changchang Yin, Yuanlong Wang, and Ping Zhang. Predictive modeling with temporal graphical representation on electronic health records. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI 2024)*, pp. 5763–5771, 2024. doi: 10.24963/ijcai.2024/637.

Junying Chen, Zhenyang Cai, Ke Ji, Xidong Wang, Wanlong Liu, Rongsheng Wang, and Benyou Wang. Towards medical complex reasoning with LLMs through medical verifiable problems. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 14552–14573, 2025. doi: 10.18653/v1/2025.findings-acl.751.

Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine learning for healthcare conference*, pp. 301–318. PMLR, 2016a.

Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in neural information processing systems*, 29:3504–3512, 2016b.

Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. GRAM: Graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 787–795, 2017.

Gordon V Cormack, Charles LA Clarke, and Stefan Buettcher. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 758–759, 2009.

DeepSeek-AI. DeepSeek-V4: Towards highly efficient million-token context intelligence. *arXiv preprint arXiv:2606.19348*, 2026.

Yue Fang, Yuxin Guo, Jiaran Gao, Hongxin Ding, Xinke Jiang, Weibin Liao, Yongxin Xu, Yinghao Zhu, Zhibang Yang, Liantao Ma, Junfeng Zhao, and Yasha Wang. Toward better EHR reasoning in LLMs: Reinforcement learning with expert attention guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 30690–30698, 2026. doi: 10.1609/aaai.v40i36.40325.

Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M Glass, and Jimeng Sun. Stagenet: Stage-aware neural networks for health risk prediction. In *Proceedings of the web conference 2020*, pp. 530–540, 2020.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081):633–638, 2025. doi: 10.1038/s41586-025-09422-z.

Anthony GX-Chen, Jatin Prakash, Jeff Guo, Rob Fergus, and Rajesh Ranganath. KL-regularized reinforcement learning for generative modelling is designed to mode collapse. In *International Conference on Learning Representations*, volume 2026, pp. 50665–50693, 2026.

Xiaoke Huang, Juncheng Wu, Hui Liu, Xianfeng Tang, and Yuyin Zhou. m1: Unleash the potential of test-time scaling for medical reasoning with large language models. In *Proceedings of the Fifth Machine Learning for Health Symposium*, volume 297 of *Proceedings of Machine Learning Research*, pp. 369–383. PMLR, 2026.

Pengcheng Jiang, Cao Xiao, Adam Cross, and Jimeng Sun. GraphCare: Enhancing healthcare predictions with personalized knowledge graphs. In *International Conference on Learning Representations*, volume 2024, pp. 16839–16866, 2024.

Pengcheng Jiang, Cao Danica Xiao, Minhao Jiang, Parminder Bhatia, Taha Kass-Hout, Jimeng Sun, and Jiawei Han. Reasoning-enhanced healthcare predictions with knowledge graph community retrieval. In *International Conference on Learning Representations*, volume 2025, pp. 81785–81830, 2025.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):160035, 2016.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1, 2023.

Feifei Ke, Deli Hua, and Kaichao Yang. Hieramed: A hierarchical representation framework for next-visit diagnosis prediction from longitudinal electronic health records. *MethodsX*, 17:104087, 2026. ISSN 2215-0161. doi: https://doi.org/10.1016/j.mex.2026.104087. URL https://www.sciencedirect.com/science/article/pii/S2215016126003031.

Heejoon Koo. Next visit diagnosis prediction via medical code-centric multimodal contrastive EHR modelling with hierarchical regularisation. In *Findings of the Association for Computational Linguistics: EACL 2024*, pp. 41–55, 2024.

Zeljko Kraljevic, Dan Bean, Anthony Shek, Rebecca Bendayan, Harry Hemingway, Joshua Au Yeung, et al. Foresight—a generative pretrained transformer for modelling of patient timelines using electronic health records: a retrospective modelling study. *The Lancet Digital Health*, 6(4):e281–e290, 2024.

Chenyi Li, Yuan Zhang, Bo Wang, Guoqing Ma, Wei Tang, Haoyang Huang, and Nan Duan. SetPO: Set-level policy optimization for diversity-preserving LLM reasoning. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=tYKGzVE1g7.

Yikuan Li, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. BEHRT: Transformer for electronic health records. *Scientific Reports*, 10(1):7155, 2020.

Che Liu, Haozhe Wang, Jiazhen Pan, Zhongwei Wan, Yong Dai, Fangzhen Lin, Wenjia Bai, Daniel Rueckert, and Rossella Arcucci. Beyond distillation: Pushing the limits of medical LLM reasoning with minimalist rule-based RL. *arXiv preprint arXiv:2505.17952*, 2025.

Zhanyu Liu, Shiyao Wang, Xingmei Wang, Rongzhou Zhang, Jiaxin Deng, Honghui Bao, Jinghao Zhang, Wuchao Li, Pengfei Zheng, Xiangyu Wu, et al. OneRec-Think: In-text reasoning for generative recommendation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2664–2681, 2026. doi: 10.18653/v1/2026.acl-long.123.

Chang Lu, Chandan K Reddy, Prithwish Chakraborty, Samantha Kleinberg, and Yue Ning. Collaborative graph learning with auxiliary text for temporal event prediction in healthcare. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, pp. 3529–3535, 2021. doi: 10.24963/ijcai.2021/486.

Chang Lu, Tian Han, and Yue Ning. Context-aware health event prediction via transition functions on dynamic disease graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 4567–4574, 2022.

Junyu Luo, Muchao Ye, Cao Xiao, and Fenglong Ma. HiTANet: Hierarchical time-aware attention networks for risk prediction on electronic health records. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 647–656, 2020.

Hang Lv, Kaisong Zhang, Yanchao Tan, and Xing Chen. Towards efficient and interpretable medical concept representation via ontology-driven residual vector quantization. In *Proceedings of the ACM Web Conference 2026*, pp. 8645–8648, 2026.

Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1903–1911, 2017.

Fenglong Ma, Quanzeng You, Houping Xiao, Radha Chitta, Jing Zhou, and Jing Gao. Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pp. 743–752, 2018.

Mingyu Derek Ma, Xiaoxuan Wang, Yijia Xiao, Anthony Cuturrufo, Vijay S Nori, Eran Halperin, and Wei Wang. Memorize and rank: Elevating large language models for clinical diagnosis prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 24786–24794, 2025.

Shashank Rajput, Nikhil Mehta, Anima Singh, Raghunandan Hulikal Keshavan, Trung Vu, Lukasz Heldt, Lichan Hong, Yi Tay, Vinh Tran, Jonah Samost, Maciej Kula, Ed Chi, and Maheswaran Sathiamoorthy. Recommender systems with generative retrieval. In *Advances in Neural Information Processing Systems*, volume 36, pp. 10299–10315, 2023.

Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digital Medicine*, 4(1):86, 2021.

Tom Röhr, Thomas Maximilian Josef Steffek, Roman Teucher, Keno Bressem, Alexei Figueroa, Paul Grundmann, Peter Troeger, Felix Alexander Gers, and Alexander Löser. DeepICD-R1: Medical reasoning through hierarchical rewards and unsupervised distillation. In *Proceedings of the Fifteenth Language Resources and Evaluation Conference*, pp. 10764–10775. ELRA Language Resource Association, 2026. doi: 10.63317/5ntn3gnmv9cy. URL https://aclanthology.org/2026.lrec-1.843/.

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

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.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R Pfohl, Heather Cole-Lewis, et al. Toward expert-level medical question answering with large language models. *Nature Medicine*, 31(3):943–950, 2025.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling parameters for reasoning. In *International Conference on Learning Representations*, volume 2025, pp. 10131–10165, 2025.

Ali Soroush, Benjamin S Glicksberg, Eyal Zimlichman, Yiftach Barash, Robert Freeman, Alexander W Charney, Girish N Nadkarni, and Eyal Klang. Large language models are poor medical coders—benchmarking of medical code querying. *NEJM AI*, 1(5):AIdbp2300040, 2024.

Xiaorui Su, Shvat Messica, Yepeng Huang, Ruth Johnson, Lukas Fesser, Shanghua Gao, Faryad Sahneh, and Marinka Zitnik. Multimodal medical code tokenizer. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 57030–57045. PMLR, 2025.

Yanchao Tan, Hang Lv, Yunfei Zhan, Guofang Ma, Bo Xiong, and Carl Yang. Boxlm: Unifying structures and semantics of medical concepts for diagnosis prediction in healthcare. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 58603–58620. PMLR, 2025.

OneRec Team, Biao Yang, Boyang Ding, Chenglong Chu, Dunju Zang, Fei Pan, Han Li, Hao Jiang, Honghui Bao, Huanjie Wang, et al. Onereason technical report. *arXiv preprint arXiv:2606.06260*, 2026.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*, 2023.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022.

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.

Chaoqi Yang, Zhenbang Wu, Patrick Jiang, Zhen Lin, Junyi Gao, Benjamin P Danek, and Jimeng Sun. Pyhealth: A deep learning toolkit for healthcare applications. In *KDD'23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5788–5789. ACM Association for Computing Machinery, 2023a.

Pengcheng Yang, Fuli Luo, Shuming Ma, Junyang Lin, and Xu Sun. A deep reinforced sequence-to-set model for multi-label classification. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pp. 5252–5258, 2019.

Zhichao Yang, Avijit Mitra, Weisong Liu, Dan Berlowitz, and Hong Yu. Transformehr: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records. *Nature communications*, 14(1):7857, 2023b.

Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Yang Yue, Shiji Song, and Gao Huang. Does reinforcement learning really incentivize reasoning capacity in LLMs beyond the base model? In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Neil Zeghidour, Alejandro Luebs, Ahmed Omran, Jan Skoglund, and Marco Tagliasacchi. Soundstream: An end-to-end neural audio codec. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:495–507, 2022. doi: 10.1109/TASLP.2021.3129994.

Bowen Zheng, Yupeng Hou, Hongyu Lu, Yu Chen, Wayne Xin Zhao, Ming Chen, and Ji-Rong Wen. Adapting large language models by integrating collaborative semantics for recommendation. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pp. 1435–1448, 2024.

A Semantic Structure of Medical SIDs

A.1 Randomly Sampled First-Token Cases

We sampled four distinct first SID tokens uniformly from the 256 used by the name plus ICD path variant, using a fixed seed of 2024. Figure 7 shows how their diseases are distributed across ICD ontology subgroups. Each count represents one distinct disease.

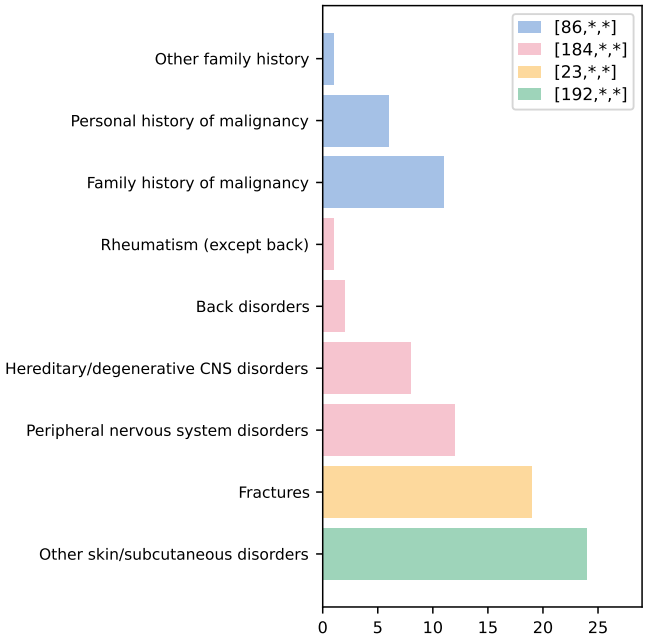

Figure 7: ICD ontology subgroup counts for four randomly sampled first SID tokens. Subgroups are the second nodes below the ontology root.

The sampled prefixes have recognizable clinical themes. Of the 18 diseases beginning with $\langle a_{86} \rangle$, 17 describe a personal or family history of malignancy. The $\langle a_{184} \rangle$ prefix contains 20 nervous system diagnoses among 23 diseases, mainly peripheral and hereditary or degenerative disorders. All 19 diseases under $\langle a_{23} \rangle$ are spinal cord injuries, and all 24 under $\langle a_{192} \rangle$ are chronic skin ulcers. The three musculoskeletal diagnoses under $\langle a_{184} \rangle$ also illustrate that a first token can connect related conditions across ontology branches.

A.2 Semantic Organization Across the Disease Vocabulary

We measured how well a shared first token groups the 4,491 distinct diseases by their top-level ICD ontology group and single-level CCS category. The vocabulary spans 52 top-level groups and 264 CCS categories. For each token, majority purity is the fraction of its diseases in the most common category. Weighted purity averages this fraction over diseases. Pair agreement is the probability that two diseases with the same first token also have the same category. Both SID variants use the same disease labels.

Table 3: Semantic organization of first SID tokens over 4,491 distinct diseases. ICD groups are the first nodes below the ontology root.

Metric	Name only	Name + ICD path
First-token groups pure by ICD / 256	58	**143**
First-token groups $\geq 80\%$ pure by ICD / 256	108	**212**
Top-level ICD weighted purity (%)	69.36	**90.51**
Same ICD group given shared token (%)	57.60	**86.49**
First-token groups pure by CCS / 256	13	**44**
First-token groups $\geq 80\%$ pure by CCS / 256	36	**103**
CCS weighted purity (%)	47.63	**69.78**
Same CCS category given shared token (%)	32.74	**57.39**

Table 3 shows stronger semantic organization for SIDs learned from names plus ICD paths. The share of diseases assigned to their token's dominant top-level group rises from 69.36% to 90.51%, and 143 rather than 58 first-token groups contain a single top-level group. The same pattern holds at the finer CCS level: diseases sharing a first token have the same CCS category with probability 57.39% for names plus paths, compared with 32.74% for names alone. These results extend the sampled cases in Figure 7 to the full disease vocabulary.

B Extended Related Work

We expand the discussion in Section 2, following the progression from clinical prediction and medical representations to reasoning models and the optimization of diagnosis coverage.

B.1 Longitudinal Diagnosis Prediction

Longitudinal EHR models learn how diagnoses and other clinical events evolve across visits. Doctor AI and Dipole use recurrent networks to model these sequences, while RETAIN uses attention to identify influential visits and clinical variables (Choi et al., 2016a; Ma et al., 2017; Choi et al., 2016b). HiTANet accounts for irregular time intervals, and StageNet models changes in patient status (Luo et al., 2020; Gao et al., 2020).

Medical knowledge and relations between patients provide additional structure. GRAM and KAME incorporate medical ontologies into the representations used for prediction (Choi et al., 2017; Ma et al., 2018). CGL learns from patient and disease graphs, while subsequent methods model transitions on dynamic disease graphs or encode EHRs as temporal heterogeneous graphs (Lu et al., 2021; 2022; Chen et al., 2024). NECHO combines medical codes, demographic information, and clinical notes through contrastive learning and hierarchical regularization (Koo, 2024). BoxLM combines the textual semantics and hierarchical structure of medical concepts using box embeddings (Tan et al., 2025). HieraMed uses embeddings of medical code descriptions within a hierarchy of code, visit, and patient representations (Ke et al., 2026).

Pretraining on patient records offers another route to clinical prediction. BEHRT and Med-BERT learn contextual representations of structured EHRs that can be adapted to disease prediction (Li et al., 2020; Rasmy et al., 2021). Generative approaches also exist: TransformEHR predicts future medical codes with a Transformer encoder and decoder, and Foresight models patient timelines through autoregressive generation (Yang et al., 2023b; Kraljevic et al., 2024). CPLLM adapts pretrained LLMs to clinical prediction by expressing patient histories as text, with evaluations including specific disease prediction and readmission (Ben Shoham & Rappoport, 2024). These settings differ

in the information available to the model and the outputs being evaluated. Predicting a selected outcome, generating future clinical concepts, and recovering the full set of diagnoses at the next visit place different demands on the prediction model.

B.2 Medical Code Representations and Generation

Medical codes identify standardized concepts, but their character sequences do not express disease meanings in ordinary language. Evaluations of LLMs on medical code queries report errors in mapping descriptions to exact codes (Soroush et al., 2024). This motivates representations that connect code identities to their descriptions and clinical relations. MedTok encodes textual descriptions and relational context, then quantizes these representations into a shared token space (Su et al., 2025). Its evaluations include both EHR predictors and medical question answering, where the learned tokens are supplied as prefixes to LLMs. MedRQ uses residual quantization and supervision from medical ontologies to learn discrete medical representations (Lv et al., 2026).

MERA incorporates disease identities directly into a language model's vocabulary, with a dedicated token for each disease (Ma et al., 2025). It learns mappings between disease tokens and definitions, varies the order of diagnoses within a visit, and uses the medical hierarchy in candidate ranking. This treatment accounts for the unordered structure of visit diagnoses, although a dedicated token for each disease requires the added vocabulary to grow with the diagnosis catalog.

Our compositional representation draws on semantic IDs in generative recommendation, including systems that generate a natural language rationale before an identifier (Rajput et al., 2023; Liu et al., 2026; Team et al., 2026). We adapt this representation to diseases by quantizing their descriptions and ICD ontology paths into sequences of shared codebook tokens. Alignment with clinical text and patient histories makes these SIDs usable as both inputs and diagnosis outputs. The resulting interface allows a single autoregressive model to generate a clinical rationale followed by a diagnosis SID.

B.3 Reasoning for Clinical Prediction

Medical question answering has provided a setting for adapting and evaluating LLM reasoning. Med-PaLM and Med-PaLM 2 study the use of LLMs for answering medical questions (Singhal et al., 2023; 2025). HuatuoGPT-o1 trains on medical problems with verifiable answers, while AlphaMed studies reinforcement learning from answer correctness without an initial stage of supervised training on reasoning traces (Chen et al., 2025; Liu et al., 2025). The m1 study examines computation at inference time for medical reasoning and finds that longer reasoning alone does not consistently improve performance; medical knowledge and the reasoning budget also matter (Huang et al., 2026).

For prediction from patient records, knowledge retrieval and reasoning supervision help connect clinical evidence to target outcomes. GraphCare constructs patient-specific knowledge graphs for healthcare prediction (Jiang et al., 2024). KARE retrieves relevant knowledge graph communities and distills reasoning for mortality and readmission prediction (Jiang et al., 2025). ReMedi uses known outcomes to guide the construction and regeneration of rationale–answer pairs for supervised and preference tuning (Cao et al., 2026). EAG-RL initializes the policy with reasoning trajectories guided by specialized EHR models. It then uses their attention patterns as an additional reward during reinforcement learning (Fang et al., 2026).

Reasoning has also been applied to medical coding. DeepICD-R1 combines distilled reasoning traces with rewards at several levels of the ICD hierarchy to predict codes from admission notes (Röhr et al., 2026). Coding the conditions described in a clinical document and forecasting a future visit use different evidence and supervision. Our setting requires reasoning from earlier visits to recover multiple diagnoses in the next visit. We therefore evaluate diagnosis coverage across sampled trajectories, in addition to whether an individual trajectory predicts a correct diagnosis.

B.4 Large Reasoning Models and Policy Collapse

Chain-of-thought prompting elicits intermediate reasoning steps before an answer (Wei et al., 2022). Self-consistency samples several reasoning paths and aggregates their answers (Wang et al., 2023),

while work on scaling computation at inference time studies how to allocate search and refinement to different problems (Snell et al., 2025). Reinforcement learning with verifiable rewards trains reasoning policies from feedback on their outputs. DeepSeekMath introduces GRPO, which compares sampled trajectories within a group, and DeepSeek-R1 applies reinforcement learning to develop reasoning capabilities (Shao et al., 2024; Guo et al., 2025).

Higher accuracy for an individual sample need not imply broader coverage across samples. In the settings evaluated by Yue et al. (2025), RLVR improves pass@1, but base models can attain higher pass@k when k is large. GX-Chen et al. (2026) analyze how rewards, the reference policy, and KL regularization determine the target distribution. Under conditions including weak regularization and equal verifiable rewards, that distribution can concentrate on a single region of high reward. These findings motivate examining both correctness and coverage when several outputs can be valid. For diagnosis prediction, repeated success on one diagnosis can coexist with poor recall of the other diagnoses recorded at the same visit.

B.5 Correctness and Diversity in Reinforcement Learning

Optimizing the quality of a group of outputs requires assigning credit for what each sample contributes to that group. GAPO adjusts rewards using the frequency of valid outputs within a rollout group, giving more credit to underrepresented valid outputs (Anschel et al., 2025). SetPO measures diversity through similarities between trajectories and uses each trajectory's marginal contribution to group diversity in policy optimization (Li et al., 2026). These methods connect the training signal for one trajectory to the other trajectories sampled for the same input.

Rewards defined over sets also appear in multi-label classification. Yang et al. (2019) use reinforcement learning with a reward independent of label order to reduce the sensitivity of sequence generation to arbitrary label orderings. Their model generates a label set within one output sequence. In our formulation, each reasoning trajectory ends with one diagnosis, and the rollout group is evaluated against the complete set of diagnoses at the next visit.

Our objective measures diversity in terms of distinct correct diagnosis labels. The coverage reward divides a correct prediction's credit by the number of times that diagnosis appears in the group. Consequently, each distinct correct diagnosis contributes one unit to the raw group reward, as shown in Appendix D. This gives the frequency adjustment a direct interpretation as diagnosis coverage. An auxiliary multi-positive loss supervises the SID answer tokens to provide training signal for correct diagnoses absent from a sampled group. Together, the two terms address repetition among sampled positives and the limited exposure to positives that sampling can miss.

C Task Formats

This appendix provides representative input–output examples for the task formats used throughout training. We first present the six tasks used for multi-task SID–language alignment. The examples below follow the prompts in the implementation; only the assistant response is supervised.

C.1 Multi-Task SID–Language Alignment

Tasks T1–T4 align textual or SID-based longitudinal histories with disease-name or SID prediction. Tasks T5–T6 establish bidirectional grounding between disease names and compositional SIDs. Diagnoses within each visit are unordered, whereas visits are listed chronologically.

Let $\mathcal{D}_\tau$ denote the training set for task $\tau \in \{1, \ldots, 6\}$ and let $\mathcal{A}(y)$ contain the assistant-response token positions. All tasks are jointly optimized under the standard next-token prediction objective.

$$\mathcal{L}_{\mathrm{MT}} = -\frac{1}{\sum_\tau |\mathcal{D}_\tau|} \sum_{\tau=1}^{6} \sum_{(x,y)\in\mathcal{D}_\tau} \sum_{j\in\mathcal{A}(y)} \log p_\theta(y_j \mid x, y_{<j}). \tag{7}$$

Task T1: SID History to SID

System:
Predict one possible diagnosis SID in the next visit from the chronological SID history.
User:
The patient's chronological diagnosis history is:
<visit_start>
<a_117><b_59><c_117>, <a_199><b_76><c_251>
<visit_end>
<visit_start>
<a_117><b_0><c_82>, <a_87><b_184><c_39>
<visit_end>
Return exactly one disease SID.
Assistant:
<a_117><b_0><c_82>

Task T2: Textual History to SID

System:
Predict one possible diagnosis SID in the next visit from the chronological diagnosis history.
User:
The patient's chronological diagnosis history is:
Visit 1: Subarachnoid hemorrhage; Pneumonitis due to inhalation of food or vomitus
Visit 2: Cerebral aneurysm, nonruptured; Unspecified essential hypertension
Return exactly one disease SID.
Assistant:
<a_117><b_0><c_82>

Task T3: SID History to Disease Name

System:
Predict one possible diagnosis in the next visit from the chronological SID history.
User:
The patient's chronological diagnosis history is:
<visit_start>
<a_117><b_59><c_117>, <a_199><b_76><c_251>
<visit_end>
<visit_start>
<a_117><b_0><c_82>, <a_87><b_184><c_39>
<visit_end>
Return one possible disease description for the next visit.
Assistant:
Cerebral aneurysm, nonruptured

Task T4: Textual History to Disease Name

System:
Predict one possible diagnosis in the next visit from the chronological diagnosis history.
User:
The patient's chronological diagnosis history is:
Visit 1: Subarachnoid hemorrhage; Pneumonitis due to inhalation of food or vomitus
Visit 2: Cerebral aneurysm, nonruptured; Unspecified essential hypertension
Return one possible disease description for the next visit.
Assistant:
Cerebral aneurysm, nonruptured

Task T5: Disease Name to SID

System:
You map ICD disease names to semantic identifiers.
User:
Which disease has the name “Cerebral aneurysm, nonruptured”?
Assistant:
<a_117><b_0><c_82>

Task T6: SID to Disease Name

System:
You map semantic identifiers to ICD disease names.
User:
What disease does <a_117><b_0><c_82> represent?
Assistant:
Cerebral aneurysm, nonruptured

C.2 Reasoning-Enriched Training

The disease-level and case-level enriched corpora are trained as plain-text next-token prediction sequences, without a system or user prompt. Reasoning activation instead uses the chat-formatted history-only prompt shown in the third example; both its reasoning trace and final SID are supervised.

Training Case R1: Disease-Level Semantic Enrichment

Training text:
<a_117><b_0><c_82> is a clinically significant vascular abnormality characterized by localized ballooning of an intracranial artery that has not yet ruptured. The CCI body-system classification places this entity within Diseases of the circulatory system, and the CCI chronicity classification designates it as chronic, reflecting its persistent nature and need for long-term surveillance. According to the CCS single-level category, <a_117><b_0><c_82> falls under Other and ill-defined cerebrovascular disease, and the complete CCS multi-level hierarchy is Diseases of the circulatory system > Cerebrovascular disease > Other and ill-defined cerebrovascular disease [111.]. Within the ICD ontology hierarchy, it is situated as DISEASES AND INJURIES > DISEASES OF THE CIRCULATORY SYSTEM > CEREBROVASCULAR DISEASE > Other and ill-defined cerebrovascular disease > <a_117><b_0><c_82>. This classification structure is coding information and should not be interpreted as evidence of a causal pathway between the categories and the lesion. Clinically, <a_117><b_0><c_82> often remains asymptomatic and may be discovered incidentally during brain imaging for unrelated complaints. However, the most feared complication is rupture, leading to subarachnoid hemorrhage, which carries high morbidity and mortality. Established risk factors for the development and growth of <a_117><b_0><c_82> include hypertension, cigarette smoking, alcohol misuse, and a family history of intracranial aneurysms. Certain genetic conditions, such as autosomal dominant polycystic kidney disease and Ehlers-Danlos syndrome, also predispose patients to this vascular lesion. Common comorbidities among patients harboring <a_117><b_0><c_82> may include atherosclerotic cardiovascular disease, prior ischemic stroke, or other manifestations of systemic vascular injury. Management strategies focus on preventing rupture, balancing the risks of microsurgical clipping or endovascular coiling against the natural history of the unruptured lesion. Regular imaging follow-up is often recommended for small stable lesions, while larger or symptomatic aneurysms may warrant intervention. Thus, <a_117><b_0><c_82> represents a chronic circulatory condition that demands individualized risk assessment and careful clinical surveillance.

Training Case R2: Case-Level Semantic Enrichment

Training text:
Reviewing the longitudinal record, the picture centers on chronic cerebrovascular disease. At the initial visit, <a_117><b_0><c_82> appeared alongside <a_117><b_59><c_117>, with associated symptoms <a_138><b_32><c_38> and <a_93><b_60><c_3>; <a_179><b_208><c_18> was also present but seemed less tied to this axis. By the most recent visit, recurrence of <a_117><b_59><c_117> together with <a_207><b_3><c_118> and <a_38><b_213><c_64> points to an active cerebrovascular process that required intervention and now device-related follow-up. In this context, the chronic <a_117><b_0><c_82> from the earliest encounter stands out as the likely underlying lesion being monitored, rather than an incidental finding, since it belongs to the same circulatory/cerebrovascular domain. The early presence of <a_179><b_208><c_18> contributed to the initial burden but does not share this close relationship. Going forward, continued cerebrovascular surveillance and management of potential complications, along with related neurological and vascular diagnoses, would be expected.

Training Case R3: Reasoning Activation

System:
Predict one possible diagnosis SID in the next visit from the chronological SID history. Think through the longitudinal clinical patterns before answering.
User:
The patient's chronological diagnosis history is:
<visit_start>
<a_117><b_0><c_82>, <a_138><b_32><c_38>, <a_117><b_59><c_117>,
<a_179><b_208><c_18>, <a_93><b_60><c_3>
<visit_end>
<visit_start>
<a_117><b_59><c_117>, <a_207><b_3><c_118>, <a_38><b_213><c_64>
<visit_end>
Think step by step about the longitudinal clinical patterns, then return exactly one disease SID.
Assistant:
<think>
I see a core longitudinal pattern centered on chronic cerebrovascular disease. The earliest visit already contained <a_117><b_0><c_82>, a chronic circulatory-system diagnosis, alongside <a_117><b_59><c_117> and symptoms such as <a_138><b_32><c_38> and <a_93><b_60><c_3>. The recurrence of <a_117><b_59><c_117> in the most recent visit, together with <a_207><b_3><c_118> and <a_38><b_213><c_64>, strongly suggests an active cerebrovascular process that required intervention and generated device-related follow-up. The persistence of the chronic SID <a_117><b_0><c_82> from the first visit may remain clinically relevant because it belongs to the same circulatory/cerebrovascular domain and could represent the underlying lesion being monitored. The non-chronic <a_179><b_208><c_18> from the first visit seems less tied to this axis, though it contributed to the initial overall burden. Given the recurrence and the recent device/procedure encounters, subsequent visits may continue to involve cerebrovascular surveillance, complication management, or related neurological and vascular diagnoses.
</think>
<a_117><b_0><c_82>

C.2.1 Training Objectives

Let $\mathcal{X}_{\text{disease}}$ and $\mathcal{X}_{\text{case}}$ denote the disease-level and case-level enriched corpora. Each sample $x = (x_1, \ldots, x_{|x|})$ is a plain-text sequence, with $p_\theta(x) = \prod_{j=1}^{|x|} p_\theta(x_j \mid x_{<j})$. The two corpora use next-token prediction objectives:

$$\begin{aligned} \mathcal{L}_{\text{disease}} &= -\mathbb{E}_{x\sim\mathcal{X}_{\text{disease}}}\left[\log p_\theta(x)\right], \\ \mathcal{L}_{\text{case}} &= -\mathbb{E}_{x\sim\mathcal{X}_{\text{case}}}\left[\log p_\theta(x)\right]. \end{aligned} \tag{8}$$

For reasoning activation, each training example pairs a historical SID sequence H with a teacher-generated reasoning path r and a target diagnosis d. The prompt contains only H, and the loss applies to the assistant response:

$$\mathcal{L}_{\text{act}} = -\mathbb{E}_{(H,r,d)}[\log p_\theta(\texttt{<think>}r\texttt{</think>}\, q(d) \mid H)]. \tag{9}$$

C.3 PROMPTS FOR REASONING-ENRICHED DATA

We generate the three supervision signals in two routes. Disease-level semantic enrichment directly converts the structured dossier of one coded concept into a SID-grounded clinical narrative. For case-level enrichment, we first elicit a history-supported reasoning path and then provide that path, together with the same history and the observed diagnosis, to a separate integration prompt. The resulting reasoning path is also reused as the reasoning supervision in the activation data. The prompts below omit only the mechanics of model invocation.

Prompt P1: Disease-Level Narrative Generation

User prompt (template):
You are a senior clinical medicine expert preparing an in-depth narrative for a coded clinical concept dossier.

Source information:
– Concept name, provided only for understanding and medical knowledge retrieval; never reproduce it in the output: [CONCEPT NAME]
– Semantic Identifier: [SID]
– CCI body-system classification: [CCI BODY SYSTEM]
– CCI chronicity classification: [CCI CHRONICITY]
– CCS single-level category: [CCS SINGLE-LEVEL LABEL]
– CCS optional category: [CCS OPTIONAL LABEL]
– CCS multi-level hierarchy: [COMPLETE CCS MULTI-LEVEL PATH]
– ICD ontology hierarchy: [COMPLETE ICD ONTOLOGY PATH]

Task:
1. Write a rich, coherent clinical narrative of at least ten sentences about the coded clinical concept represented by [SID].
2. Incorporate every available structured fact above. Explicitly describe the CCI body-system classification, CCI chronicity classification, CCS single-level category, complete CCS multi-level hierarchy, and complete ICD ontology hierarchy.
3. Treat the CCI, CCS, and ICD fields as coding and classification information. Do not present a classification relationship as proof of clinical causation.
4. Enrich the dossier using only widely accepted, high-confidence clinical knowledge relevant to the supplied concept name. When applicable, discuss a concise clinical description, established causes or risk factors, commonly associated complications or downstream clinical consequences, relevant comorbidities or prior diagnoses, and the broader clinical context in which the concept may be recorded.
5. If CCI chronicity or body-system information is unavailable, state only that the concept is not classified for that field. Do not infer a missing classification.
6. Adapt the narrative to the type of coded concept. If it represents an external cause, status, symptom, finding, or supplementary classification rather than a disease, do not describe it as a disease.

Entity-reference requirements:
7. Begin the paragraph with [SID].
8. Every direct reference to this specific clinical concept must use exactly [SID]. Never output the supplied concept name, ICD code, disease ID, abbreviation, synonym, or another alias.
9. Official CCS, CCI, and ICD category labels supplied above may appear only when explicitly describing those classification categories.
10. Use [SID] naturally at least four times. Do not output any other Semantic Identifier.

Output requirements:
11. Output only one flowing English paragraph. Do not use headings, bullet points, enumerations, JSON, citations, meta-commentary, or `<think>` tags.
12. Maintain a professional and medically precise tone suitable for a clinical concept dossier.

The concrete output obtained by instantiating this template with `<a_117><b_0><c_82>` is shown in Training Case R1.

Prompt P2: History-Supported Reasoning-Path Generation

User prompt (template):
You are an experienced clinical diagnostician reviewing a patient's longitudinal diagnosis history.

Your task is to reason through the patient's historical diagnoses and identify the longitudinal clinical patterns that would be most useful for a downstream prediction of one possible diagnosis in the next

Prompt P2: History-Supported Reasoning-Path Generation (continued)

visit. The observed next diagnosis is provided only as internal guidance for deciding which historical patterns may be relevant. Never mention or discuss that diagnosis.

Patient diagnosis history:
Visits are listed in chronological order, from oldest to most recent. Diagnoses within the same visit are an unordered set; never infer temporal order among them. For every diagnosis, the input supplies its complete SID, diagnosis name, CCI body system, CCI chronicity, CCS single-level category, and CCS multi-level path.
[HISTORICAL VISITS WITH PER-DIAGNOSIS METADATA]

Observed next diagnosis—internal guidance only; do not reveal:
[NEXT-DIAGNOSIS SID, NAME, CCI FIELDS, AND CCS FIELDS]

SID atomicity rule: Every SID is one atomic identifier consisting of exactly three adjacent components, `<a_NUMBER><b_NUMBER><c_NUMBER>`. Copy the complete SID exactly; never abbreviate it to one or two components.

Output requirements:
1. Output only a concise but detailed first-person reasoning monologue as one paragraph. Start directly with the reasoning.
2. Compare diagnoses across visits, while treating diagnoses within one visit as unordered and potentially co-occurring.
3. Give the greatest immediate predictive weight to recent visits. Also consider diagnoses that persist or recur across visits, especially chronic diagnoses. For a single-visit history, discuss only the observed disease burden.
4. Body systems, chronicity, CCS categories, and broad clinical relationships may be used as attributes of SID-based diagnoses, but never as alternative names for a concrete diagnosis.
5. Whenever referring to a concrete diagnosis, use only its complete SID. Never attach a diagnosis name, abbreviation, synonym, clinical gloss, or explanatory label to a SID.
6. Mention only SIDs present in the history. Never introduce an unseen SID, and do not output the observed next-diagnosis SID unless it already occurs in the history.
7. Never mention "observed diagnosis," "reference diagnosis," "target diagnosis," or equivalent language. Reason as if predicting blindly.
8. Do not claim that chronology or within-visit co-occurrence proves causation. Use appropriately cautious language such as "may," "could," or "is consistent with."
9. Adapt the depth to history length: emphasize burden and co-occurrence for short histories, and persistence, recurrence, newly appearing conditions, and recent shifts for longer histories.
10. Do not select or name a specific next diagnosis. End with a non-deterministic summary.

Applied to the two-visit history in Training Case R3, this prompt produces the paragraph enclosed by `<think>` and `</think>` in that example. That paragraph is retained as activation supervision and passed to Prompt P3 as the case-level reasoning path.

Prompt P3: Case-Level Narrative Integration

User prompt (instantiated example):
Integrate the following information into a single, coherent, natural narrative paragraph.

Patient diagnosis history:
Visits are listed in chronological order, from oldest to most recent. Diagnoses within the same visit are an unordered set.
Historical visit 1—oldest:
– `<a_117><b_0><c_82>`: Cerebral aneurysm, nonruptured
– `<a_138><b_32><c_38>`: Headache
– `<a_117><b_59><c_117>`: Subarachnoid hemorrhage
– `<a_179><b_208><c_18>`: Iron deficiency anemia, unspecified
– `<a_93><b_60><c_3>`: Visual discomfort
Historical visit 2—most recent:
– `<a_117><b_59><c_117>`: Subarachnoid hemorrhage
– `<a_207><b_3><c_118>`: Other complications due to nervous system device, implant, and graft
– `<a_38><b_213><c_64>`: Other specified procedures as the cause of abnormal reaction or later complication

Observed next diagnosis to integrate:
– SID: `<a_117><b_0><c_82>`
– Diagnosis name: Cerebral aneurysm, nonruptured

Prompt P3: Case-Level Narrative Integration (continued)

Reasoning path:
I see a core longitudinal pattern centered on chronic cerebrovascular disease. The earliest visit already contained <a_117><b_0><c_82>, a chronic circulatory-system diagnosis, alongside <a_117><b_59><c_117> and symptoms such as <a_138><b_32><c_38> and <a_93><b_60><c_3>. The recurrence of <a_117><b_59><c_117> in the most recent visit, together with <a_207><b_3><c_118> and <a_38><b_213><c_64>, strongly suggests an active cerebrovascular process that required intervention and generated device-related follow-up. The persistence of the chronic SID <a_117><b_0><c_82> from the first visit may remain clinically relevant because it belongs to the same circulatory/cerebrovascular domain and could represent the underlying lesion being monitored. The non-chronic <a_179><b_208><c_18> from the first visit seems less tied to this axis, though it contributed to the initial overall burden. Given the recurrence and the recent device/procedure encounters, subsequent visits may continue to involve cerebrovascular surveillance, complication management, or related neurological and vascular diagnoses.

Output requirements:
1. Write one natural, flowing longitudinal clinical narrative. Avoid mechanical or formulaic language.
2. Preserve the essential analytical insights and clinical logic from the reasoning path.
3. Connect the historical pattern to the observed SID as an ordinary clinical finding. Never use meta-language such as "reference," "target," "ground truth," or "next diagnosis," and never imply that it was known in advance.
4. Refer to every concrete diagnosis only by its complete SID of the form <a_XXX><b_YYY><c_ZZZ>. Do not output diagnosis names, codes, abbreviations, synonyms, or parenthetical clinical glosses. Use only SIDs appearing in the supplied history or observed diagnosis.
5. Keep the narrative clinically precise, cohesive, and engaging. Output only the paragraph.

The narrative produced for this instantiated prompt is shown in Training Case R2.

D Coverage Reward Property

For a group of N trajectories, let Y be the ground-truth SID set, y_i the SID predicted by trajectory i, and $r_i = \mathbf{1}[y_i \in Y]/c_i$, where $c_i = \sum_{j=1}^{N} \mathbf{1}[y_j = y_i]$. The total reward within a group equals its number of distinct correct predictions:

$$\sum_{i=1}^{N} r_i = |\{y_i : y_i \in Y,\ 1 \le i \le N\}|\,. \tag{10}$$

To see this, group correct rollouts by disease. If a correct SID appears c times, its total contribution is $c(1/c) = 1$; summing over distinct correct SIDs gives Equation 10. Incorrect predictions contribute zero regardless of their frequency. The reward therefore increases when the group covers another correct SID, but not when it repeats a correct SID or generates a new incorrect one.

Equation 10 describes the raw group reward. GRPO normalizes rewards within each group before updating the policy. When rewards vary, trajectories predicting less-repeated correct SIDs receive larger normalized advantages than duplicate or incorrect trajectories. If all rollouts have the same reward, their normalized advantages are zero: an all-identical correct group receives no policy-gradient reinforcement, while an all-distinct correct group has already attained the maximum coverage possible with N rollouts. Finite sampling and advantage normalization do not guarantee a globally uniform policy over all ground-truth diagnoses.

E Additional Experimental Details

E.1 Dataset Preprocessing

We process MIMIC-III and MIMIC-IV following prior diagnosis-prediction studies (Ma et al., 2018; Chen et al., 2024). For both datasets, we retain patients with at least two visits and arrange each patient's visits chronologically. Given the diagnoses observed in the historical visits, the task is to predict the complete diagnosis set recorded at the next visit. We split each dataset by patient into training, validation, and test sets with a ratio of 7:1:2, ensuring that all visits from the same patient occur in only one split.

Table 4: Dataset statistics after cohort preprocessing and before subsampling MIMIC-IV.

Dataset	MIMIC-III	MIMIC-IV
# of patients	5,449	79,393
# of visits	14,141	408,990
Avg. # of visits per patient	2.60	5.15
Max. # of visits per patient	29	170
# of unique diagnoses	3,874	37,917
# of CCS codes	285	842

Because MIMIC-IV is substantially larger than MIMIC-III, we construct the experimental MIMIC-IV cohort by sampling 10,000 patients from the patient-level partitions while preserving the 7:1:2 ratio: 7,000 patients from the training split, 1,000 from the validation split, and 2,000 from the test split. Table 4 reports the cohort statistics after the common preprocessing procedure and before this MIMIC-IV subsampling.

E.2 Baselines

We compare CARing with representative baselines spanning conventional neural predictors, language-assisted predictors, and a zero-shot large language model:

- **RETAIN** (Choi et al., 2016b) uses two reverse-time recurrent attention mechanisms to assign visit-level and variable-level importance weights for accurate and interpretable prediction from longitudinal EHRs.
- **KAME** (Ma et al., 2018) augments a recurrent patient-history encoder with medical knowledge-graph embeddings and a knowledge-attention mechanism over relevant ontology concepts for next-visit diagnosis prediction.
- **StageNet** (Gao et al., 2020) combines a stage-aware LSTM with stage-adaptive convolution to identify latent health-stage transitions and capture progression patterns for risk prediction.
- **CGL** (Lu et al., 2021) jointly learns patient and disease representations from a patient–disease observation graph and a disease ontology graph, while integrating attention-weighted clinical-text features into sequential prediction.
- **TRANS** (Chen et al., 2024) represents each longitudinal EHR as a temporal heterogeneous graph and applies a temporal graph transformer that combines temporal edge features with global positional and local structural encodings.
- **BoxLM** (Tan et al., 2025) fuses language-model semantics with ontology- and EHR-derived hierarchies through box embeddings and models evolving patient states for diagnosis prediction using volume-based similarity.
- **HieraMed** (Ke et al., 2026) encodes LLM-generated medical-code descriptions with Sentence-BERT, aggregates diagnoses, procedures, and medications in visit-centered heterogeneous graphs, and models patient-state evolution across visits with a Liquid Neural Network for next-visit diagnosis prediction.
- **DeepSeek-V4-Flash (zero-shot)** is prompted directly to perform next-visit diagnosis prediction without in-context demonstrations or task-specific parameter updates.

DeepSeek-V4-Flash zero-shot evaluation. We evaluate DeepSeek-V4-Flash on the test splits of MIMIC-III and MIMIC-IV without in-context examples or task-specific training. For MIMIC-IV, we use the 10,000-patient cohort. The prompt presents a patient's prior diagnosis codes in chronological visit order and asks for 40 unique next-visit diagnoses ranked by likelihood. It includes neither diagnosis names nor target diagnoses. Non-thinking and thinking use the same prompt and differ only in whether the API's thinking mode is enabled. Both modes use temperature 0 and a maximum generation budget of 8,192 tokens on each dataset.

We evaluate the final answer rather than the model's reasoning text. Predicted codes are normalized, matched to the dataset's diagnosis catalog, and deduplicated while preserving their order. Invalid

codes are discarded, and at most 40 predictions are retained. R@k averages the fraction of ground-truth diagnoses retrieved among the top k predictions across visits, for $k \in \{10, 20, 30, 40\}$.

E.3 Implementation Details

We use Qwen3-1.7B (Yang et al., 2025) as the backbone and optimize all model parameters in each training stage. DeepSeek-V4-Flash is used to generate the disease-level and longitudinal corpora for enriched SID–language alignment and reasoning activation. In the multi-task alignment stage, SID and visit-boundary tokens are appended to the tokenizer vocabulary and their embeddings are trained jointly with the backbone. We train with AdamW, an effective batch size of 1024, a per-GPU micro-batch size of 8, a learning rate of 3×10^{-4}, and a maximum sequence length of 1024. Training runs for at most five epochs with early stopping on validation loss. We use 20 warm-up steps followed by cosine learning-rate decay to 10% of the initial rate.

The reasoning-activation stage initializes from the multi-task alignment checkpoint and retains the same effective batch size, micro-batch size, sequence length, scheduler, and maximum number of epochs, while reducing the learning rate to 1×10^{-5}. Both supervised stages use bfloat16 precision and seed 42.

For reinforcement learning, we implement GRPO (Shao et al., 2024) with the VERL framework. We use a training batch size and PPO mini-batch size of 256, a per-GPU PPO micro-batch size of 8, and a learning rate of 1×10^{-6}. For each input, we sample 16 rollouts with temperature 1.0 and top-p 1.0; both the maximum prompt length and maximum response length are 1024. We set the KL-loss coefficient to 1×10^{-3}, the entropy coefficient to zero, and the multi-positive loss coefficient to $\lambda_{\mathrm{MP}} = 5 \times 10^{-3}$. The actor is trained for at most ten epochs with gradient checkpointing. All stages are distributed over four NVIDIA A800 GPUs.

E.4 GRPO Optimization Details

For negative-advantage tokens, we additionally apply dual clipping for stability. A KL penalty to the fixed reasoning-activation policy constrains drift.

The original rollouts contribute the GRPO and KL gradients. A second forward pass over answer-replaced branches contributes the multi-positive gradients. These gradients accumulate on the same actor parameters across microbatches, followed by gradient norm clipping and one optimizer step. At update t, eligible rollouts are ordered by rollout index and assigned successive ground-truth SIDs from a fixed enumeration of Y, starting at index $((t-1)N) \bmod |Y|$.

E.5 Catalog-Constrained Beam Search

Let $\mathcal{Q}$ be the catalog of valid three-token SIDs. We build a prefix trie from their tokenizer IDs. For a generated SID prefix a, the trie returns the tokens that extend a toward at least one SID in $\mathcal{Q}$:

$$\mathcal{A}(a) = \{v : a \parallel v \text{ is a prefix of some } q \in \mathcal{Q}\}. \tag{11}$$

Here $a \parallel v$ appends token v to prefix a. After a complete SID, only the end-of-sequence token is allowed. At each decoding step, we mask tokens outside $\mathcal{A}(a)$. Starting from the empty prefix, beam search extends each retained path with its allowed next tokens and keeps the B paths with the highest cumulative log probability. This continues through the end-of-sequence token. Completed paths are checked against the catalog and deduplicated before mapping SIDs to diagnoses.

Direct inference uses the checkpoint after reasoning activation. The prompt contains the longitudinal SID history H and disables explicit thinking. The SID-token contribution to the score of diagnosis d is

$$S_{\mathrm{dir}}(d \mid H) = \sum_{l=1}^{L} \log p_\theta\left(s_d^l \mid H, s_d^{<l}\right). \tag{12}$$

The implementation ranks completed beams by their full sequence scores, which also include the terminating end-of-sequence token. The thinking setting uses the checkpoint after coverage-reward RL with multi-positive supervision. We first sample R reasoning chains. Each chain is appended to the same history prompt, and the trie constrains only the subsequent SID tokens. The resulting local rankings are fused as in Equation 6.

E.6 General Ability Evaluation

Table 5 reports general-ability scores for the base Qwen3-1.7B model and two intermediate training stages. Enriched alignment (S2) reduces performance on both benchmarks. Adding general reasoning data (S3) partially recovers these abilities.

Table 5: General-ability evaluation across training stages. Higher scores are better.

Model	IFEval	GSM8K
Base Qwen3-1.7B	0.6820	0.8832
S2 = S1 + Enriched Alignment	0.1460	0.5238
S3 = S2 + General Reasoning	0.4713	0.6300