MOD-KG: MultiOrgan Diagnosis Knowledge Graph

Anas Anwarul Haq Khan, Pushpak Bhattacharyya
Indian Institute of Technology Bombay, India
{anaskhan@cse.iitb.ac.in, anas290816007@gmail.com, pb@cse.iitb.ac.in}

Abstract

The human body is highly interconnected, where a diagnosis in one organ can influence conditions in others. In medical research, graphs (such as Knowledge Graphs and Causal Graphs) have proven useful for capturing these relationships, but constructing them manually with expert input is both costly and time-intensive, especially given the continuous flow of new findings. To address this, we leverage the extraction capabilities of large language models (LLMs) to build the ***M**ulti**O**rgan **D**iagnosis **K**nowledge **G**raph* (**MOD-KG**). MOD-KG contains over **21,200 knowledge triples**, derived from both textbooks **(13%)** and carefully selected research papers (with an average of **444** citations each). The graph focuses primarily on the *heart, lungs, kidneys, liver, pancreas, and brain*, which are central to much of today's multimodal imaging research. The extraction quality of the LLM was benchmarked against baselines over **1000** samples, demonstrating reliability.Our dataset is publicly available[1].

1 Introduction

The human body is a deeply interconnected system, where dysfunction in one organ often cascades into effects on others. Capturing these inter-organ relationships in a structured form has long been a goal in medical informatics. Graph-based representations—most notably Knowledge Graphs (KGs) and Causal Graphs (CGs)—have emerged as powerful tools to encode relationships among diseases, risk factors, and treatments. They support exploration of associations, causal pathways, and reasoning across complex medical conditions, and have already been applied in tasks such as *clinical decision support, drug repurposing, treatment discovery, medical imaging report generation, causal drug prioritization, comorbidity network analysis, etc.* Despite their promise, building such graphs remains a bottleneck.

Manual curation requires substantial expert time, struggles to keep pace with the constant influx of biomedical knowledge, and is difficult to scale. To address this, we present the **Multi-Organ Diagnosis Knowledge Graph (MOD-KG)**, comprising **21,200+ triples** extracted from textbooks and high-quality research papers, focusing on six key organs: *heart*, *lungs*, *kidneys*, *liver*, *pancreas*, and *brain*—which are central to many clinical diagnoses and multimodal imaging studies.
MOD-KG enables a wide range of downstream applications:

1. *Diagnostic support*: for example, linking kidney disease with heart failure to prompt cardiovascular monitoring.
2. *Multimodal imaging*: contextualizing CT findings of pulmonary fibrosis with associated liver comorbidities.
3. *Causal reasoning*: tracing pathways such as diabetes $\rightarrow$ kidney disease $\rightarrow$ stroke.
4. *Comorbidity discovery*: uncovering links such as between cirrhosis and hepatic encephalopathy.
5. *Diagnosis omission detection*: flagging overlooked risks, e.g., pneumonia noted in a report but sepsis risk not considered.

Global Patient Safety Report 2024 by WHO[2], notes that most adults will experience at least one diagnostic error in their lifetime and highlights technology based systems as promising interventions. Similarly, (Panagioti et al., 2019) found that 16% of preventable patient harm is linked to diagnostic errors, with diagnosis omission being especially prevalent. *Detailed use cases are in Section 7*

[1] https://github.com/anas2908/MOD-KG

[2] https://www.who.int/publications/i/item/9789240095458

Second Workshop on Integrating NLP and AI for Multilingual and Patient-Centric Healthcare Communication, pages 9–15
December 23, 2025 ©2025 Association for Computational Linguistics

Our work makes the following key contributions:

- We introduce **MOD-KG**, the first large-scale *Multi-Organ Diagnosis Knowledge Graph*, consisting of over 21,200 high-quality knowledge triples covering six critical organs (heart, lungs, kidneys, liver, pancreas, and brain).
- We propose a pipeline for extracting medical knowledge triples from textbooks and research papers, benchmark the extraction quality against baseline methods over 1000 samples, and release MOD-KG along with all associated metadata for the community.

2 Related Work

Biomedical knowledge graphs (BKGs) integrate diverse sources such as databases, ontologies, and literature to represent entities (e.g., diseases, drugs, genes) and relations, supporting applications like question answering, drug repurposing, and decision support via path-based or embedding-based reasoning (Zhu et al., 2020; Lu et al., 2025; Arsenyan et al., 2024). In drug discovery, KG-based approaches leverage drug–disease–gene networks with path, embedding, and causal methods to prioritize candidates and explain mechanisms, exemplified by RPath (Zhu et al., 2020; Ma et al., 2023a; Zhu et al., 2023; Domingo-Fernández et al., 2022). In radiology and multimodal medicine, organ- or modality-specific KGs enhance vision–language models for accurate report generation (Kale et al., 2023b,a), while automated extraction pipelines (e.g., SemMedDB, SemRep, PubTator) and hybrid rule–ML methods improve coverage and precision for specialized biomedical relations (Kilicoglu et al., 2020; Wei et al., 2019; Lai et al., 2023; Pawar et al., 2021). Large language models have further enabled zero/few-shot and ontology-guided triplet extraction pipelines for text-to-KG construction, reducing annotation costs but facing challenges in calibration, factuality, and entity standardization (Papaluca et al., 2024; Mo et al., 2025; Khorashadizadeh et al., 2024).

MOD-KG distinguishes itself as an organ-centric graph encoding both intra- and inter-organ relations, automatically extracted from curated textbooks and highly cited research, with **21.7k triples** across six major organs, supporting applications in imaging context, comorbidity discovery, and omission detection.

3 MultiOrgan Diagnostic Knowledge Graph (MOD-KG)

3.1 Definition and representation

We represent inter- and intra-organ diagnostic knowledge initially as *quintuples* of the form

$$Q = \langle d_1, o_1, r, d_2, o_2 \rangle,$$

where d_i is a diagnosis (or clinical concept), o_i is the organ in which d_i occurs, and r is a relation (e.g., "may cause", "is associated with", "increases risk of"). Quintuples explicitly bind each diagnosis to an organ, which reduces ambiguity when the same diagnosis label can appear in multiple anatomical contexts.

For graph construction we map each quintuple to a canonical *triple* by collapsing the diagnosis+organ pair into a single node identifier via a canonicalization function $c(\cdot,\cdot)$:

$$\begin{aligned} Q &= \langle d_1, o_1, r, d_2, o_2 \rangle \\ &\longrightarrow \; t = \langle h, r, t \rangle \\ &\text{with } h = c(d_1, o_1), \quad t = c(d_2, o_2). \end{aligned}$$

The set of all canonical entities (nodes) is denoted $\mathcal{E}$ and the set of relation types is $\mathcal{R}$. The resulting knowledge graph is

$$\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}),$$

where $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ is the set of extracted triples. Representative intra- and inter-organ triples are shown in Table 1 and node examples in Table 2.

3.2 Relation to existing organ-centric work and embedding strategy

Organ-centric KGs have been shown useful for multimodal clinical tasks; in particular, Kaveri Kale *et al.* construct abdominal-organ knowledge representations and demonstrate benefits when these triples are injected into vision–language pipelines for radiology report generation (Kale et al., 2023b,a). Following the same spirit of converting structured text extractions into an embeddable graph, we produce **MOD-KG** triples and compute translational embeddings using TransE (Bordes et al., 2013) so that MOD-KG is immediately amenable to downstream neural integration.

Concretely, for each triple $(h, r, t) \in \mathcal{T}$ we learn low-dimensional vectors $\mathbf{e}_h, \mathbf{e}_t, \mathbf{r} \in \mathbb{R}^k$ with the TransE scoring function

$$f(h, r, t) = \|\mathbf{e}_h + \mathbf{r} - \mathbf{e}_t\|_2,$$

Inter Organ Triples	**Intra Organ Triples**
Cirrhosis in **Liver**, may cause, Impaired Ventricular Ejection in **Heart**	Pneumothorax in **Lung**, may cause, Hypoxia in **Lung**
NAFLD in **Liver**, is associated with, Myocardial infarction in **Heart**	COPD in **Lung**, contributes to, Emphysema in **Lung**
COPD in **Lung**, may lead to, Glomerular Injury in **Kidney**	Honeycomb **Lung**, associated with, Rheumatoid Arthritis in **Lung**
Osteoporosis in **Bone**, is related to, Emphysema in **Lung**	Tumor embolism in **Heart**, associated with, Mild cardiomegaly in **Heart**
Emphysema in **Lung**, linked to, Elastolytic changes of the **skin**	Valvular **Heart** disease, may cause, Hypoeffective **Heart**
Severe PLD in **Liver**, may cause, Elevation in **Diaphragm**	Aortic Regurgitation in **Heart**, may cause, Diastolic Murmur in **Heart**
Sarcoidosis in **Spleen**, can involve, Cardiac Sarcoidosis in **Heart**	Glomerulonephritis in **Kidney**, may lead to, Chronic Inflammation in **Kidney**
Type 2 diabetes in **Pancreas**, is associated with, Reduced **Lung** function	Portal Hypertension in **Liver**, can lead to, Ascites in **Liver**
Cancer in **Bladder**, may cause, Aortic endocarditis in **Heart**	Pancreatitis in **Pancreas**, may be caused by, ERCP in **Pancreas**
Drooling in **Mouth**, may lead to, Aspiration in **Lung**	Neurovascular dysfunction in **Brain**, may cause, Oligemia in **Brain**

Table 1: Example intra- and inter-organ knowledge triples.

Source Diagnosis	Inter-Organ Relation	Inter-Organ Target Diagnosis	Intra-Organ Relation	Intra-Organ Target Diagnosis
Liver Cirrhosis	may cause	Cardiac Dysfunction in **Heart**	may induce	Cardiac **Liver** cirrhosis
	may cause	Q-T Interval Prolongation in **Heart**	may lead to	Portal Hypertension
	may be associated with	decreased heart rate variability in **Heart**	may lead to	Biliary Cyst (BC)
	may be involved in	Cirrhotic Cardiomyopathy in **Heart**	may lead to	Fibrosis in **Liver**
	may cause	Biliary Cyst (BC) in **Gallbladder**	may be caused by	Chronic Hepatitis B (CHB)
	may be associated with	Pulmonary hypertension in **Lung**	may be caused by	Hepatocellular Necrosis
	may lead to	Hepatorenal Syndrome in **Kidney**	may be caused by	Hepatocellular Regeneration
Sarcoidosis in **Lung**	may involve	Cardiac Sarcoidosis in **Heart**	may cause	Pleural effusions
	may accumulate in	Hilar Lymph Node Sarcoidosis	is similar to	Talc granulomatosis
	may cause	Congestive heart failure	may be associated with	Pulmonary hypertension
	may cause	Pulmonary Hypertension in **Heart**	may lead to	Pneumothorax in **Lung**
	may cause	Granulomatous Vasculitis in **Heart**	increase risk of	Pulmonary embolism
	may cause	Right Ventricular Hypertrophy in **Heart**	may cause	Aspergillus **Lung** disease
	may cause	Cardiac Involvement in **Heart**	may cause	Bronchiectasis

Table 2: Organ-centric source–target triple examples.

Description	**Statistics**
Total Triples Curated	21770
Redundant Triples	564
Total Unique Triples	21206
Number of Intra-Organ Triples	16039
Number of Inter-Organ Triples	5167
Number of Unique Relation in Triples	2794
Number of Unique Diagnosis in Triples	20581
Number of Unique Organs	62

Table 3: Summary statistics of MOD-KG triples.

trained using a margin ranking loss with negative sampling (standard TransE procedure). These embeddings (stored for all $h, t \in \mathcal{E}$) convert MOD-KG from a collection of symbolic triples into a continuously parameterized graph representation. In future work the learned node/edge features can be consumed by graph neural modules (e.g., Graph Attention Networks, GATs (Veličković et al., 2018)) and injected into model decoders (via cross-attention or concatenated latent features) for tasks such as multimodal generation or graph-aware reasoning.

3.3 Methodology

Corpus curation & target coverage We curated a high-quality corpus of **422** well-cited research papers (avg. 444 citations) from **219** distinct journals, covering **109** clinically relevant conditions across the target organs. The organ keyword set used for retrieval and filtering is summarized in Table 5, and the per-organ frequency distribution is reported in Table 6.

Segmentation and chunking Each document was segmented into overlapping chunks to reduce boundary artifacts during extraction. We used chunks of length 300 tokens with a 100-token overlap (heuristically chosen through pilot experiments). This segmentation balances local context size with the need to avoid splitting relations across chunk boundaries.

Prompted LLM extraction (2-shot) The Prompt used in extraction is mentioned in the section 8. The extraction output examples and selected triples are shown in Table 1.

Post-processing and canonicalization Raw quintuples were normalized and canonicalized before conversion to triples. Canonicalization included (i) string normalization, (ii) mapping high-confidence synonyms to a single canonical node label, and (iii) light clustering to unify near-duplicate entities arising from surface variation. After canonicalization each quintuple was mapped to a triple as shown above and duplicate triples were collapsed.

Embedding and storage The deduplicated triple set $\mathcal{T}$ (MOD-KG) was embedded with TransE to produce node and relation vectors for all canonical entities and relations. These embeddings are stored alongside the symbolic graph, enabling either (i) direct graph-based queries over $\mathcal{G}$ or (ii) neural consumption (e.g., as initial node features for GATs) for downstream models.

Summary Statistics Table 3 presents the overall statistics of **MOD-KG**. Out of 21,770 curated triples, 564 were redundant, yielding 21,206 unique triples. The graph captures both *intra-organ* (16,039) and *inter-organ* (5,167) relations, spanning 2,794 unique relation types, 20,581 unique diagnoses, and 62 organ categories. These numbers highlight the medium scale of MOD-KG while ensuring high coverage across diverse diagnostic contexts.

4 Extraction Evaluation

The quality of a knowledge graph is fundamentally constrained by the quality of its extraction pipeline. Since **MOD-KG** was curated from well-cited papers and textbooks sourced from reputable journals and publishers, the limiting factor becomes the accuracy of the extraction itself. We therefore systematically evaluated whether large language model (LLM)–based extraction, specifically GPT-4o (Achiam et al., 2023), can reliably operate in the medical domain.

Setup. We compared GPT-4o extraction against classical IE pipelines, including *spaCy*, *DREEAM* (Ma et al., 2023b), and *OpenIE* (Vasiliev, 2020; Zhou et al., 2022). For each method we sampled **1000 quintuples**, stratified across organs, and asked a practicing medical doctor to annotate correctness with respect to both medical faithfulness and relation accuracy. This provided a controlled human benchmark for extraction quality.

Results. Table 4 summarizes the comparative results. GPT-4o achieved the highest faithfulness, substantially outperforming both heuristic IE baselines and the smaller LLM. Classical pipelines often failed to capture domain-specific terminology or produced fragmented triples. In contrast, GPT-4o consistently generated medically coherent relations, though with some errors in rare disease contexts.

Cost. The full extraction across the corpus required approximately **$730** of OpenAI API usage for GPT-4o, which was acceptable given the quality gains relative to baselines.

Method	**Faithfulness (% correct)**
GPT-4o (ours)	96.2
spaCy	39.1
DREEAM	48.9
OpenIE	66.9

Table 4: Faithfulness comparison of extraction methods (1000-sample evaluation with human annotation). GPT-4o achieves the highest medical accuracy.

5 Conclusion

In this work, we presented **MOD-KG**, a multi-organ diagnostic knowledge graph constructed from high-quality biomedical corpora, comprising both textbooks and well-cited research papers. By extracting quintuples and converting them into triples, MOD-KG captures both *intra-* and *inter-organ* relationships across six major organ systems. Through post-processing and embedding with TransE, we produced a resource that is both interpretable and readily usable for neural consumption. Our evaluation, based on 1000 expert-annotated samples, demonstrated that GPT-4o substantially outperforms classical IE pipelines in medical extraction quality, albeit at a higher computational cost.

6 Limitations and Ethical Considerations

MOD-KG, built from high-quality textbooks and research papers, is limited by the scope of its source corpus, which may omit rare conditions, emerging knowledge, or community-specific diagnostic practices. Although LLM-based extraction achieves high accuracy, it can occasionally hallucinate, particularly for underrepresented terminologies, and decisions may collapse medical subtypes into broader categories. As a research resource, not a clinical decision support system, MOD-KG is not intended for direct patient care. Additionally, biases present in published literature, such as overrepresentation of certain populations, diseases, or treatment paradigms, may propagate into the graph. Therefore, its use is intended for research, benchmarking, and as a substrate for developing multimodal models.

References

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

Vahan Arsenyan, Spartak Bughdaryan, Fadi Shaya, Kent Wilson Small, and Davit Shahnazaryan. 2024. Large language models for biomedical knowledge graph construction: information extraction from emr notes. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 295–317.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Daniel Domingo-Fernández et al. 2022. Rpath: causal reasoning over knowledge graphs leveraging transcriptomic signatures for drug prioritization. *PLOS Computational Biology*. PMC8906585.

Kaveri Kale, Pushpak Bhattacharyya, Milind Gune, Aditya Shetty, Rustom Lawyer, et al. 2023a. Kgvl-bart: Knowledge graph augmented visual-language bart for radiology report generation. In *EACL / ACL Workshop or Proceedings (conference version)*. See ACL Anthology / authors' project pages for PDF and bib.

Kaveri Kale, Pushpak Bhattacharyya, Aditya Shetty, Milind Gune, Kush Shrivastava, Rustom Lawyer, and Spriha Biswas. 2023b. "knowledge is power": Constructing knowledge graph of abdominal organs and using them for automatic radiology report generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Industry Track)*, pages 11–24.

Hanieh Khorashadizadeh, Fatima Zahra Amara, Morteza K. Ezzabady, Frédéric Ieng, Sanju Tiwari, Nandana Mihindukulasooriya, Jinghua Groppe, Soror Sahri, Farah Benamara, and Sven Groppe. 2024. Research trends for the interplay between large language models and knowledge graphs. *arXiv / VLDB workshop proceedings*. ArXiv:2406.08223; VLDB workshop LLM+KG version available.

Halil Kilicoglu, Graciela Rosemblat, Marcelo Fiszman, and Dongwook Shin. 2020. Broad-coverage biomedical relation extraction with semrep. *BMC Bioinformatics*, 21:188.

Po-Ting Lai et al. 2023. Biorex: a rich biomedical relation extraction dataset. *Journal of Biomedical Informatics / arXiv*. See: BioRED / BioREx resources; DOI / arXiv entry.

Yuxing Lu, Sin Yee Goi, Xukai Zhao, and Jinzhuo Wang. 2025. Biomedical knowledge graph: A survey of domains, tasks, and real-world applications. *arXiv preprint*. ArXiv:2501.11632.

Chunyu Ma, Zhihan Zhou, Han Liu, and David Koslicki. 2023a. Kgml-xdtd: A knowledge graph-based machine learning framework for drug treatment prediction and mechanism description. *GigaScience*, 12:giad057. Preprint / arXiv:2212.01384.

Youmi Ma, An Wang, and Naoaki Okazaki. 2023b. Dreeam: Guiding attention with evidence for improving document-level relation extraction. *arXiv preprint arXiv:2302.08675*.

Belinda Mo, Kyssen Yu, Joshua Kazdan, Proud Mpala, Lisa Yu, et al. 2025. Kggen: Extracting knowledge graphs from plain text with language models. *arXiv preprint*. ArXiv:2502.09956.

Maria Panagioti, Kanza Khan, Richard N Keers, Aseel Abuzour, Denham Phipps, Evangelos Kontopantelis, Peter Bower, Stephen Campbell, Razaan Haneef, Anthony J Avery, et al. 2019. Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *bmj*, 366.

Andrea Papaluca, Daniel Krefl, Sergio Méndez Rodríguez, Artem Lensky, Hanna Suominen, et al. 2024. Zero- and few-shots knowledge graph triplet extraction with large language models. *Proceedings / arXiv*. See ACL / arXiv entry for PDF and bib.

Sachin Pawar, Ravina More, Girish K. Palshikar, Pushpak Bhattacharyya, and Vasudeva Varma. 2021. Knowledge-based extraction of cause-effect relations from biomedical text. *arXiv preprint*. ArXiv:2103.06078.

Yuli Vasiliev. 2020. *Natural language processing with Python and spaCy: A practical introduction*. No Starch Press.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *6th International Conference on Learning Representations (ICLR 2018)*.

Chih-Hsuan Wei, Alexis Allot, Robert Leaman, and Zhiyong Lu. 2019. Pubtator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Research*, 47(W1):W587–W593.

Shaowen Zhou, Bowen Yu, Aixin Sun, Cheng Long, Jingyang Li, Haiyang Yu, Jian Sun, and Yongbin Li. 2022. A survey on neural open information extraction: Current status and future directions. *arXiv preprint arXiv:2205.11725*.

Chaoyu Zhu et al. 2023. Rdkg-115: a knowledge graph to assist drug repurposing for rare diseases. *Computers in Biology and Medicine*. Example/representative KG-for-drug-repurposing work (RDKG family / 2023).

Yongjun Zhu, Chao Che, Bo Jin, Ningrui Zhang, Chang Su, and Fei Wang. 2020. Knowledge-driven drug repurposing using a comprehensive drug knowledge graph. *Health Informatics Journal*.

Appendix

7 Use Cases of MOD-KG

The **MultiOrgan Diagnostic Knowledge Graph (MOD-KG)** offers a structured representation of inter- and intra-organ diagnostic relationships, making it applicable to a wide range of clinical and computational settings. Below, we outline several key use cases where MOD-KG can contribute to improved results and insights.

7.1 Diagnostic Omission Detection

A common challenge in clinical workflows is the inadvertent omission of potential diagnoses. By encoding *inter-organ dependencies* (e.g., "Liver Cirrhosis → Kidney Failure"), MOD-KG can flag missing diagnoses in structured or free-text reports. For example, if a patient record documents *Cirrhosis* but omits possible *Renal Dysfunction*, MOD-KG can highlight the omission, prompting physicians to investigate further. This can reduce diagnostic errors and improve patient safety.

7.2 Multimodal Imaging Report Augmentation

MOD-KG can be paired with vision–language models for radiology report generation. For instance, in chest X-ray interpretation, if a model predicts *Cardiomegaly*, MOD-KG can suggest related findings such as *Pulmonary Edema* or *Pleural Effusion*, thereby producing more complete and consistent reports. Such augmentation mirrors the use of organ-centric KGs in models like KGVL-BART (Kale et al., 2023b,a), but extends coverage across multiple organs.

7.3 Comorbidity Analysis and Patient Stratification

By representing co-occurrence and causal relationships among diagnoses, MOD-KG can support stratification of patient cohorts. For example, in a hospital database, patients diagnosed with *Diabetes Mellitus* and *Hypertension* can be linked to MOD-KG's paths leading to *Chronic Kidney Disease*, enabling earlier identification of at-risk populations. This is particularly valuable for designing preventive interventions and population-scale studies.

7.4 Causal Reasoning in Disease Progression

MOD-KG encodes not only co-occurrence but also *directional relationships*. This enables causal reasoning over progression paths. For instance, a path such as *Hypertension → Left Ventricular Hypertrophy → Heart Failure* allows models to infer plausible progressions and to simulate hypothetical interventions. This could support clinical decision-making by providing mechanistic explanations rather than surface-level associations.

7.5 Clinical Decision Support Systems (CDSS)

CDSS often rely on isolated rules or black-box predictions. MOD-KG provides an interpretable layer of structured knowledge that can complement predictive models. For example, when a CDSS flags a risk of *Stroke*, MOD-KG can provide context by surfacing associated conditions such as *Atrial Fibrillation* or *Carotid Atherosclerosis*. This improves both physician trust and actionability of CDSS outputs.

7.6 Education and Training

Medical students and residents often struggle with connecting knowledge across organ systems. MOD-KG can serve as a visual and interactive learning resource, showing how diagnoses in one system cascade into others (e.g., *COPD in Lungs → Pulmonary Hypertension → Right Heart Failure*). This supports a systems-based approach to clinical education.

7.7 Foundation for Multimodal Extensions

Beyond text, MOD-KG could be extended to integrate imaging or lab-test signals. For example, embedding MOD-KG into a multimodal pipeline could allow a model to jointly reason over lab abnormalities (e.g., elevated creatinine), imaging findings (e.g., renal cysts), and clinical diagnoses, providing a holistic diagnostic assistant.

8 LLM Extraction Prompt

Extraction was performed with an LLM using a 2-shot prompting strategy. For each chunk we asked the model to emit structured quintuples in a fixed JSON format.

```
Please analyze the following text to
identify organ-to-organ diagnosis
relationships,whether they occur between
different organs or within the same organ,
using only the information provided in the
text. Structure the output strictly in the
JSON format specified below with a dummy
2-shot example. If no such relationships
can be derived from the text, return an
```

Heart Keywords	**Lungs Keywords**	**Kidney Keywords**	**Liver Keywords**	**Brain Keywords**	**Pancreas Keywords**
pericarditis	COPD	acute kidney injury	hepatitis b	Encephalitis	cystic fibroma
angina pectoris	asthma	alport syndrome	cirrhosis	Huntington’s disease	pancreatic cancer
atrial fibrillation	emphysema	amyloidosis	liver cancer	Epilepsy	pancreatitis
hypertension	chronic bronchitis	ADPKD	fatty liver	Cerebral palsy	hemorrhagic pancreatitis
cardiomyopathy	pneumonia	ESRD	liver fibrosis	Diabetic neuropathy	glucagonoma
heart failure	pulmonary hypertension	FSGS	hemochromatosis	Vascular dementia	diabetes mellitus
endocarditis	pulmonary embolism	chronic kidney disease	wilsons disease		ascites
myocardial infarction	goodpasture syndrome	HUS	gilbert syndrome		annular pancreas
tetralogy of fallot	lung cancer	HSP	crigler-najjar syndrome		pancreatic agenesis
coronary heart disease	pneumothorax	hypertensive nephrosclerosis	primary biliary cholangitis		pancreatic fistula
mitral valve regurgitation	cystic fibrosis	lupus nephritis	drug-induced liver injury		
atrial septal defect	pleuritis	kidney cancer	amebic liver abscess		
tricuspid regurgitation	hydropneumothorax	kidney stones	portal vein thrombosis		
pulmonary embolism	silicosis	nephrotic syndrome	caroli’s disease		
ventricular septal defect	histoplasmosis	obstructive nephropathy	choledochal cysts		
cardiac sarcoidosis	bronchiectasis	vasculitis	polycystic liver disease		
patent foramen ovale	ARDS	pyelonephritis	viral hepatitis d		
patent ductus arteriosus	tuberculosis	post-cystic kidney disease	budd-chiari syndrome		
wolff-parkinson white syndrome	pulmonary sarcoidosis	papillary necrosis	acute hepatic failure		
cardiac tamponade	pulmonary hypertension	proteinuria	hepatoblastoma		
aortic stenosis	cor pulmonale		hepatitis e		
mitral valve prolapse	mesothelioma				
cardiomegaly	atelectasis				
enlarged cardiomediastinum	consolidation				
	edema				
	lung lesion				
	lung opacity				
	pleural effusion				

Table 5: Keywords used for MOD-KG corpus curation.

Organ	**Frequency**	**Organ**	**Frequency**	**Organ**	**Frequency**	**Organ**	**Frequency**
Heart	16044	**Kidney**	6401	**Lung**	5676	**Liver**	4833
Brain	3774	**Pancreas**	1960	Skin	423	Eye	342
Bone	302	Skeletal Muscle	246	Thyroid	244	Stomach	211
Artery	165	Spleen	160	Joint	117	Nose	115
Colon	102	Bladder	100	Spinal Cord	97	Adrenal Gland	96
Testis	78	Hypothalamus	70	Bone Marrow	67	Uterus	65
Small Intestine	62	Gallbladder	60	Cerebellum	43	Nerve	39
Mouth	39	Vein	38	Pituitary Gland	38	Diaphragm	34
Ovary	32	Cervix	31	Lymph Node	30	Bronchus	27
Ear	25	Large Intestine	25	Prostate	24	Rectum	24
Parathyroid Gland	18	Salivary Gland	16	Ureter	14	Penis	13
Tooth	13	Placenta	12	Mesentery	8	Appendix	8
Capillary	8	Scrotum	8	Vagina	6	Fallopian Tube	6
Larynx	5	Subcutaneous Tissue	5	Urethra	4	Nasal Cavity	2
Trachea	2	Tonsil	1	Pharynx	1	Nail	1
Seminal Vesicle	1	Tongue	1	Others	0		

Table 6: Organ-wise distribution of entities in MOD-KG.

```
empty JSON object.
[
  {
    "organ1": "Heart",
    "diagnosis1": "Pericarditis",
    "relation": "may cause",
    "organ2": "Lungs",
   "diagnosis2": "Retrosternal Chest Pain"
  },
  {
    "organ1": "Heart",
   "diagnosis1": "Pericardial Effusion",
    "relation": "may lead to",
    "organ2": "Heart",
    "diagnosis2": "Cardiac Tamponade"
  }
]
```

We operated the extractor at the chunk level across the corpus and collected the resulting quintuples for downstream processing. We used GPT-4o as the extraction engine and compared its output against heuristic and classical IE pipelines (e.g., spaCy, DREEAM, OpenIE) and literature mining baselines (SemRep / PubTator) to guide our choice of extractor (Achiam et al., 2023; Vasiliev, 2020; Ma et al., 2023b; Zhou et al., 2022; Kilicoglu et al., 2020; Wei et al., 2019).