Medical Knowledge Harmonization: A Graph-based, Entity-Selective Approach to Multi-source Diagnoses

Andrea Bianchi
DISIM Department
University Of L'Aquila
L'Aquila, Italy
andrea.bianchi@graduate.univaq.it

Antinisca Di Marco
DISIM Department
University Of L'Aquila
L'Aquila, Italy
antinisca.dimarco@univaq.it

***Abstract*—The paper discusses a novel system for medical diagnostics that integrates patient data from various sources to address the fragmentation of healthcare information. By generating and merging knowledge graphs from raw medical texts focused on key biomedical entities (Gene, Disease, Chemical, Species, Mutation, Cell Type), the system facilitates a comprehensive understanding of a patient's medical history. It accurately extracts and connects critical entities, creating individual and combined knowledge graphs that elucidate a patient's medical journey. This approach helps bridge diagnostic gaps, offering a visual tool for practitioners to detect patterns and discrepancies in patient data. Despite limitations such as language dependency and validation scope, this system sets the stage for future enhancements toward a more universally accessible and clinically useful healthcare system.**

***Index Terms*—medical diagnostics, multi-source diagnosis**

I. Introduction

In modern healthcare systems, a patient often consults with multiple specialists across different institutions, leading to multiple diagnostic records. These records, though rich in information, can often be fragmented and inconsistent [1]. As a result, for chronic or complex illnesses, a single individual may have many diagnoses, sometimes different and spanning different time periods and institutions. While this multitude of data sources should, in theory, provide a comprehensive view of a patient's health, it often results in the opposite: a fragmented, and occasionally contradictory puzzle of information [2]. This overwhelming and fragmented landscape of patient data can lead to gaps in understanding, potentially causing misdiagnoses, redundant testing, and even treatment errors [3]. Knowledge graph (KG) is a systematic way to connect information and data points to knowledge. These graphs may effortlessly combine intricate patient data in the context of medical diagnostics, making them an appropriate solution for managing discussed challenges [4].

This paper introduces an approach to tackle the problem of multi-source diagnostic data integration, a process that involves combining diagnostic information from various healthcare sources to create a cohesive patient health profile. The problem is intriguing because resolving it has the potential to significantly enhance diagnostic accuracy and treatment efficacy. It's particularly vital in genetic information and rare diseases, where integrating scattered and specialized data can lead to breakthroughs in understanding and treatment. While previous efforts have made strides in improving data quality and developing data exchange standards, they often fall short in addressing the semantic integration of complex medical data comprehensively. Our work aims to bridge this gap by not only generating but also merging knowledge graphs from various diagnostic sources, thereby offering a panoramic and unified view of a patient's medical history. This approach stands to revolutionize how medical professionals access, interpret, and utilize patient data for more informed decision-making.

In light of these considerations, the primary contributions of this work are framed around three key research questions: (*RQI*) How can individual knowledge graphs be generated from raw medical texts? (*RQII*) What mechanism allows for the merging of these individual graphs while highlighting unique entities? And (*RQIII*) How can a visualization tool assist medical professionals in understanding a patient's comprehensive medical history? Addressing these questions, our paper outlines the methodology for generating and merging knowledge graphs, followed by an exploration of a visualization tool designed for medical professionals.

The paper is structured as follows: Section II reviews related works, while Section III presents the motivating scenario behind our work. The technical details of our approach are explored in Section IV while Section V and Section VI discuss the experimental settings and results, respectively. Section VII provides a discussion on our findings. Finally, we conclude and discuss future work in the Section VIII.

II. Related Works

Different approaches and goals have been seen in the field of building knowledge graphs from medical and biological texts. In order to provide a more comprehensive representation of medical situations, some research projects aim to augment textual data with multiple notations that include genetics, proteomics, symptoms, and more [5] [6]. Others are focused on

 ISBN: 978-1-68558-136-7

developing knowledge graphs that are specialised to particular illness types and provide in-depth insights into their complex dynamics [7]. Moreover, some initiatives, such as [8] and [9], aim to generate knowledge graphs straight from spoken dialogues or utterances recorded in-context clinical encounters. Authors of [8] proposed a method to construct a medical knowledge graph directly from clinical conversations between doctors and patients. Unlike this work, our approach aim at providing a unified visualization that emphasizes patient's whole medical journey rather than predictive analysis from singular clinical conversations. PrimeKG [5] serves as a multimodal knowledge graph for precision medicine, integrating data from 20 resources to offer insights across ten biological scales, from protein perturbations to therapeutic drug actions. [6] introduces the Clinical Knowledge Graph (CKG), an expansive platform designed to integrate diverse biomedical data, including proteomics, to facilitate precision medicine. CKG, encompassing over 16 million nodes and 220 million relationships, aims to represent experimental data, public databases, and literature while implementing advanced statistical and machine learning tools to enhance proteomics workflows. Differently from [5] and [6], our research is tailored towards unifying diagnostic data from multiple healthcare centres, providing a comprehensive visual picture of a patient's medical trajectory.

III. Motivating Scenario

Consider a scenario where distinct diagnostic reports, generated at different times and by different institutions, capture varied aspects of a patient's health. A report from one hospital might highlight specific findings that were either not observed or not considered pertinent in another [10], [11]. The proposed system ingests diagnostic texts from various sources and generates individual knowledge graphs. These graphs, each representing a unique diagnostic perspective, are then merged into a unified knowledge graph, as illustrated in Figure 1. This integrated visualization accentuates common entities and relationships using consistent colours and distinctly highlights unique entities or pieces of evidence from each diagnostic source. By offering this consolidated view, healthcare professionals are equipped with a panoramic understanding of an individual's health trajectory. This enables more informed decisions, ensures no detail is missed, and potentially avoids redundant or misguided medical interventions, ensuring the best possible patient care and improving personalized medicine [5].

IV. Medical Knowledge Harmonization

The goal of our system is to transform fragmented diagnostic texts into a unified knowledge graph, providing a holistic understanding of a patient's medical history. This transformation is achieved through a series of systematic steps, as depicted in Figure 1, General Workflow. The figure delineates our workflow through four pivotal macro-steps: 1) Named Entity Recognition (NER), where entities are identified from the raw texts; 2) Relationship Extraction (RE), where relationships between identified entities are extracted; 3) Single Source Graph Generation, which involves creating individual knowledge graphs for each diagnostic source of each patient; and 4) Knowledge Graph Integration, where these individual graphs are amalgamated into a unified, comprehensive knowledge graph.

A. Input Source Determination and Preprocessing

The system processes multiple diagnostic texts from varied healthcare environments (Figure 1, Tools), reflecting different stages of a patient's medical history. It operates in two modes: **Data Ingestion Mode**, which uses a structured dataset to generate and integrate knowledge graphs, and **Manual Mode**, where users manually input diagnostic reports for ad-hoc analysis. In Manual Mode, reports are uploaded to a specific folder (*diagnostic_reports*), and the system then extracts and integrates data into the knowledge graph, similar to the Data Ingestion Mode. This flexibility allows for both comprehensive and targeted analyses of patient diagnostics.

B. Entity Recognition and Normalization

Each diagnostic text (T1, T2, ... Tn) undergoes NER to identify medically relevant entities. This step utilizes NER techniques and tailored for medical and biological texts, ensuring accurate extraction of entities. For this critical task, our system employs BERN2 [12], a state-of-the-art tool in the biomedical domain, which is capable of recognizing and normalizing nine different entities: Gene, Disease, Chemical, Species, Mutation, Cell Line, Cell Type, DNA, and RNA. BERN2 adopts distinct strategies for multi-task NER, ensuring accurate extraction of entities by navigating through the intricate and domain-specific language of medical and biological texts. Subsequent to the entity recognition, BERN2 proceeds with the normalization of these entities, utilizing dedicated methods that enhance the precision and reliability of the identified entities within the diagnostic texts.

C. Relation Extraction

After the entities have been recognized and normalized, the system advances to the RE stage, which aims to decipher the relationships between the identified entities within the diagnostic text. For this endeavor, we use the capabilities of Bio_ClinicalBERT [13]. Bio_ClinicalBERT is a model developed for processing clinical text. It combines BioBERT's pretraining on biomedical literature with further training on MIMIC-III notes, a database of electronic health records from ICU patients. The model, trained on a variety of notes, is designed to capture the nuances of clinical language [14]. Despite not being originally designed to discern relationships between entities, the embeddings from Bio_ClinicalBERT, enriched with substantial biomedical and clinical contextual information, can be leveraged to infer potential relationships among the identified entities through a heuristic approach. It's worth noting that our experiments also leverage the MIMIC database, aligning our experimental setup with the intrinsic knowledge and understanding embedded within Bio_ClinicalBERT, thereby ensuring a coherent setting.

 ISBN: 978-1-68558-136-7

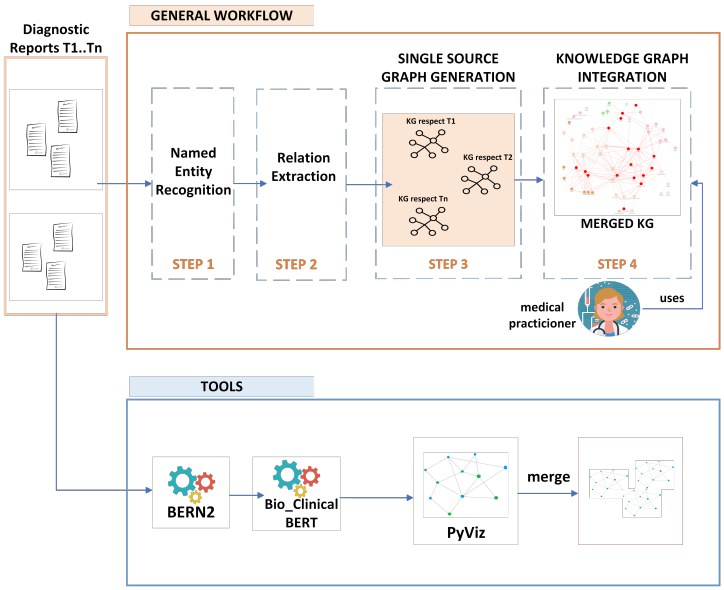

Figure. 1. High-Level Workflow for the system.

D. Knowledge Graph Generation

After extracting entities and their respective relationships, the system leverages on these to construct individual knowledge graphs for each diagnostic text, utilizing entities as nodes and their relationships as edges to graphically illustrate the information embedded within each text. Following the generation of these individual knowledge graphs, the system goes to the integration phase, wherein it amalgamates these multiple graphs into a unified knowledge graph. This consolidated graph stands as a coherent synthesis of information, amalgamating insights from all diagnostic sources and providing a comprehensive visual depiction of a patient's entire medical history. The visually integrated knowledge graph also highlights common entities and relationships with consistent colours.

V. Experimental Settings

Here we delve into the specifics of how our research was conducted, ensuring transparency and reproducibility.

A. Hardware Configuration

The study utilized the Caliban cluster at the University of L'Aquila, which has multiple nodes with 40 processing units for parallel execution in the "mpi" environment. The tests ran on a CentOS Linux 7.4.1708 system with an Intel Xeon E5-2698 v4 CPU at 2.20GHz and 141GB RAM.

B. Dataset

Our study uses the MIMIC-IV-Note dataset (version 2.2), featuring *331,794* discharge summaries and *2,321,355* radiology reports, all de-identified for patient confidentiality ([14]). We focus on discharge summaries to analyze patients' medical histories. For ethics and replication, our dataset and code are available at PhysioNet [14] and [15], respectively. To replicate our preprocessed dataset, specific steps are required.

- **Filtering for Discharge Notes:** We selected *discharge* notes from our database for their detailed summaries of hospital stays, including diagnoses, treatments, and medical histories.
- **Extracting *History of Present Illness*:** We used regex parsing to extract this section, providing a detailed narrative of the patient's condition at the time of a specific hospitalization.
- **Adapting for Multiple Hospitalizations:** For patients with several hospital stays, we adjusted the data structure to isolate each hospitalization, enabling analysis of medical condition progression across visits.
- **Selecting Patients with Multiple Diagnoses:** Our dataset only includes patients with multiple diagnoses to focus on complex or rare medical histories.

This process resulted in a dataset of 59,051 unique patients, each with detailed hospitalization records and *'History of Present Illness'*.

C. Software Configuration

For the **NER step** we used **(BERN2 [12])**, an advanced biomedical entity recognition service. The *load_bern2_model* function processes diagnostic reports to extract and structure named entities for further use. For the **RE step** we selected *(Bio_ClinicalBERT [13])*, a variant of BERT that's specialized for clinical and biological texts. This model's embeddings are pivotal in our approach to relation extraction. For each pair of entities in a report, embeddings are generated. Then, the cosine similarity between entity pairs determines if a relation exists, creating it if the similarity surpasses a predetermined threshold fixed to *0,85*. For the **Knowledge Graph Generation** we considered **PyVis** v.0.3.1 Accession date: 02/06/2023.

 ISBN: 978-1-68558-136-7

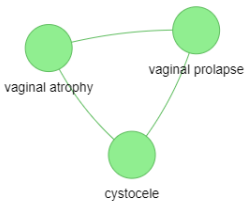

Figure. 2. Knowledge graph resulting from Medical Report 1 of the Experiment 1

The entities and relations derived from the aforementioned steps are organized into individual knowledge graphs using **Networkx** v.3.1 Accession date: 02/06/2023.

VI. Results

The experiment we show (detailed in Section VI-A) presents where the tool analyses the preprocessed dataset of 59,051 patients to produce knowledge graphs for numerous patients. This hypothetical situation might be similar to situations in which healthcare systems try to automatically create and preserve knowledge graphs for a large number of patients to aid in future consultations and plan creation.

A. Experiment 1: Automated Knowledge Graph Generation in Data Ingestion Mode

Example Case: Patient #10001876. Number of associated medical reports: 2.

Medical Report 1: Ms. ___ presented for evaluation of urinary complaints and after review of records and cystoscopy was diagnosed with a stage III cystocele and stage I vaginal prolapse, both of which were symptomatic. She also had severe vaginal atrophy despite being on Vagifem. Treatment options were reviewed for prolapse including no treatment, pessary, and surgery. She elected for surgical repair. All risks and benefits were reviewed with the patient and consent forms were signed.

Knowledge Graph for Medical Report 1. In Figure 2, we report the Individual Knowledge Graph generated by the system for Medical Report 1. Here only 3 interrelated entities have been extracted. Such graph is non informative with 7 entities and 13 relations.

Medical Report 2: She is a ___ patient who presents with ___ rectocele after having a sacral colpopexy and supracervical hysterectomy in ___ for uterine prolapse and cystocele. At that time, she had no rectocele at all. She has symptoms of bulge and pressure in the vagina that has gotten worse over the past few months. She also complains of feeling of incomplete emptying. She states that after she goes to the bathroom, she could go back and urinate some more. She had some frequency, urgency symptoms, which had resolved postoperatively. She also has resolved diarrhea after being started on Zenpep. She is followed by Dr. ___ and her fecal incontinence has resolved as well as resolved diarrhea."

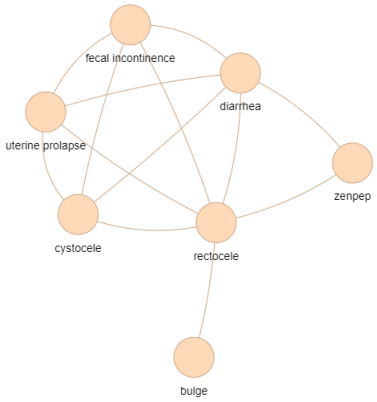

Figure. 3. Knowledge graph resulting from Medical Report 2 of Experiment 1

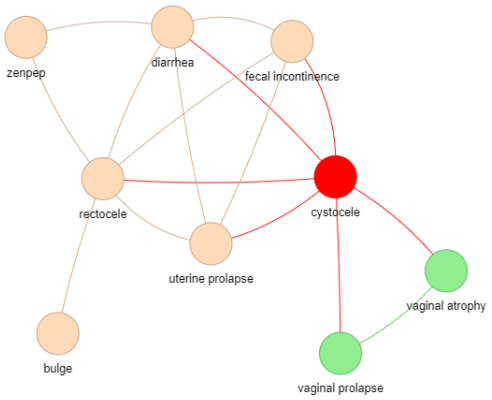

Figure. 4. Knowledge graph representing the merging of Medical Reports 1 and 2 for Experiment 1.

Knowledge Graph for Medical Report 2. In Figure 3 we report the Individual Knowledge Graph generated by the system for Medical Report 2

Merged Knowledge Graph of Experiment 1. Figure 4 shows the combined knowledge graph from Experiment 1, highlighting the shared entity *cystocele*, found in both reports, as a key connection point. This shared diagnosis suggests an ongoing or recurrent condition, emphasizing the importance of continuous monitoring and management. Recognizing such common conditions is essential for tracking disease progression or recurrence, aiding healthcare professionals in tailoring treatment plans to the patient's long-term medical history and current condition. Unique entities across reports, representing different medical conditions and treatments, are equally critical. For instance, *'vaginal prolapse'* noted in the first report, and *'rectocele'* and *'fecal incontinence'* in the second, highlight separate medical issues the patient has faced. These conditions—cystocele, vaginal prolapse, and fecal incontinence—are interconnected pelvic floor disorders. They involve the bladder bulging into the vagina, pelvic organ descent, and loss of bowel control, respectively, often due to weakened pelvic support ([16], [17], [18]). This information is vital for

 ISBN: 978-1-68558-136-7

TABLE I
COMPUTATIONAL TIMES.

Experiment	Time Required
Single patient (between 2-6 reports)	20-50 seconds
Entire dataset (59,051 patients)	12-14 days

TABLE II
SPACE USAGE.

Graph Type	Memory average	Memory - range	Memory (all patients)
Single Knowledge Graph	24 KB	[6 KB - 42 KB]	6.39 GB
Merged Knowledge Graph	56 KB	[8 KB - 110 KB]	3.9 GB

understanding the comprehensive scope of the patient's health challenges and planning appropriate interventions.

B. Computational Time and Space Usage

Creating knowledge graphs for each patient in our large dataset poses computational challenges. Our methodology accelerates graph generation, yet processing time escalates with dataset size, complexity of reports, and the quantity of entities and relationships. We utilized BERN2 API for NER, adhering to a 300 request limit per 100 seconds by incorporating 3-second pauses, prolonging processing for our dataset of 59,051 patients. The specific processing times are outlined in Table I. Storage requirements also significantly impact our experiments, with the space needed for individual and merged graphs dependent on the complexity and details of the diagnostic reports. Table II provides space usage statistics, showing that individual graphs require a total of 6.39 GB, while merged graphs need 3.9 GB.

VII. DISCUSSION

Addressing the complexity of healthcare information, our system autonomously creates and combines knowledge graphs from raw medical texts, navigating this crucial and challenging domain. Given the enormous variety of medical and biological entities present in healthcare, it was practical for us to narrow our primary attention to a small number of biomedical entities. This emphasis was seen in the studies, which showed the system's skill at locating, extracting, and connecting these chosen elements to create knowledge graphs that depict a clear and insightful narrative of a patient's medical journey. Focusing on a particular group of entities at this point allowed for deeper and more accurate knowledge as well as opened the door for methodical extension and inclusion of a wider variety of entities in the system's subsequent iterations. The experiments demonstrated the system's capability to accurately and coherently navigate medical texts, generating individual and merged knowledge graphs that highlight key entities and recurring illnesses, essential for understanding a patient's medical history and refining therapeutic strategies. The visualization tool emerged as a vital asset, offering medical professionals an intelligible visual narrative of a patient's medical journey, enhancing understanding and diagnostic ability.

A. Bridging Health Gaps: Societal Benefits of Comprehensive Medical Views

During brief appointments, some patients may find it difficult to remember and describe every medical exam, symptom, or medicine they have ever experienced (older people or people who are naturally reticent to retell every aspect of their medical history). Our system addresses these challenges by integrating multiple diagnostic reports into a unified visual representation. This ensures that every patient, irrespective of their background or communicative abilities, benefits from a comprehensive record that encapsulates their entire health journey.

B. Cost Efficiency

The proposed approach decreases the risk of unnecessary medical exams by giving a comprehensive perspective of a patient's health, which saves public and private money. Patients with complex medical histories, such as rare diseases, benefit most from the system since it makes sure they receive timely and effective care regardless of how many healthcare professionals they consult.

C. Global Scalability and Integration into Existing Infrastructure

The system showed excellent scalability, handling a dataset of 59,051 patients effectively, essential for managing the expanding volume of medical data. It's modular, allowing updates or replacements of components (e.g., entity extraction, relation prediction) without affecting the overall workflow.

D. Limitations and Threats to Validity

Input Accuracy: One of the foundational premises of our system is the reliance on accurate and relevant input. It's necessary that users (namely, doctors) provide diagnostic texts pertaining to the same patient. The system is designed to compare and integrate these texts, and any discrepancy in the input, such as including texts from unrelated patients, can lead to misleading results.

Natural Language Dependency: Our current implementation is tailored for the English language. This is largely because we utilize pre-trained tools, which are predominantly trained on English medical and biomedical terminologies. While the system demonstrates efficacy with English texts, its applicability could be limited in regions with different native languages. Expanding the system's capability to cater to diverse languages remains a future target.

Lack of Direct Baselines: It's challenging to compare our system directly with existing tools. While many tools extract entities from biomedical text, there are no tools aiming at integrating multiple texts into a unified knowledge graph.

E. Future Directions

As our system continues to evolve, one of our primary goals is to ensure its accessibility and usability worldwide. To achieve this, we are actively considering the incorporation of multilingual models, which would enable the system to

 ISBN: 978-1-68558-136-7

process and understand medical reports in various languages, catering to a global audience. Moreover, a promising frontier for our system lies in leveraging the intricate patterns within the knowledge graphs. Our vision is to utilize dedicated pattern recognition techniques that systematically analyze these graphs, pinpointing recurring sequences or clusters of entities and relations that could be indicative of specific medical conditions or trajectories [19]. For instance, by analyzing a vast number of knowledge graphs and tracing back the diagnostic journeys of patients with a particular condition, we might discern that certain entity relationships frequently precede the diagnosis of that condition [20].

VIII. Conclusion

Healthcare, at its core, revolves around accurate and timely information. In our study, we demonstrate the power of software engineering to bridge gaps, connect dots, and provide a comprehensive view of a patient's medical journey. By integrating fragmented medical reports into a unified knowledge graph, we ensure that no detail is missed. This holistic approach amplifies the quality of care, particularly for those who might struggle to articulate their medical experiences. This research underscores the synergy between software engineering and medical informatics, demonstrated through a system adept at autonomously generating and merging knowledge graphs from medical texts. The targeted focus on specific biomedical entities showcased the system's precision in narrating a patient's medical journey. The experiments reflected not only the accuracy and utility of this system but also its potential to significantly impact healthcare by aiding in timely and informed decision-making. The potential healthcare ramifications are profound. By reducing redundant medical exams, we envision a path towards more efficient and cost-effective healthcare. Moreover, we are committed to utilizing the knowledge graphs to gain valuable insights, which will help us develop proactive healthcare strategies and enable early interventions.

Acknowledgments

European Union - NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) - Project: "SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021. LIFEMAP-Dalla patologia pediatrica alle malattie cardiovascolari e neoplastiche nell'adulto: mappatura genomica per la medicina e prevenzione personalizzata Traiettoria 3 "Medicina rigenerativa, predittiva e personalizzata" - Linea di azione 3.1 "Creazione di un programma di medicina di precisione per la mappatura del genoma umano su scala nazionale" of the Ministry of Health.

References

[1] F. C. Bourgeois, K. L. Olson, and K. D. Mandl, "Patients treated at multiple acute health care facilities: quantifying information fragmentation," *Archives of internal medicine*, vol. 170, no. 22, pp. 1989–1995, 2010.

[2] W.-Q. Wei, C. L. Leibson, J. E. Ransom, A. N. Kho, P. J. Caraballo, H. S. Chai, B. P. Yawn, J. A. Pacheco, and C. G. Chute, "Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 219–224, 2012.

[3] D. Dong, R. Y.-N. Chung, R. H. Chan, S. Gong, and R. H. Xu, "Why is misdiagnosis more likely among some people with rare diseases than others? insights from a population-based cross-sectional study in china," *Orphanet journal of rare diseases*, vol. 15, no. 1, pp. 1–12, 2020.

[4] L. Murali, G. Gopakumar, D. M. Viswanathan, and P. Nedungadi, "Towards electronic health record-based medical knowledge graph construction, completion, and applications: A literature study," *Journal of Biomedical Informatics*, p. 104403, 2023.

[5] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," *Scientific Data*, vol. 10, no. 1, p. 67, 2023.

[6] A. Santos, A. R. Colaço, A. B. Nielsen, L. Niu, P. E. Geyer, F. Coscia, N. J. W. Albrechtsen, F. Mundt, L. J. Jensen, and M. Mann, "Clinical knowledge graph integrates proteomics data into clinical decision-making," *bioRxiv*, pp. 2020–05, 2020.

[7] A. Rossanez, J. C. Dos Reis, R. d. S. Torres, and H. de Ribaupierre, "Kgen: a knowledge graph generator from biomedical scientific literature," *BMC medical informatics and decision making*, vol. 20, no. 4, pp. 1–24, 2020.

[8] R. Kulkarni and Y. Haribhakta, "Building the knowledge graph from medical conversational text data and its applications," in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. IEEE, 2022, pp. 1508–1513.

[9] M. R. Kamdar, W. Dowling, M. Carroll, C. Fitzgerald, S. Pal, S. Ross, K. Scranton, D. Henke, and M. Samarasinghe, "A healthcare knowledge graph-based approach to enable focused clinical search." in *ISWC (Posters/Demos/Industry)*, 2021.

[10] M. Schmidli, "Outcome of patients with acute coronary syndrome in hospitals of different sizes. a report from the amis plus registry," *Swiss medical weekly*, vol. 140, no. 2122, pp. 314–322, 2010.

[11] T. Hewitt, S. Chreim, and A. Forster, "Incident reporting systems: a comparative study of two hospital divisions," *Archives of Public Health*, vol. 74, no. 1, pp. 1–19, 2016.

[12] D. Kim, J. Lee, C. H. So, H. Jeon, M. Jeong, Y. Choi, W. Yoon, M. Sung, and J. Kang, "A neural named entity recognition and multi-type normalization tool for biomedical text mining," *IEEE Access*, vol. 7, pp. 73 729–73 740, 2019.

[13] E. Alsentzer, J. R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, and M. McDermott, "Publicly available clinical bert embeddings," *arXiv preprint arXiv:1904.03323*, 2019.

[14] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow *et al.*, "Mimic-iv, a freely accessible electronic health record dataset," *Scientific data*, vol. 10, no. 1, p. 1, 2023.

[15] B. Andrea, "Medical Knowledge Harmonization: A Graph-based, Entity-Selective Approach to Multi-source Diagnoses," https://github.com/anbianchi/knowledge_frombio, [Accessed 27-02-2024].

[16] C. Aboseif and P. Liu, "Pelvic organ prolapse," 2020.

[17] C. Reisenauer, A. Kirschniak, U. Drews, and D. Wallwiener, "Anatomical conditions for pelvic floor reconstruction with polypropylene implant and its application for the treatment of vaginal prolapse," *European Journal of Obstetrics & Gynecology and reproductive Biology*, vol. 131, no. 2, pp. 214–225, 2007.

[18] P. Abrams, K.-E. Andersson, A. Apostolidis, L. Birder, D. Bliss, L. Brubaker, L. Cardozo, D. Castro-Diaz, P. O'connell, A. Cottenden *et al.*, "6th international consultation on incontinence. recommendations of the international scientific committee: evaluation and treatment of urinary incontinence, pelvic organ prolapse and faecal incontinence," *Neurourology and urodynamics*, vol. 37, no. 7, pp. 2271–2272, 2018.

[19] X. Tao, T. Pham, J. Zhang, J. Yong, W. P. Goh, W. Zhang, and Y. Cai, "Mining health knowledge graph for health risk prediction," *World Wide Web*, vol. 23, pp. 2341–2362, 2020.

[20] H. Wang, X. Miao, and P. Yang, "Design and implementation of personal health record systems based on knowledge graph," in *2018 9th international conference on information technology in medicine and education (ITME)*. IEEE, 2018, pp. 133–136.

 ISBN: 978-1-68558-136-7