D^{2}KGMed: Dynamic Diagnostic Knowledge Graphs for Medical Diagnosis Prediction

Jie Zhang[1], Gaoyang Zheng[1], Hang Lv[1], Linhao Luo[2], Guofang Ma[3], Zhigang Lin[4], Xiping Chen[5], Yanchao Tan[1*]
[1] Fuzhou University, Fuzhou, China
[2] Monash University, Melbourne, Australia
[3] Zhejiang Gongshang University, Hangzhou, China
[4] Fujian Medical University, Fuzhou, China
[5] SKEMA Business School, Suzhou, China

2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3315-1557-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/BIBM66473.2025.11357158

***Abstract*—Accurate diagnosis prediction using Electronic Health Records (EHRs) is essential for personalized healthcare. Clinical knowledge graphs (KGs) can enrich EHRs by structuring medical knowledge, and recent work integrates large language models (LLMs) with KGs to enhance reasoning. However, these approaches often depends on static, expensive global graph construction and one-time retrieval, yielding noisy or irrelevant subgraphs that hinder effective diagnosis prediction in real-world clinical scenarios. To this end, we propose D^{2}KGMed, a diagnosis prediction framework that constructs a patient-specific Dynamic Diagnostic Knowledge Graph guided by LLMs. It consists of two stages: constructing an initial graph from diagnostic entities and multi-source medical knowledge; refining its construction via supervised fine-tuning to better align with the ideal graph for conciseness and relevance, and subsequently leveraging it for interpretable predictions. This design reduces graph construction costs and retrieval noise common in KG+LLM methods, enabling more accurate diagnosis prediction. Extensive experiments on two real-world EHR datasets demonstrate that D^{2}KGMed outperforms state-of-the-art baselines, especially in few-shot learning scenarios, showcasing its practical utility in real-world clinical settings.**

***Index Terms*—Electronic Health Records, Diagnosis Prediction, Knowledge Graph, Large Language Models.**

I. Introduction

Accurate diagnosis prediction from Electronic Health Records (EHRs) is crucial for personalized treatment and improved outcomes. However, privacy concerns and data scarcity limit patient-specific EHRs, creating a few-shot learning challenge [1] that demands robust methods for reliable prediction with minimal data. To enhance predictive performance, clinical knowledge graphs (KGs) were adopted to complement EHR modeling [2]–[4]. These KGs structure medical concepts and their relationships, enabling the learning of latent patterns and dependencies within clinical data. By providing structured, evidence-based knowledge, KGs support accurate diagnosis prediction. However, conventional approaches that combine deep learning (DL) with KGs to predict diagnoses by modeling KG structures and diagnostic representations often lack sufficient interpretability and face greater limitations in few-shot settings due to the inherent sparsity of patient data.

*Corresponding author: yctan@fzu.edu.cn

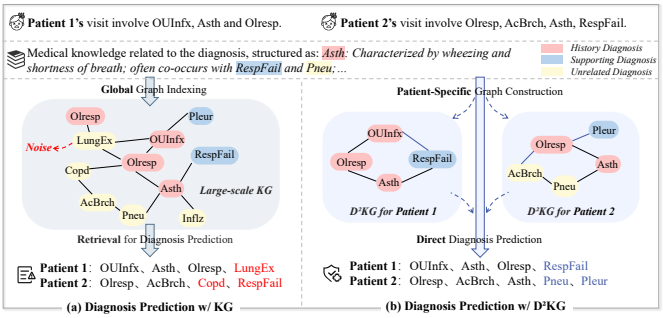

Fig. 1. An illustrative example of (a) the limitation of conventional methods leveraging KG for diagnosis prediction and (b) our proposed patient-specific D²KG for diagnosis prediction.

Recent studies explore integrating KGs with large language models (LLMs) and retrieval-augmented generation (RAG) to improve prediction accuracy. One category of KG+LLM methods retrieves information from existing clinical KGs (e.g., UMLS [5]) to enhance diagnostic support. While these KGs offer comprehensive coverage, they often lack patient-specific context, and their mappings to real-world EHR data can be noisy or ambiguous. Another category of KG+LLM approaches typically organizes medical knowledge as text-associated KGs. Representative work like GraphRAG [6] leverage LLMs to extract entities and relations for global graph indexing. Subsequently, retrieval and community detection aggregate information to generate summaries that support diagnosis prediction, as illustrated in Fig. 1(a).

However, these approaches rely on a single retrieval pass, which often results in noisy and redundant diagnoses. As illustrated in Fig. 1(a) for Patient 1, the global KG contains reasoning paths linking historical diagnoses to "Lung disease due to external agents" (LungEx), causing erroneous retrieval and incorrect predictions. In contrast, leveraging patient-specific KGs, as shown in Fig. 1(b), effectively removes noise and correctly predicts all diagnoses. In addition to producing noisy graphs, existing KG+LLM-based approaches also incur substantial time and token costs. Such inefficiencies significantly hinder their practical effectiveness in clinical settings.

To this end, we propose D^{2}KGMed, a novel approach that leverages **D**ynamic **D**iagnostic **K**nowledge **G**raph (D²KG) for

Medical diagnosis prediction. D²KGMed enhances predictive accuracy by integrating LLMs with the most informative D²KG mined from patient-specific knowledge, as illustrated in Fig. 1(b). Specifically, the D²KG framework comprises two key components: (1) **D²KG Construction** that extracts diagnostic entities from historical and candidate diagnoses as key entities, enriches them with multi-source medical knowledge, and uses LLMs to identify their relations to form the initial graph. (2) **Diagnosis Prediction with Refined D²KG**, which explicitly refines the graph via supervised fine-tuning (SFT) to better align with the ideal KG distribution, thereby guiding the model toward accurate and interpretable predictions.

Evaluation on two real-world EHR datasets against 10 competitive baselines shows that D²KGMed significantly enhances prediction accuracy, outperforming state-of-the-art (SOTA) models. Our main contributions are as follows:

- We propose the D²KGMed framework, which integrates patient-specific dynamic diagnostic knowledge graphs into diagnosis prediction. This approach mitigates the high costs of KG construction and the low patient relevance in retrieval typical of traditional KG+LLM methods, while effectively reducing noise interference in diagnosis prediction.
- We conduct extensive experiments on real-world EHR datasets, demonstrating that D²KGMed consistently outperforms state-of-the-art models in both visit-level and code-level diagnosis prediction.
- D²KGMed effectively expanding the model's reasoning scope and mitigating KG content insufficiency caused by sparse EHR data in few-shot settings, underscoring its practical value in clinical applications.

II. Related Work

A. Diagnosis Prediction

Diagnosis prediction has been extensively studied due to its pivotal role in healthcare. Traditional DL-based methods like RETAIN [7] ,TRANS [2] and BoxLM [8] rely on structured EHRs to model patient visit sequences for diagnosis prediction. However, these models face challenges from data sparsity and limited external knowledge integration.

LLM-based methods like MedReason [9] and Medical-CoT [10] show strong diagnostic ability but still suffer from hallucinations due to limited domain knowledge integration.

B. Knowledge Graph-enhanced Healthcare

KGs have emerged as promising tools for improving healthcare task. Previous approaches like CGL [3] incorporated KGs to improve performance of DL-based models. Although these methods embed KG structural knowledge into models, they are lack sufficient structure for accurate reasoning.

Recent studies focus on strategies integrating KGs to enhance LLMs and RAG. Some general-task KG+LLM methods [6], [11] showing strong potential for healthcare applications. These methods follow a two-stage graph construction and retrieval approach, but one-shot retrieval often yields noisy, incurring high time and token costs.

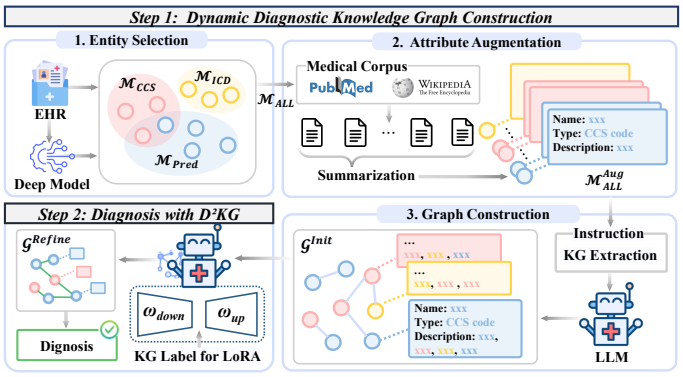

Fig. 2. The overall framework of D²KGMed, illustrating the process of training the model to construct the most informative D²KG for diagnosis prediction.

III. Methodology of D²KGMed

In this section, we define the problem and then present our proposed D²KGMed, with the overall framework illustrated in Fig. 2. D²KGMed enhances diagnosis prediction by integrating LLMs with the most informative D²KG derived from patient-specific knowledge. It comprises two key steps: (1) D²KG Construction and (2) Diagnosis Prediction with Refined D²KG.

A. Problem Definition

Given a patient's EHR data, consisting of a sequence of visits where each medical event is represented by a unique code (e.g., diagnoses and CCS codes) and associated with a name in the form of a short text snippet, the task is to predict the relevant CCS codes for the patient's next visit.

B. Step 1: D²KG Construction

To address the limitations of leveraging existing KGs such as UMLS [5] or those constructed from the entire EHR database [12], which often introduce irrelevant or misleading knowledge and lead to hallucinated predictions, we construct a patient-specific D²KG derived from individual EHR visits to provide more relevant and reliable context.

*1) **Entity Selection**:* Constructing an all-diagnosis medical KG is costly and usually yields a static structure that lacks patient specificity. To address this, we first perform entity selection for the D²KG by adopting a hybrid strategy that integrates both history-driven and prediction-driven perspectives.

History-driven entity selection. Let a patient i's EHR data be represented as a sequence of visits $\{v_1, v_2, \ldots, v_t\}$, where each visit v_t includes diagnosis codes $c_{\text{ICD}} \in \mathcal{M}_{\text{ICD}}$ and CCS codes $c_{\text{CCS}} \in \mathcal{M}_{\text{CCS}}$. We collect all such codes from historical visits to construct the history-driven entity set, denoted as $\mathcal{M}_{\text{Hist}} = \mathcal{M}_{\text{ICD}} \cup \mathcal{M}_{\text{CCS}}$.

Prediction-driven entity selection. To capture potential future diagnoses, we employ a clinical deep learning model (denoted as $\mathcal{F}_{\text{pred}}$; we use TRANS [2] in this work) to model the patient's EHR sequence and predict likely future diagnoses. The top-k predictions (with $k = 50$ in our implementation) are selected as candidate entities for KG construction:

$$\mathcal{M}_{\text{Pred}} = \text{Top-}k\left(\mathcal{F}_{\text{pred}}(\{v_1, v_2, \ldots, v_t\})\right). \quad (1)$$

Although the prediction accuracy of $\mathcal{F}_{\text{pred}}$ may be limited in few-shot scenarios, its output remains useful for broadening the coverage of relevant knowledge in downstream tasks.

By combining both sets, we construct the complete entity set for D²KG construction: $\mathcal{M}_{\text{ALL}} = \mathcal{M}_{\text{Hist}} \cup \mathcal{M}_{\text{Pred}}$.

*2) **Attribute Augmentation**:* Diagnosis names alone are insufficient for prediction as they miss complex relationships among medical entities. We enrich entity attributes as follows:

Multi-source knowledge retrieval. To enhance the coverage of clinical information, we retrieve external knowledge from diverse sources such as PubMed [13] and Wikipedia [14]. For each medical code $c \in \mathcal{M}_{\text{ALL}}$, we use its surface name $s \in \mathcal{S}_{\text{ALL}}$ as a query to construct a text corpus $\mathcal{D}_{\text{med}}$. We then employ Dragon [15] (denoted as Ret$(\cdot)$), a dual-encoder retriever, to obtain the top-n most relevant passages (n = 50 in this work), denoted as $\mathcal{T}_c = \text{Ret}(s, n, \mathcal{D}_{\text{med}})$, which serve as external knowledge for the medical code c.

Task-specific knowledge summarization. Retrieved passages $\mathcal{T}_c$ provide external context for code c, but their direct use often introduces lengthy and noisy inputs with limited task relevance. To improve utility, we choose to use strong LLM-based summarizer [16] to condense the retrieved passages $\mathcal{T}_c$ into a task-specific summary e based on the prompt $prom_{\text{sum}}$, i.e., $e = Summarizer(prom_{\text{sum}}, \mathcal{T}_c)$. The prompt enables the model to extract code c's information and its relations with other codes mentioned in the retrieved content.

With the attribute augmentation for each code $c \in \mathcal{M}_{\text{ALL}}$ completed, we obtain the augmented code set:

$$\mathcal{M}^{\text{Aug}}_{\text{ALL}} = \bigcup_{t \in \{\text{ICD,CCS,Pred}\}} \mathcal{M}^{\text{Aug}}_{t}, \tag{2}$$

where $\mathcal{M}^{\text{Aug}}_{t} = \{(c_t, e_t)\}$. Based on this, we construct the diagnosis prediction dataset $\mathbb{D} = \{(x_i, y_i)\}$, where each x_i is a diagnosis instruction for patient i generated from the flattened $\mathcal{M}^{\text{Aug}}_{\text{ALL}}$, and y_i is the ground-truth CCS label.

*3) **Graph Construction**:* Given x_i, an unstructured diagnostic knowledge containing medical entities and relations, we aim to extract a structured KG. To this end, we design a multi-step prompt for graph construction, denoted as $prom_{\text{kg}}$.

Leveraging the LLM's semantic understanding and reasoning capabilities, we extract structured triples from x_i for codes in $\mathcal{M}^{\text{Aug}}_{\text{ALL}}$, where each triple (s_c, t_c, d_c) consists of an entity name, type, and a contextualized description [6], [12]. Unlike the static attributes in the previous step, these descriptions are patient-adaptive, capturing context-specific features and relations such as *"co-occurs with"* grounded in actual diagnoses.

This process yields the initial construction of patient i's D²KG, denoted as: $\mathcal{G}_i^{init} = \bigcup_{c \in \mathcal{M}^{\text{Aug}}_{\text{ALL}}} \{(s_c, t_c, d_c)\}$.

C. Step 2: Diagnosis Prediction with Refined D²KG

The initial graph $\mathcal{G}_i^{\text{init}}$ generated by the backbone model often contains redundant entities and verbose descriptions, introducing noise and reducing its utility. To address this issue, we adopt a supervised training strategy that guides the model to construct concise and informative D²KGs.

TABLE I
STATISTICS OF THE DATASETS USED IN OUR EXPERIMENTS.

Dataset	MIMIC-III	MIMIC-IV
# of patients	5,449	79,393
# of visits	14,141	329,605
Avg. # visits per patient	2.60	4.15
Max. # visits per patient	29	169
Avg. # all CCS per visit	12.08	10.62
# of unique diagnoses	3,874	37,917
# of CCS codes	285	842

Specifically, given $prom_{\text{kg}}$, we leverage a strong LLM (gpt-4o in our setup) to generate doctor-verified golden KGs, which are used to fine-tune the model via LoRA [17] on $\mathbb{D}_{train}$.

Through supervision from these ideal KGs, the model learns to refine the graph in a diagnosis-oriented manner, resulting in $Model_{\text{kg}}$ with the ability to construct a patient-adaptive refined graph $\mathcal{G}^{\text{refine}}$. Given the diagnostic information of patient i, the next-diagnosis prediction Ans is formulated as:

$$\text{Ans} = Model_{\text{kg}}\big(prom_{\text{pred}}, \mathcal{G}^{\text{refine}}, \mathcal{M}^{\text{Aug}}_{\text{ALL}}\big), \tag{3}$$

where $prom_{\text{pred}}$ denotes the diagnosis prediction prompt.

IV. EXPERIMENT

A. Experimental Setup

Datasets and Evaluation Protocols. We use two real-world EHR dataset MIMIC-III [18] and MIMIC-IV [19]. The statistics are summarized in Table I. We use visit-level Precision@K (P@K) and code-level Accuracy@K (A@K), which are consistent with prior work [2], [8].

Baselines. To comprehensively evaluate `D²KGMed`, we compare it with 10 representative SOTA methods across four categories: (1) DL-based methods: Transfomer [20], RETAIN [7], Stagenet [21], KAME [22]; (2) KG+DL-based methods: TRANS [2], CGL [3]; (3) LLM-based methods: MedicalCoT [10], MedReason [9]; (4) KG+LLM-based methods: GraphRAG [6], PathRAG [11].

B. Overall Diagnosis Prediction Results

As shown in Table II, `D²KGMed` consistently outperforms all baselines on both the MIMIC-III and MIMIC-IV datasets.

TABLE II
RESULTS OF DIAGNOSIS PREDICTION (%) ON MIMIC-III/IV WITH 5% TRAINING DATA. BEST IN BOLD; SECOND BEST UNDERLINED.

Dataset	MIMIC-III				MIMIC-IV			
Metric	Visit-Level		Code-Level		Visit-Level		Code-Level	
	P@10	P@20	A@10	A@20	P@10	P@20	A@10	A@20
Transformer	35.09	43.94	26.25	42.90	36.27	40.34	25.35	43.92
RETAIN	38.54	46.25	28.61	44.78	43.19	49.05	31.49	<u>45.60</u>
StageNet	35.63	43.72	26.75	42.94	37.69	43.45	27.69	40.89
KAME	35.61	44.36	26.59	43.18	30.54	37.16	22.29	34.10
TRANS	36.67	44.92	27.44	43.78	36.00	41.95	26.52	39.83
CGL	38.98	45.98	28.06	44.89	32.48	38.01	22.42	36.72
MedicalCoT-8B	41.66	44.77	29.17	43.91	42.53	44.03	30.37	42.57
MedReason-8B	42.36	45.84	28.61	43.58	43.96	45.28	30.28	42.36
GraphRAG	41.04	<u>47.45</u>	31.19	<u>45.33</u>	43.51	48.01	<u>32.67</u>	44.26
PathRAG	39.21	43.77	26.42	43.41	39.71	46.11	27.07	43.11
Base	40.86	44.46	28.47	42.91	44.73	45.55	29.91	43.88
+Knowledge	40.09	44.56	28.34	42.94	44.14	45.39	28.82	43.67
+Original KG	<u>43.65</u>	47.38	<u>32.61</u>	44.58	<u>46.35</u>	<u>49.99</u>	32.65	44.35
`D²KGMed`	**45.23**	**50.87**	**32.82**	**45.65**	**47.88**	**50.96**	**33.47**	**46.17**

Comparison with DL-based methods. RETAIN is a strong diagnostic baseline but relies on static EHR knowledge, limiting semantic richness and interpretability. In contrast, D^{2}KGMed integrates external medical knowledge to construct patient-specific D²KG, enabling traceable reasoning. As a result, D^{2}KGMed outperforms RETAIN by 173.59% (P@10) and 147.15% (A@10) on MIMIC-III, demonstrating its strength in constructing and leveraging dynamic knowledge structures.

Comparison with KG+DL-based methods. Unlike KG-based models that rely on EHR-specific structures (e.g., TRANS), D^{2}KGMed constructs diagnostic graphs dynamically using LLMs and external knowledge, making it more robust in sparse or few-shot scenarios. Although guided by TRANS predictions, D^{2}KGMed still achieves substantial P@10 gains of 23.43% on MIMIC-III and 33.00% on MIMIC-IV, demonstrating the strong diagnostic capability of D²KG.

Comparison with LLM-based methods. LLM-based methods such as MedicalCoT and MedReason perform well on certain metrics but lack explicit knowledge structures. By integrating patient-specific D²KG, D^{2}KGMed expands the reasoning space beyond candidate answers. On MIMIC-III and MIMIC-IV, D^{2}KGMed improves P@10 by an average of 9.21% over these methods, showing superior diagnostic reasoning.

Comparison with KG+LLM-based methods. GraphRAG and PathRAG effective on some metrics, but they rely on large pre-built index graphs and heavy retrieval, limiting adaptability to personalized clinical contexts and causing unstable predictions. In contrast, D^{2}KGMed allows more precise and controllable knowledge use. On MIMIC-III, improves A@10 by an average of 14.73% over these methods.

C. Ablation Study

We evaluate the impact of medical knowledge and D²KG in D^{2}KGMed: (1) "Base": Prompts the LLM with patient history and candidate diagnoses; (2) "+ Knowledge": Enhances candidates via NaiveRAG on external knowledge; (3) "+ Original KG": Injects the initial D²KG from (2) as structured knowledge; (4) "+ Refined KG (D^{2}KGMed)": Uses fine-tuned LLM to extract optimal D²KG. Results in Table II show that:

External Medical Knowledge substantially improves CCS prediction, though long contexts (7k/12k tokens in MIMIC-III/IV) slightly degrade P@10. *The introduction of structured D²KG* mitigates long-context burden, yielding average performance gains of +29.02% on MIMIC-III and +21.72% on MIMIC-IV. *The design of Refined D²KG* yields notable gains, achieving SOTA on all metrics across both datasets, highlighting its interpretability in complex clinical scenarios.

V. Conclusion

In this paper, we propose D^{2}KGMed, a patient-specific D²KG-guided framework for diagnosis prediction. By dynamically constructing diagnostic KGs, D^{2}KGMed enhances accuracy and interpretability over traditional KG+LLM methods. Experiments on real-world EHR data confirm its effectiveness, highlighting the promise of dynamic knowledge modeling for clinical practice. Future work will extend this approach to broader healthcare tasks and real-world deployment.

Acknowledgment

This work was supported in part by the Fujian Provincial Artificial Intelligence Industry Development Technology Project under Grant (2025H0042) and Talent Foundation of Fuzhou University (No. XRC-23027 and No. XRC-23091).

References

[1] X. Xu, B. Yao, Y. Dong, S. Gabriel, H. Yu, J. Hendler, M. Ghassemi, A. K. Dey, and D. Wang, "Mental-llm: Leveraging large language models for mental health prediction via online text data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–32, 2024.

[2] J. Chen, C. Yin, Y. Wang, and P. Zhang, "Predictive modeling with temporal graphical representation on electronic health records," in *IJCAI: proceedings of the conference*, vol. 2024, 2024, p. 5763.

[3] C. Lu, C. K. Reddy, P. Chakraborty, S. Kleinberg, and Y. Ning, "Collaborative graph learning with auxiliary text for temporal event prediction in healthcare," *arXiv preprint arXiv:2105.07542*, 2021.

[4] P. Jiang, C. Xiao, and other, "Graphcare: Enhancing healthcare predictions with personalized knowledge graphs," in *12th International Conference on Learning Representations, ICLR 2024*, 2024.

[5] O. Bodenreider, "The unified medical language system (umls): integrating biomedical terminology," *Nucleic acids research*, vol. 32, no. suppl_1, pp. D267–D270, 2004.

[6] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody *et al.*, "From local to global: A graph rag approach to query-focused summarization," *arXiv preprint arXiv:2404.16130*, 2024.

[7] E. Choi *et al.*, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," *Advances in neural information processing systems*, vol. 29, 2016.

[8] Y. Tan *et al.*, "Boxlm: Unifying structures and semantics of medical concepts for diagnosis prediction in healthcare," in *Forty-second International Conference on Machine Learning*.

[9] J. Wu, W. Deng, X. Li, S. Liu, T. Mi, Y. Peng, Z. Xu, Y. Liu, H. Cho, C.-I. Choi *et al.*, "Medreason: Eliciting factual medical reasoning steps in llms via knowledge graphs," *arXiv preprint arXiv:2504.00993*, 2025.

[10] E. Karataş, "Deepseek-r1-medical-cot," https://huggingface.co/emredeveloper/, 2025.

[11] B. Chen, Z. Guo, Z. Yang, Y. Chen, J. Chen, Z. Liu, C. Shi, and C. Yang, "Pathrag: Pruning graph-based retrieval augmented generation with relational paths," *arXiv preprint arXiv:2502.14902*, 2025.

[12] P. Jiang, C. Xiao, M. Jiang, P. Bhatia, T. Kass-Hout, J. Sun, and J. Han, "Reasoning-enhanced healthcare predictions with knowledge graph community retrieval," *arXiv preprint arXiv:2410.04585*, 2024.

[13] K. Canese and S. Weis, "Pubmed: the bibliographic database," *The NCBI handbook*, vol. 2, no. 1, p. 2013, 2013.

[14] D. Vrandečić *et al.*, "Wikidata: a free collaborative knowledgebase," *Communications of the ACM*, vol. 57, no. 10, pp. 78–85, 2014.

[15] S.-C. Lin, A. Asai, M. Li, B. Oguz, J. Lin, Y. Mehdad, W.-t. Yih, and X. Chen, "How to train your dragon: Diverse augmentation towards generalizable dense retrieval," *arXiv preprint arXiv:2302.07452*, 2023.

[16] OpenAI, "Hello gpt - 4o," https://openai.com/index/hello-gpt-4o/, 2024.

[17] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen *et al.*, "Lora: Low-rank adaptation of large language models." *ICLR*, vol. 1, no. 2, p. 3, 2022.

[18] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[19] A. E. Johnson, Stone *et al.*, "The mimic code repository: enabling reproducibility in critical care research," *Journal of the American Medical Informatics Association*, vol. 25, no. 1, pp. 32–39, 2018.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[21] J. Gao, C. Xiao, Y. Wang, W. Tang, L. M. Glass, and J. Sun, "Stagenet: Stage-aware neural networks for health risk prediction," in *Proceedings of the web conference 2020*, 2020, pp. 530–540.

[22] F. Ma, Q. You, H. Xiao, R. Chitta, J. Zhou, and J. Gao, "Kame: Knowledge-based attention model for diagnosis prediction in healthcare," in *Proceedings of the 27th ACM international conference on information and knowledge management*, 2018, pp. 743–752.