ClinPath: A General-Purpose Knowledge Graph with LLM Reasoning For Understanding Clinical Interactions

Sahithi Ankireddy* SANKIRED@CALTECH.EDU
California Institute of Technology, USA

Purvi Sehgal* PSEHGAL@ALUMNI.CALTECH.EDU
California Institute of Technology, USA

Adam Wierman ADAMW@CALTECH.EDU
California Institute of Technology, USA

Abstract

We present ClinPath, a holistic multimodal framework that combines knowledge graph modeling with large language model (LLM) reasoning to comprehensively represent and and analyze longitudinal patient clinical journeys. Built on the MIMIC-IV database, ClinPath introduces ClinKG, a large-scale clinical knowledge graph that integrates diagnoses, symptoms, medications, procedures, demographics, and provider interactions into a unified representation of patient care. Unlike prior work that constructs narrow, diagnosis-centered graphs, ClinKG captures the full spectrum of patient–provider interactions across time and care settings. The LLM reasoning layer demonstrates ClinPath's versatility through two key applications: (1) patient similarity analysis, where this pipeline significantly improved performance on our custom benchmark, ClinPath-SimBench, and (2) provider behavior analysis, a novel downstream task. Together, these results illustrate how combining graph-structured representations with LLM-based reasoning yields clinically meaningful, multi-perspective insights.

Keywords: Healthcare AI, Knowledge Graphs, Large Language Models, MIMIC-IV, Clinical Interactions, Patient Similarity, Multimodal Data Integration

Data and Code Availability We repurposed MIMIC-IV dataset (Johnson et al., 2024) and the code is made publicly available here. Our patient similarity benchmark dataset is publicly available here.

Institutional Review Board (IRB) Our research does not require IRB approval

* These authors contributed equally.

1. Introduction

The medical field generates a wide range of data distributed across various modalities and sources. Clinical information may take the form of unstructured physician notes, structured electronic health records (EHRs), medical imaging (e.g., X-rays, CT scans), laboratory test results, and genomic data. These modalities span a different format—text, images, time-series, or structured tabular records. They each capture distinct aspects of a patient's health status and together offer a more holistic view of the clinical journey.

However, effectively integrating this multimodal data remains a key challenge in healthcare. The data ranging from free-text to high-dimensional omics data makes it difficult to create unified representations. Bridging these sources in a way that preserves semantic meaning and clinical utility is a key step toward enabling more effective care. This challenge has led to an active area of research focused on multimodal machine learning, representation learning, and the development of frameworks that can reason across diverse data types in a meaningful way (Acosta et al., 2022; Kline et al., 2022; Artsi et al., 2024; Cai et al., 2019).

Knowledge graphs (KGs) have emerged as a promising approach for organizing this fragmented information. By representing clinical concepts—such as diagnoses, symptoms, medications, procedures, and providers—as nodes, and their interactions as edges, KGs unify diverse data sources into a structured representation. Biomedical KGs have been successfully applied to various downstream tasks, including drug repurposing, clinical decision support, and adverse event prediction (Chandak et al., 2023; Wu et al.,

2023; Mishra and Shridevi, 2024). However, most existing KGs are either domain-specific or built around a narrow set of entities, limiting their generality. Moreover, the construction of new knowledge graphs for different use cases introduces redundancies that reduce KG research efficiency and can be simplified with a more general knowledge graph that can enable broader re-usability. However, building a comprehensive, general-purpose KG that accurately captures the full scope of a patient's clinical journey remains an open challenge.

ClinPath introduces ClinKG, a large-scale, general-purpose knowledge graph that captures information across the entire clinical journey, extending beyond patient or diagnosis-centric data. It includes details about providers, pharmacists, procedures, medications, and more, enabling a comprehensive view of healthcare interactions. Unlike existing graphs that are narrowly focused, ClinKG is designed to be broadly applicable.

LLMs were then employed on top of ClinKG as the primary reasoning layer in the pipeline, operating over both the structured graph representations and raw calculated metrics (e.g., degree centrality, Jaccard scores). Traditional approaches for analyzing knowledge graphs rely on only graph algorithms and various neural network encoders, which learn latent representations from node features and the underlying graph structure and oftentimes have difficulty generalizing (Bai et al., 2023; Xu et al., 2016).

More recently, LLMs have shown strong performance in processing natural language representations of structured data. By converting relational triples into text, LLMs were applied on top of ClinKG to support a wide range of downstream applications, including the established task of patient similarity analysis and the novel task of provider behavior analysis. For patient similarity analysis, we introduce a custom benchmark, ClinPath-SimBench, and evaluate it against state-of-the-art baselines.

The key contributions of this project are as follows:

- Construction of ClinKG, a large, general-purpose clinical knowledge graph that models diverse healthcare entities and their interactions.
- Integration of LLMs as a reasoning layer on top of ClinKG to demonstrate its usability across multiple downstream tasks.
- Empirical evaluation through two representative applications: (i) patient similarity analysis and (ii) provider behavior analysis, highlighting the framework's versatility and state-of-art performance compared to existing patient similarity models
- A new benchmark dataset, ClinPath-SimBench, for evaluating patient similarity models with clinical relevance

2. Literature Review

Knowledge graphs have been used in the biomedical field for years. These graphs have traditionally been very specialized and only designed for a small range of specific downstream applications. Cui et al. (2023) note that most biomedical knowledge graphs contain fewer than 20 nodes (implying limited feature diversity) and all were built for their specific downstream use cases.

Halamka and Cerrato (2023) used knowledge graphs to extract diagnosis and procedure codes for more accurate coding, but they only contained enough information for that task. Collectively, such graphs are currently being used to solve a wide range of tasks within biomedicine, underscoring their importance across various domains within the field. However, because they are highly specialized for their own tasks, each graph cannot be applied to many other problems besides their own.

Patient-centric knowledge graphs built from Electronic Health Records (EHRs) have similarly been developed for narrow, application-specific tasks. PharmKG (Zheng et al., 2021) only identified relationships between genes, drugs, and diseases, and DRKF knowledge graph (Zhang and Che, 2021) only linked drugs and Parkinson's disease. Mishra and Shridevi (2024) used EHRs to create a patient similarity graph using diagnoses, procedures, and medications information. This graph only had a limited number of features (diagnoses, procedures, and medications) and was tailored only for the downstream patient similarity task but none others. ClinKG was also applied to this patient similarity task to show its effectiveness in solving problems previously solved by specialized graphs.

Overall, knowledge graphs utilized for biomedical and patient-centric healthcare tasks were previously specialized and independent of each other, which added a layer of redundancy. To resolve these issues, this work introduces a general purpose knowledge graph that follows the patient's journey including

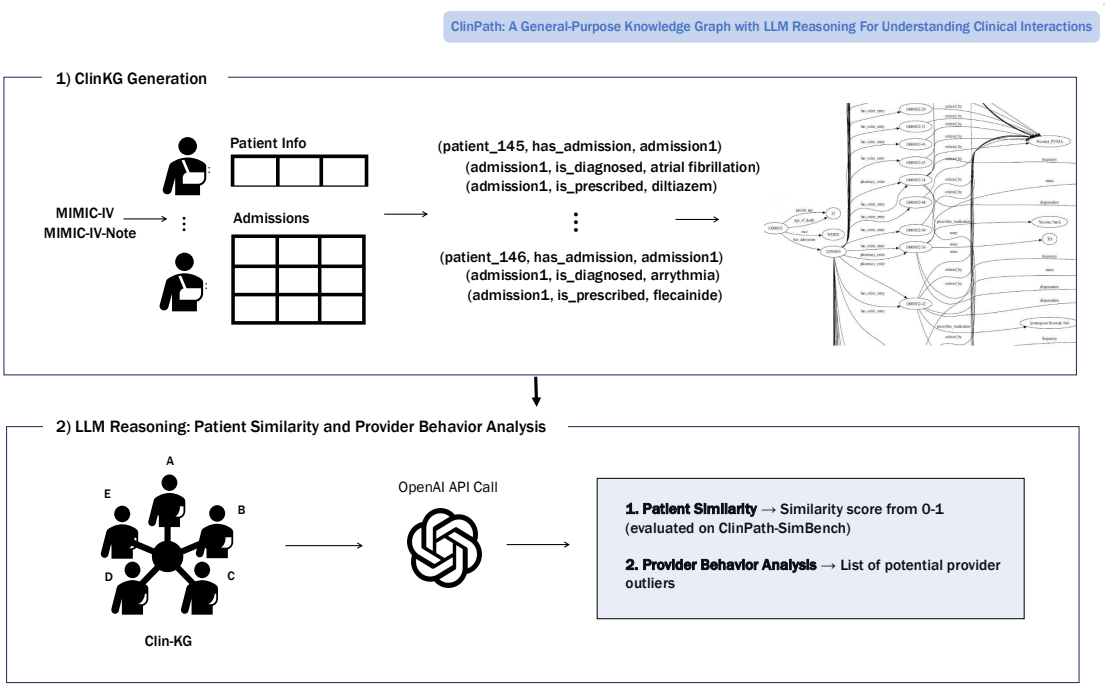

Figure 1: **ClinPath pipeline.** Overview of the ClinPath framework. Heterogeneous EHR data from MIMIC-IV is integrated into a unified knowledge graph ClinKG. An LLM reasoning layer operates over relation triples to support various downstream applications, including patient similarity (evaluated on ClinPath-SimBench) and provider behavior analysis.

symptoms, procedures, provider and pharmacy orders, diagnoses, and subsequent admissions, etc. This comprehensive view of the patient journey from start to end provides a unified framework which can help researchers solve a multitude of downstream tasks and understand clinical interactions more holistically.

3. ClinKG Generation

3.1. Overview

Relevant data from MIMIC-IV (Johnson et al., 2024) was processed and aggregated to generate temporally ordered relational information across multiple data modalities. Each patient's longitudinal clinical history from all their hospital admissions was structured into a unified clinical knowledge graph, ClinKG, that integrates tabular data (e.g., diagnoses, procedures), unstructured free text (e.g., symptoms and other information extracted from clinical notes), and numerical information (e.g., dosage and prescription details). Each admission was represented as a distinct node connected to the patient through temporal relationships, enabling the graph to encode both multimodal and longitudinal clinical data.

3.2. Dataset

MIMIC-IV (Johnson et al., 2024) database, contains tabular, textual, and imaging data for more than 300,000 patients across all hospital admissions. This covers information such as diagnoses, procedures, and other clinical events, as illustrated in Figure 2. The MIMIC-IV-Note (Johnson et al., 2023) database provides clinical notes associated with hospital stays, from which symptom information was extracted using GPT-4o (OpenAI, 2023). Altogether, the relevant information was originally distributed across 26 decentralized files. We concatenated, extracted, and reorganized these data sources into two standardized tables per patient, as described in Section 3.3.

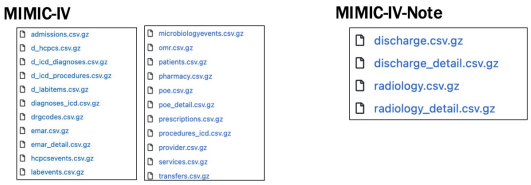

Figure 2: MIMIC-IV database tables used for constructing the clinical knowledge graph

3.3. Feature Extraction

For each patient, relevant information was aggregated into two key tables. The first table included patient information such as demographics and other commonalities across hospital admissions. The second table included patient admissions information where each row in the table contained the patient's admission information such as symptoms, prescribed medications, and provider name for that admission. In addition to extracting information relevant to this ClinKG graph from the 26 files mentioned above, some other necessary associations were created between features. For example, two provider orders were connected when one was discontinued by another and linked medications listed in a different file to provider orders. All codes were cross-referenced with a lookup table to retrieve their descriptions, which were then added as features to the table. All these features were organized in a dictionary by patient and admission numbers.

3.4. Knowledge Graph Creation

Once these features were extracted, a knowledge graph was generated. For each patient and admission, GPT-4o (OpenAI, 2023) was queried with the feature dictionary and feature descriptions. This produced a list of relational triples which linked one entity to another entity. An example of this is the patient linked to the admission number with the label, "has admission" or the admission number linked to a medicine with the label, "is prescribed."

These relational triples were concatenated and then iterated through to create a digraph using the Graphviz library. This graph captured the complex relationships between patients, providers, and other healthcare professionals, tracking each patient's full clinical journey—including orders placed, medications prescribed or discontinued, and other intermediate steps— across time.

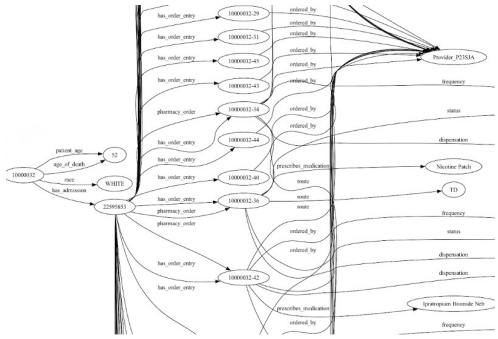

Figure 3: Segment of the knowledge graph for a single patient and admission

While this knowledge graph includes information for all patients and all admissions, Figure 3 displays only a segment of the knowledge graph for one patient and one admission for clarity.

3.5. Knowledge Graph Analytics

After constructing the longitudinal ClinKG, we performed a series of graph analyses to characterize its temporal structure and to provide interpretable signals for the LLM reasoning layer described in Section 4. Unlike static graphs that capture a single patient current status, ClinKG encodes each patient's full clinical trajectory across time. Each admission is a distinct node connected chronologically to prior and subsequent admissions, enabling temporal reasoning over evolving diagnoses, medications, and procedures. Thus, we computed graph-level and node-level statistics to understand connectivity and influence within this temporal network. This included degree distributions across node types (patients, admissions, diagnoses, medications, and procedures). Centrality measures, such as degree, betweenness, and eigenvector centrality, were used to identify clinically influential nodes (e.g., key diagnoses or medications that frequently mediate transitions across admissions) and to quantify their role in information flow over time. Similarity-based measures, including Jaccard similarity, were also used to quantify the overlap between patients' neighborhoods of diagnoses and treatments. The resulting analytics were not only used to characterize the graph but were also in-

tegrated as contextual signals for the LLM reasoning layer.

4. LLM Reasoning: Patient Similarity

Once constructed, the knowledge graph ClinKG can be applied to a wide range of downstream tasks. Its large scale captures information on patients, providers, medications, procedures, and symptoms, making it a powerful tool for clinical decision support.

A primary application and focus of this research was patient similarity analysis, where the goal was to identify patients with similar medical profiles. This is particularly valuable for providers aiming to better understand complex cases or for junior medical professionals who can more easily learn by referencing similar past patient cases. By identifying patients most similar to the current patient, clinicians can gain insights into likely disease trajectories, effective treatment responses, and potential risks. This, in turn, enables more personalized care and improves clinical decision-making and outcomes.

4.1. Experimental Set-up

To implement this, ClinKG triples for all patients were retrieved and saved to a pickle file. These relation triples were then reformatted into natural language strings in a two-step process. First, a triple, in the format (99384712, is_diagnosed, asthma), for example, was converted to the natural language sentence: 99384712 is diagnosed with asthma. These strings were then further organized by category (e.g., diagnoses, prescriptions, procedures) to introduce more structure and facilitate analysis.

This transformation was applied to all triples for each patient to create a consolidated textual representation per patient. This master string was then sent to an LLM, where it extracted similarity scores for each feature (symptoms, prescribed medications, and miscellaneous) between every pairwise combination of patients. Each similarity score was calculated on a scale from 0 to 1 based on the number of overlapping items. A final overall similarity score was computed by averaging the sub-scores across all features. The resulting information was structured into a JSON format for each pairwise set of patients, enabling the final similarity scores to be easily sorted in order to identify the most similar patients for any given individual.

4.2. Benchmark Creation

Because no suitable dataset existed to evaluate ClinPath, a novel benchmark dataset, ClinPath-SimBench, was constructed containing 120 patient profiles mapping to their pairwise similarity scores.

To build this dataset, a system of unique codes representing medical diagnoses (ICD-10-CM) was utilized. ICD-10-CM is organized as a hierarchy, where codes that are closer together generally represent to more similar conditions and codes that are further apart typically represent less similar conditions (Girardi et al., 2016). Leveraging this property, diagnosis code similarity derived from the hierarchy was used as an objective ground truth to evaluate patient similarity.

The authors began by randomly selecting ICD-10-CM diagnosis codes across ICD-10-CM chapters to ensure broad coverage across disease categories. For each selected diagnosis code, five different codes were identified corresponding to five levels of similarity relative to the original code. These similarity levels were derived from the ICD-10-CM hierarchy and conclusions from prior literature.

The similarity levels were:

- **1.00 similarity** → Exact code match (ex: I26.9, I26.9)
- **0.75 similarity** → Same 3 character prefix (ex: I26.9, I26.0)
- **0.50 similarity** → Same chapter (same letter) (ex: I26.9, I82.409)
- **0.25 similarity** → Different chapter but same organ system (ex: I26.9, J96.01)
- **0.0 similarity** → Different chapter and different organ system (ex: I26.9, H54.0)

For each diagnosis code, a patient profile was created from clinical research and computational methods. Patient profiles simulated LLM-readable natural language representations of the MIMIC-IV data tables and included clinical features such as symptoms, medications, physician orders, and procedures while explicitly excluding diagnosis-related information. Thus, patient profile pairs were mapped to corresponding ground truth similarity scores in this benchmark.

ClinPath-SimBench provided an objective framework to evaluate patient similarity and can be used to

enhance clinical models for improved treatment recommendations, clinical decision support, and clinical trial cohort discovery.

4.3. Baseline Models

The benchmark described in Section 4.2 was utilized to evaluate performance of ClinKG against baselines. A diverse group of baselines were selected spanning lexical, graph, and embedding-based models.

A traditional text similarity technique, *TF-IDF* (Salton and Buckley, 1988), was used to establish a performance floor. *Node2Vec* (Grover and Leskovec, 2016), a graph-embedding model, was included to evaluate the contribution of ClinKG's graph structure. *BioClinicalBERT*(Alsentzer et al., 2019), a transformer model pretrained on biomedical literature and MIMIC III clinical notes, was selected for its ability to capture deep contextual semantics within clinical text. *SapBERT* (Liu et al., 2021), a transformer model initialized from PubMedBERT and trained with a self-alignment objective on UMLS biomedical ontologies, was included because of its strength in identifying semantic similarity between medical terms, especially synonymous diagnoses. The *Patient Similarity Network (PSN)* was also evaluated as it has been extensively applied in patient similarity research (Navaz et al., 2022). It involves similarity driven by feature overlap (procedures, symptoms, etc) that is highly interpretable and thus heavily used in healthcare. Two PSN variants were evaluated: *PSN Learned*, in which the model learns optimal weights of features, and *PSN Uniform*, a simpler model in which all features have equal weights. Lastly, *ClinPathNoGraph*, a variant of ClinPath that excluded graph analytics augmentation described in Section 3.5, was included as a baseline to demonstrate the contribution of the knowledge graph component of ClinPath.

4.4. Evaluation

4.4.1. Benchmark

Once the benchmark was created, ClinPath and the baseline models were evaluated on it.

As shown in Table 1, ClinPath achieved the highest performance across almost all metrics, with consistent results and minimal variance (¡0.01 for all key metrics) across ten independent seeds. It obtained the highest Top1=1.0 score and Top1≥0.75 (fraction of queries where the highest ranked patient had a similarity of 1.0 or ≥0.75, respectively) indicating strong performance on exact matches and clinically relevant close matches. It also had the highest performance on the Normalized Discounted Cumulative Gain (NDCG) metric (measures ranking accuracy with emphasis on patients with high rankings), highlighting that the model outperformed baselines in prioritizing clinically similar patients in the ranking. ClinPath also had the highest Pearson correlation coefficient with statistically significant p-values, representing that the model produces similarity scores that more accurately match clinical expectations. While ClinPath had a slightly lower pairwise order accuracy than BioClinicalBERT, it was only by 0.005, which can be negligible given its superior performance across other metrics.

On the other hand, multiple models outperformed ClinPathNoGraph, a variant of ClinPath that excluded graph analytics augmentation. This underperformance demonstrated the importance of having graph-based reasoning for patient similarity tasks and emphasized the value of integrating a knowledge-graph component in ClinPath.

ClinPath not only outperformed other patient similarity models in accuracy, but also in ranking quality and score alignment.

4.4.2. ClinPath Inference Evaluation

In addition to the benchmark-based evaluation described in earlier sections, we conducted a separate controlled experiment to directly evaluate the ClinPath pipeline on the real constructed ClinKG. This setup was designed to test ClinPath's ability to infer patient similarity from ClinKG itself, rather than relying solely on the benchmark dataset.

For each real patient node in Clin-KG, three synthetic patient profiles were generated with predefined ground-truth similarity scores of 1.0, 0.5, and 0.0, using the same methodology as in the benchmark setup. These profiles were then appended to the existing Clin-KG, creating an augmented version of the real-world graph that retained its temporal and multimodal structure.

The LLM reasoning layer was applied to this augmented Clin-KG to predict similarity scores between each real patient and their corresponding synthetic profiles. This process was repeated for all real patients, enabling a direct evaluation of ClinPath's entire pipeline including the actual ClinKG. Spearman correlation and Pearson correlation values were then

Table 1: Performance comparison of ClinKG and baseline models on the new benchmark dataset, sorted from best to worst.

Model	Top1=1.0	Top1≥0.75	Pairwise Order	NDCG@all	Mean Rank	Spearman	Pearson
ClinPath	**0.80**	**1.00**	0.935	**0.983**	**1.20**	0.872	**0.867**
BioClinicalBERT	**0.80**	0.95	**0.940**	0.981	**1.20**	0.857	0.630
PSN (Learned)	0.75	0.95	0.920	0.976	1.25	**0.873**	0.601
Node2Vec	0.75	0.95	0.920	0.976	1.25	0.872	0.601
ClinPathNoGraph	0.70	0.90	0.910	0.971	1.30	0.811	0.795
PSN (Uniform)	0.65	0.90	0.915	0.970	1.40	0.808	0.597
TF-IDF	0.65	0.80	0.905	0.959	1.40	0.790	0.694
SapBERT Concepts	0.35	0.45	0.930	0.844	2.70	-0.038	-0.038

determined between predicted and ground truth similarity scores as shown in Table 2.

Table 2: ClinPath Inference Patient Similarity Evaluation: correlation coefficients

Metric	Result
Spearman Correlation	0.7728
P-value	0.0001698
Pearson Correlation	0.7608
P-value	0.0002461

A Spearman correlation of 0.7728 demonstrated the pipeline's effectiveness in ranking patient pair similarities accurately, and a Pearson correlation of 0.7608 indicated the pipeline's effectiveness in determining the magnitude of similarity scores.

Low p-values demonstrated that these results were not likely a result of randomness. Together, both indicated a strong positive correlation between predicted and ground truth similarity scores, highlighting that this pipeline can be utilized to effectively rank and estimate patient similarity. Consequently, the ClinPath framework can be effectively applied to a wide range of clinical similarity tasks, including clinical trial matching and ICU mortality prediction.

5. LLM Reasoning Extended: Provider Behavior Analysis

5.1. Introduction

In addition to patient-centric analysis, we examined provider orders and behaviors across patients as a complementary direction. Understanding provider-level patterns allows us to detect potential outliers or anomalies in ordering behaviors.

5.2. Experimental Setup

First, all provider actions were extracted from the physician order entry data represented in the knowledge graph. These actions were categorized into distinct clinical groups, including medications, laboratory tests, procedures, and related categories. Each provider's ordering profile was then represented as a collection of these grouped orders. To assess provider-level variation, GPT-4o (OpenAI, 2023) was applied to compare each provider's ordering profile against all others. This comparison highlighted unique elements in ordering patterns and produced a list of providers whose behaviors deviated significantly from their peers.

For evaluation, we constructed a benchmark dataset of 20 provider profiles derived from raw knowledge graph triples. Each profile included associated clinical information such as diagnoses, laboratory tests, prescriptions, and procedures. Ground truth outliers within this cohort were identified through manual curation, with large language models assisting in the annotation process.

Our evaluation focused solely on whether the method could correctly identify the ground truth outliers. The method successfully detected 100% of the provider outliers, demonstrating strong baseline accuracy.

Nonetheless, this application remains exploratory, and future work will focus on developing a more comprehensive and rigorous benchmark.

6. Case Studies

We present two illustrative case studies on patient similarity and provider behavior analysis that highlight the potential clinical impact of use cases enabled by ClinPath.

6.1. Patient Similarity: Pulmonary Embolism

Query Patient (10001234, 68F): Pulmonary embolism with recurrence requiring thrombolysis. *Medications:* Apixaban, Heparin, Alteplase. *Procedures:* CT angiography, Doppler, Echocardiogram.

Patient	Shared Meds	Shared Procedures	Unique Insight
20001234 (74M)	Apixaban, Heparin, Acetaminophen, Ondansetron	CT Angiography, Echocardiogram	Structured follow-up imaging (repeat CT/echo).
30001234 (59F)	Heparin, Apixaban / Rivaroxaban, Acetaminophen, Albuterol	CT Angiography, Echocardiogram	Monitor for infection; consider IV antibiotics if septic PE suspected.
40001234 (66F)	Apixaban, Heparin, Furosemide, Metoprolol	Doppler, Catheter Thrombolysis, IVC Filter	Escalate to advanced interventions if anticoagulation fails.

Table 3: Similar patients to 10001234 with actionable insights.

As seen in Table 3, identifying patients similar to the query patient is valuable. Clinical analyses of these comparable cases can inform the primary patient's care and support subsequent clinical decision-making.

6.2. Provider Behavior Analysis

As an extended downstream application of ClinPath beyond patient-centric analysis, we present a sample case study on provider behavior analysis from our benchmark results.

Outlier Provider	Reason for Outlier Behavior
Provider_P003	Prescribed *Experimental Drug X* and ordered genomic sequencing without standard procedures.
Provider_P007	Relied on obsolete antibiotic and outdated lab panels.
Provider_P010	Used *Unapproved Compound Y* with rare laboratory tests.
Provider_P014	Ordered alternative therapies ("Energy Healing") and nonstandard lab panels.
Provider_P018	Focused on experimental immunotherapy and tumor DNA sequencing, omitting conventional procedures.
Provider_P020	Relied on herbal compounds with nonstandard herbal-related procedures and assays.

Table 4: Outlier providers identified from benchmark dataset with reasons for deviation.

The case study from the benchmark in Table 4 demonstrates how provider-level similarity analysis can flag atypical behavior. Outlier insights are valuable for potentially highlighting unsafe, obsolete, or nonstandard prescribing practices.

7. Conclusion

In this work, we introduce ClinPath, a framework that integrates knowledge graph modeling with LLM reasoning to analyze clinical interactions across time. The constructed ClinKG unifies diagnoses, symptoms, procedures, medications, demographics, and provider interactions into a single representation of patient care. In patient similarity analysis, evaluated on our curated benchmark ClinPath-SimBench, ClinPath surpassed baseline methods across nearly all metrics. The provider behavior study further illustrates the expandability of ClinPath beyond patient-centric tasks. Together, these evaluations show how combining graph-structured representations with LLM reasoning yields clinically meaningful insights. Overall, ClinPath provides a reusable foundation for reasoning over diverse clinical interactions through a multistep integration of knowledge graphs and LLMs.

References

Julián N Acosta, Guido J Falcone, Pranav Rajpurkar, and Eric J Topol. Multimodal biomedical AI. *Nat. Med.*, 28(9):1773–1784, September 2022.

Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78, 2019. doi: 10.18653/v1/W19-1909. URL https://aclanthology.org/W19-1909.

Yaara Artsi, Vera Sorin, Benjamin S. Glicksberg, Girish N. Nadkarni, and Eyal Klang. Advancing clinical practice: The potential of multimodal technology in modern medicine. *Journal of Clinical Medicine*, 13(20):6246, October 2024. ISSN 2077-0383. doi: 10.3390/jcm13206246. URL http://dx.doi.org/10.3390/jcm13206246.

Jiaxin Bai, Tianshi Zheng, and Yangqiu Song. Sequential query encoding for complex query answering on knowledge graphs, 2023. URL https://arxiv.org/abs/2302.13114.

Qiong Cai, Hao Wang, Zhenmin Li, and Xiao Liu. A survey on multimodal data-driven smart healthcare systems: Approaches and applications. *IEEE Access*, 7:133583–133599, 2019. doi: 10.1109/ACCESS.2019.2941419.

Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1), February 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-01960-3. URL http://dx.doi.org/10.1038/s41597-023-01960-3.

Hejie Cui, Jiaying Lu, Ran Xu, Shiyu Wang, Wenjing Ma, Yue Yu, Shaojun Yu, Xuan Kan, Chen Ling, Liang Zhao, Zhaohui S. Qin, Joyce C. Ho, Tianfan Fu, Jing Ma, Mengdi Huai, Fei Wang, and Carl Yang. A review on knowledge graphs for healthcare: Resources, applications, and promises, 2023.

Dominic Girardi, Sandra Wartner, Gerhard Halmerbauer, Margit Ehrenmüller, Hilda Kosorus, and Stephan Dreiseitl. Using concept hierarchies to improve calculation of patient similarity. *Journal of Biomedical Informatics*, 63:66–73, 2016. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2016.07.021. URL https://www.sciencedirect.com/science/article/pii/S1532046416300752.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks, 2016. URL https://arxiv.org/abs/1607.00653.

John Halamka and Paul Cerrato. Knowledge graphs can move healthcare into the future, December 2023. URL https://www.mayoclinicplatform.org/2023/12/21/knowledge-graphs-can-move-healthcare-into-the-future/,note={Accessed:2025-06-06}.

Alistair Johnson, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. Mimic-iv-note: Deidentified free-text clinical notes, 2023. URL https://physionet.org/content/mimic-iv-note/.

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Brian Gow, Benjamin Moody, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV, 2024.

Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5(1), November 2022. ISSN 2398-6352. doi: 10.1038/s41746-022-00712-8. URL http://dx.doi.org/10.1038/s41746-022-00712-8.

Fangyu Liu, Ehsan Shareghi, Zaiqiao Meng, Marco Basaldella, and Nigel Collier. Self-alignment pretraining for biomedical entity representations, 2021. URL https://arxiv.org/abs/2010.11784.

Rajat Mishra and S. Shridevi. Knowledge graph driven medicine recommendation system using graph neural networks on longitudinal medical records. *Scientific Reports*, 14(1), October 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-75784-5. URL http://dx.doi.org/10.1038/s41598-024-75784-5.

A. N. Navaz, H. T. El-Kassabi, M. A. Serhani, A. Oulhaj, and K. Khalil. A novel patient similarity network (psn) framework based on multi-model deep learning for precision medicine. *Journal of Personalized Medicine*, 12(5):768, May 2022. doi: 10.3390/jpm12050768.

OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523, 1988. doi: 10.1016/0306-4573(88)90021-0.

Xuehong Wu, Junwen Duan, Yi Pan, and Min Li. Medical knowledge graph: Data sources, construction, reasoning, and applications. *Big Data Mining and Analytics*, 6(2):201–217, June 2023. ISSN 2097-406X. doi: 10.26599/bdma.2022.9020021. URL http://dx.doi.org/10.26599/BDMA.2022.9020021.

Jiacheng Xu, Kan Chen, Xipeng Qiu, and Xuanjing Huang. Knowledge graph representation with jointly structural and textual encoding, 2016. URL https://arxiv.org/abs/1611.08661.

Xiaolin Zhang and Chao Che. Drug repurposing for parkinson's disease by integrating knowledge graph completion model and knowledge fusion of medical literature. *Future Internet*, 13(1):14, January 2021.

Shuangjia Zheng, Jiahua Rao, Ying Song, Jixian Zhang, Xianglu Xiao, Evandro Fei Fang, Yuedong Yang, and Zhangming Niu. PharmKG: a dedicated knowledge graph benchmark for bomedical data mining. *Brief. Bioinform.*, 22(4), July 2021.