Patient-centric knowledge graphs: a survey of current methods, challenges, and applications

Hassan S. Al Khatib*, Subash Neupane,
Harish Kumar Manchukonda, Noorbakhsh Amiri Golilarz,
Sudip Mittal, Amin Amirlatifi and Shahram Rahimi

Department of Computer Science and Engineering, Mississippi State University, Starkville, MS, United States

Patient-Centric Knowledge Graphs (PCKGs) represent an important shift in healthcare that focuses on individualized patient care by mapping the patient's health information holistically and multi-dimensionally. PCKGs integrate various types of health data to provide healthcare professionals with a comprehensive understanding of a patient's health, enabling more personalized and effective care. This literature review explores the methodologies, challenges, and opportunities associated with PCKGs, focusing on their role in integrating disparate healthcare data and enhancing patient care through a unified health perspective. In addition, this review also discusses the complexities of PCKG development, including ontology design, data integration techniques, knowledge extraction, and structured representation of knowledge. It highlights advanced techniques such as reasoning, semantic search, and inference mechanisms essential in constructing and evaluating PCKGs for actionable healthcare insights. We further explore the practical applications of PCKGs in personalized medicine, emphasizing their significance in improving disease prediction and formulating effective treatment plans. Overall, this review provides a foundational perspective on the current state-of-the-art and best practices of PCKGs, guiding future research and applications in this dynamic field.

KEYWORDS

knowledge graph, patient-centric, personalized healthcare, natural language processing, generative AI

1 Introduction

The healthcare industry has experienced a significant shift, transitioning from traditional, provider-centric models toward patient-centered care. This shift highlights the critical role of engaging patients as active participants in their healthcare journeys. At the center of this transformation are PCKGs, which significantly advance personalized, data-driven care. PCKGs facilitate the integration of diverse data types, including medical history, genetics, lifestyle choices, and real-time data from health technology devices, fostering a comprehensive view of patient health essential for customizing treatments to individual needs (Mesko, 2022; Blobel, 2007).

The evolution of Knowledge Graphs (KG) in healthcare into sophisticated networks reflects an increasing acknowledgment of the complexity of human health and the insufficiency of

siloed data systems in addressing multifaceted health issues. This evolution is propelled by the necessity for a holistic understanding of patient health, enabling personalized care and applying advanced data analytics to enhance healthcare outcomes (Blobel, 2011; Shirai et al., 2021). PCKGs stand at the forefront of healthcare innovation, signifying a crucial step toward integrated, patient-focused knowledge networks. This shift from fragmented data systems to cohesive KGs enables healthcare providers to employ machine learning and analytical technologies to help improve precision medicine, diagnostic accuracy, and treatment efficacy. Such a transition represents a technological leap that aligns with the broader objectives of healthcare reform aimed at delivering more personalized, efficient, and patient-centered care (Albannai et al., 2019; Almunawar and Anshari, 2012). However, incorporating KGs into healthcare presents technical, methodological, and ethical challenges, including data interoperability, privacy concerns, and the complexities of modeling diverse health outcomes. These hurdles pose significant barriers to the widespread adoption of PCKGs. Yet, the potential of these systems to revolutionize healthcare by offering a nuanced and comprehensive understanding of patient health is unquestionable (MacLean, 2021; Alagar et al., 2017).

PCKGs' primary aim is to enhance the quality of patient care, improve treatment outcomes, and increase the efficiency of healthcare delivery. By integrating disparate data sources and utilizing advanced analytical models, PCKGs promise to deliver personalized, efficient, and effective healthcare services tailored to each patient's unique needs. This goal emphasizes the shift toward a healthcare system that values and prioritizes individual patient experiences and needs, marking the beginning of a new era of patient-centric, data-driven care (Harper and Honour, 2015; Goniewicz et al., 2021). Despite the inherent challenges in integrating KGs into healthcare, the critical need for the advanced application of PCKGs to achieve personalized care and enhance healthcare delivery systems is undeniable. As the healthcare landscape continues to evolve, PCKGs exemplify the industry's commitment to leveraging technology to meet patients' complex and varied needs, thus marking a significant milestone in the journey toward more personalized and effective healthcare solutions.

The motivation for this paper stems from the growing need to consolidate disparate healthcare data into a unified, holistic view for improved patient care. The key contributions of this survey paper are:

- A foundational explanation of knowledge graphs serves as the theoretical basis for the remainder of the paper.
- Presentation of survey findings and introduction of a taxonomy developed for PCKGs.
- An in-depth review of methodologies designed explicitly for PCKGs, shedding light on the most effective techniques currently employed.
- Exploration of real-world applications and use cases that have successfully implemented PCKG methodologies, providing evidence of their utility.

The rest of the paper is organized as follows: Section 2 explains the principles of knowledge graphs, providing a foundation for the following discussions. Section 3 presents the findings from our survey and introduces the taxonomy we have developed. Moving on to Section 4, we review methodologies explicitly designed for PCKGs. Section 5 explores real-world applications and examples that benefit from these methodologies. Section 6 highlights research challenges

and provides targeted recommendations for future scholars in this field. Finally, in Section 7, we summarize the key findings of this paper and outline directions for work in PCKGs.

2 Background

The development of knowledge representation has a rich history in the realms of logic and AI. The notion of graphical knowledge representation can be traced back to 1956 when Richens (1956) introduced the concept of semantic nets. Similarly, symbolic logic knowledge finds its roots in the General Problem Solver (Newell et al., 1959) of 1959. Initially, knowledge bases were employed in knowledge-based systems for reasoning and problem- solving. Notably, MYCIN (Shortliffe, 2012), an expert system renowned for medical diagnosis, utilized a knowledge base containing approximately 600 rules. However, it was in 2012 that the concept of Knowledge Graph (KG) gained immense popularity, thanks to Google's search engine and its introduction of the Knowledge Vault framework (Dong et al., 2014). This framework aimed to construct large-scale KGs through knowledge fusion. There is currently no consensus on the definition of the term, with several authors proposing different definitions. Table 1 illustrates some of these definitions presently available in the literature.

A KG assumes the form of a directed graph (G), characterized by vertices and edges, where $G = (V, E)$. Vertices (V) indicate an entity, and Edge (E) between the two vertices convey the semantic relationship between two entities (Bordes et al., 2011). Within the graph, vertices, also referred to as entities or nodes, are linked through relationships represented as edges, and facts are depicted using RDF (Cyganiak et al., 2014) triples such as (subject, predicate, object) or (head, relation, tail), denoted by $< h, r, t >$ (Abu-Salih, 2021). Figure 1 depicts a simple KG with vertex v_9 and v_8 linked by the relation r_1, which goes from v_9 to v_8, making a triplet (v_9, r_1, v_8). In this scenario, v_9 represents the head, and v_8 represents the tail. In the context of PCKGs, nodes typically represent entities like patients, drugs, diseases, or genes, whereas edges denote relationships or associations between these entities (see Figure 2).

Over the past few years, especially within the domains of health and biomedicine, various forms of KGs have been proposed, often stemming from literature or Electronic Health Records (EHR). However, they are typically not individual-centered. Recent work on disease-centric knowledge graphs is available in Chandak et al. (2023) and Lin et al. (2023). These KGs facilitate clinicians' ability to study the relationships between diseases to answer practical clinical problems. Nevertheless, because relationships between disease concepts are frequently distributed across numerous datasets, integrating multimodal data sources is critical for developing comprehensive disease knowledge graphs (Chandak et al., 2023; Lin et al., 2023). On the other hand, their robustness of knowledge in terms of diseases and symptoms must be evaluated to assess their accuracy and quality (Chen et al., 2019).

In hospitals, patient information collected during visits and details of test results, clinical notes, symptoms, diagnoses, and drugs is frequently maintained as an EHR. These records include both unstructured data in the form of free notes, such as progress notes, admission notes, discharge summaries, medical histories, procedures notes, etc., and structured data, such as medical codes, medications,

TABLE 1 Various definitions of KGs are in the available literature.

| Source	Definition
Paulheim (2017)	A knowledge graph (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains
Ehrlinger and Wolfram (2016)	A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge
Wang et al. (2017)	A knowledge graph is a multi-relational graph composed of entities and relations which are regarded as nodes and different types of edges, respectively
Smirnova et al. (2018)	A Knowledge Graph, also known as a Knowledge Base, is a directed graph formed by triples that connect nodes (subjects) to other nodes (objects) through properties (predicates), where these connections represent semantic relationships and the graph's structure illustrates the subjects, objects, and their interrelations
Duan et al. (2017)	A knowledge graph organizes items, entities, and users as nodes interconnected by edges, providing rich semantics and comprehensive information in a structure akin to natural language
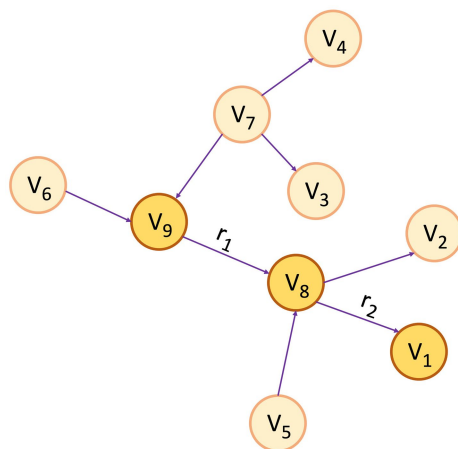
FIGURE 1
An example of a knowledge graph, where the triplet (v_9, r_1, v_8) serves as an illustration of the link between entities v_9 and v_8 through the relationship r_1 and (v_8, r_2, v_1) through r_2 for the relationship between v_8 and v_1.

administrative data, vital signs, and laboratory test report data (Sarwar et al., 2022). Numerous KGs have been introduced in the literature, utilizing EHRs as their foundation. As an illustration, Finlayson et al. (2014) constructed a graph including diseases, drugs, procedures, and devices. They leveraged 1 million clinical concepts from 20 million clinical notes to do so. Similarly, another study in Rotmensch et al. (2017) formulated a KG covering 156 diseases and 491 symptoms based on the insights drawn from a dataset of 273,174 patient visits to the emergency department. Although some efforts have been made to incorporate patient-specific information into these graphs, they are not yet patient-centric.

PCKGs can be linked to Personal Knowledge Graphs (PKG), an emerging concept in this field. Although PKGs are relatively new, some notable contributions have been made in this area. One of the pioneering papers that formally introduces the concept of PKG is authored by Balog and Kenter (2019). They define PKG *"to be a source of structured knowledge about entities and the relation between them, where the entities and the relations between them are of personal, rather than general, importance. The graph has a particular "spiderweb" layout, where every node in the graph is connected to one central node: the user."* Within the scope of this definition, the PKG only incorporates knowledge pertinent to the specific user.

The emergence of KGs has brought about a substantial transformation in the healthcare domain, providing numerous advantages to this industry. Integrating disparate health data to provide a comprehensive and holistic view of patient health is a significant application of KGs. In the era of precision medicine, wherein therapeutic strategies are customized to suit individual patients' distinctive genetic composition and lifestyle variables, KGs assume a crucial function. Integrating various data types, such as genetic profiles, medical history, lifestyle behaviors, and information from health technology devices, is facilitated to enhance efficiency. Implementing this all-encompassing methodology enables more accurate diagnoses, treatment strategies, and potentially enhanced health results (Johson et al., 2021).

Moreover, KGs are a formidable instrument for deciphering intricate biomedical data, resulting in amplified research findings (Bauer-Mehren et al., 2010). They provide a systematic depiction of connections among different entities, such as genes, diseases, drugs, and pathways, which can be utilized to formulate novel hypotheses and uncover concealed patterns. For example, KGs can elucidate the intricate relationship between a genetic mutation and its association with a specific disease or unveil the influence of environmental factors in the progression of said disease. As mentioned earlier, the discoveries possess the potential to significantly contribute to the development of groundbreaking therapeutic approaches and expand the horizons of biomedical investigation (Yuan et al., 2011).

KGs play a crucial role in developing effective decision-support systems in the clinical domain. By effectively amalgamating and analyzing individualized patient data, KGs can assist medical practitioners in formulating accurate diagnostic determinations, devising optimal treatment strategies, and forecasting patient prognoses. Consequently, the implementation of KGs can effectively mitigate the occurrence of medical errors (Raghupathi and Raghupathi, 2014).

KGs have found significant applications across various sectors beyond healthcare. For example, in e-commerce, KGs are extensively used to enhance user experience through personalized recommendations, search result ranking, and semantic search. Prominent e-commerce platforms, such as Walmart, Amazon and eBay, have been utilizing KGs to link items to their respective attributes, thereby providing more relevant product suggestions to customers (Xu et al., 2020). Moreover, KGs are increasingly used in Natural Language Processing (NLP), where they can improve the performance of machine translation, question-answering, and text summarization systems (Li et al., 2022). KGs also have substantial potential in cybersecurity, as they can aid in mapping relationships between cyber threats, vulnerabilities, and affected systems, facilitating better threat prediction and prevention (Liu et al., 2022).

FIGURE 2
Illustration of patient's clinical visits knowledge graph.
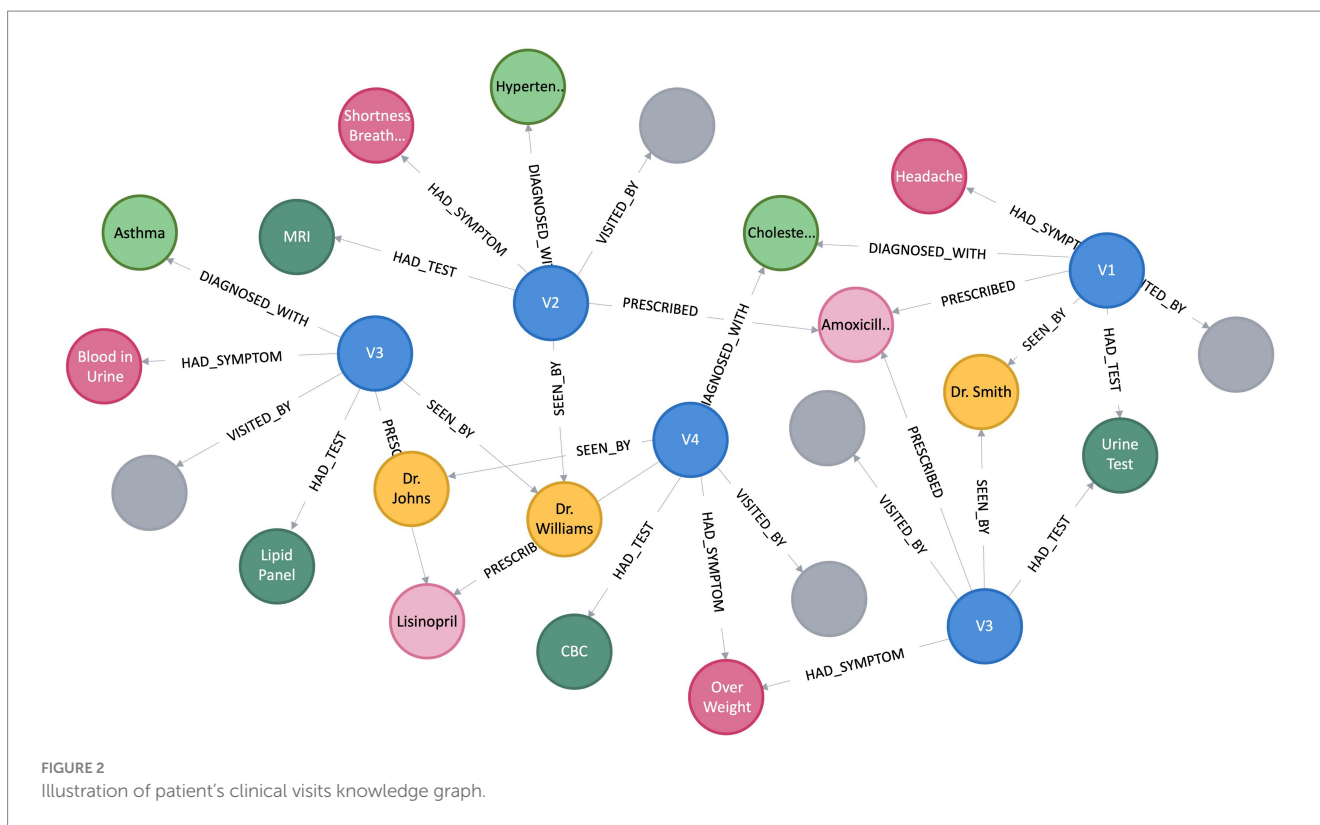
Social networking platforms like Meta and LinkedIn also leverage KGs to understand user relationships and interests, enhancing content recommendation and advertising targeting strategies (He et al., 2020). Hence, the applications of KGs are diverse and significantly impact various domains, reinforcing their relevance and potential.

3 Developed taxonomy of patient-centric knowledge graph

Having introduced the concept of KG in the previous section, we now explore the survey's core. This section will examine the various approaches, categorizations, and details influencing PCKGs. PCKGs are characterized by a variety of methodologies and studies. To navigate this diverse landscape, we need a well-structured framework. In this context, a taxonomy is a guide that enables readers to comprehend and classify the complexities involved in PCKG construction, evaluation, processing, and applications.

As depicted in Figure 3 our developed taxonomy is derived from an exhaustive survey and analysis of existing literature and practices (Eggert and Alberts, 2020). Based on our findings, we classify PCKG taxonomy into four major categories, including *construction, evaluation, process, and applications*. Given the heterogeneity of studies in the domain, it provides a structured representation and insight into the why and how of each categorization. The taxonomy presented in this paper outlines essential relationships and highlights the critical aspects of each of these categories. Its uniqueness lies in its exhaustive nature, informed by the foundational literature and cutting-edge research. In the following paragraphs, we delve into each category and subcategory.

The first category pertains to the *construction* of a PCKG and is further divided into four sub-categories: *ontology, knowledge sources, knowledge extraction, and knowledge representation*.

The initial step, *Ontology*, revolves around designing structured frameworks custom-tailored to meet patient-centered needs, ensuring that the graph accurately captures the essence of patient data while adhering to industry standards and best practices. It is crucial to articulate distinctions among closely related terms in this context: a Knowledge Graph (KG), which represents interconnected real-world facts; a Knowledge Base (KB), which stores this structured information; and an ontology, which provides a structured framework that illustrates and organizes concepts, ensuring consistent understanding and interpretation of the data within the KG and KB. Second, considering the diverse nature of healthcare data, our *Knowledge Sources* segment highlights the differences and integration methods for three primary sources of data: structured, semi-structured, and unstructured data. Third, the *Knowledge Extraction* phase employs techniques like Named Entity Recognition (NER) to pinpoint essential entities such as drugs or symptoms (Keretna et al., 2015). The process is complemented by Relationship Extraction (RE), which accurately maps the relationships among these entities, giving the graph its characteristic interconnected structure (Wang et al., 2021). The last sub-category in construction is *Knowledge Representation*, which focuses on detailing the methodological approach of encapsulating explicit and implicit knowledge in the PCKG. This involves defining the schema for entities and relationships, establishing constraints, and encoding domain knowledge through axioms and rules, thus constructing a robust framework that effectively translates raw data into a comprehensible and interconnected knowledge graph that accurately mirrors the complexity and depth of patient-centric data.

FIGURE 3
Proposed taxonomy of patient-centric knowledge graph.
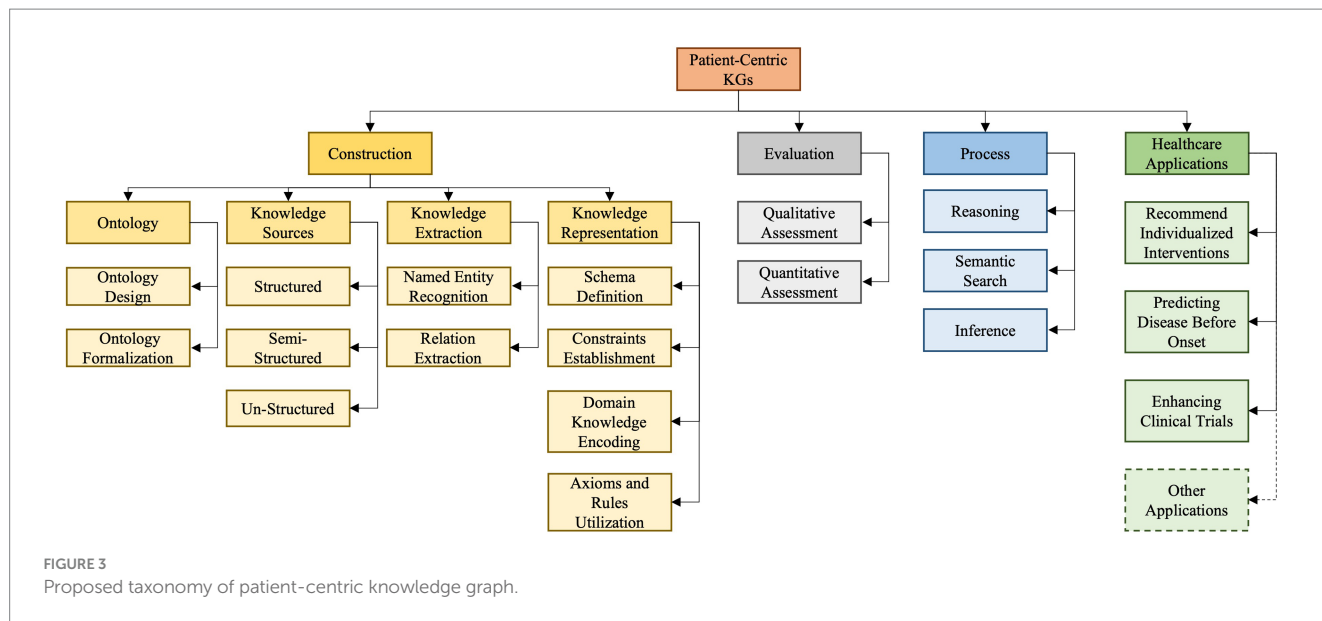
The second category involves the evaluation of the PCKG, emphasizing the need to assess the efficacy and accuracy of PCKGs post-construction. This assessment can be facilitated by two principal methodologies: *qualitative* and *quantitative*. The former involves a thorough examination of the graph to ensure that it adheres to clinical best practices and contains relevant and correct medical relationships. The latter uses statistical and computational methodologies to evaluate the graph's performance metrics, including accuracy, recall, and precision. It also offers a numerical assessment of its ability to represent and infer knowledge accurately and effectively. Once the PCKG has been constructed and evaluated, the next step is its *utilization (processing PCKG)* to derive actionable insights. To attain this objective, various approaches are crucial, including using methods such as *reasoning, semantic search, and inference. Reasoning* enables the PCKG to logically extract new information and knowledge from the existing knowledge within the graph, providing a basis for making intelligent decisions and predictions about healthcare outcomes and strategies (Rajabi and Kafaie, 2022). In contrast, *Semantic Search* allows structured searches to retrieve precise information from the graph (Lytras et al., 2009). *Inference*, on the other hand, uses the inherent structure of the graph to derive new insights and conclusions (Wang et al., 2018).

The final category of the our taxonomy relates to the *application* of PCKG in the healthcare domain. We explore four primary healthcare use cases, encompassing *recommending individualized interventions, clinical trials, predicting disease before onset, and others.* The first personalizes patient care treatments based on insights from the graph. The second application aims to enhance the efficiency of *Clinical Trials*, aiding in developing trial designs or optimizing patient selection. On the other hand, a particularly innovative application is *Predicting Disease Before Onset*, where the complex patterns and relationships mapped in the graph can be harnessed to detect potential health issues proactively. Although these are highlighted applications, the potential of PCKGs extends even further, covering areas such as *drug discovery, personalized medication regimens,* and *preventive health strategies,* to name a few.

In the following sections, we will examine each component of our taxonomy in more detail to better understand its significance and provide an overview of the suggested implementation steps.

4 Construction, evaluation and processing of PCKGS

This section explores the complex processes and methodologies in developing, assessing, and utilizing PCKG. This comprehensive exploration is structured into four critical categories, as illustrated in Figure 3. The first category, *construction* of PCKG, explains the complex process of building PCKG, highlighting the significance of ontology, various data sources, knowledge extraction techniques, and knowledge representation. Next the *evaluation* covers qualitative and quantitative methods for assessing the efficacy and accuracy of these graphs. Similarly, we discuss advanced techniques like *reasoning, semantic search*, and *inference*, which are essential to harnessing the full potential of PCKGs to derive actionable insights and inform healthcare decisions.

4.1 PCKG construction

The process of creating PCKGs involves four steps. To begin with, the *ontology* phase involves developing structured frameworks based on patient data, which ensure accuracy and adherence to standards. *Knowledge sources* are the second step, which addresses the collection and integration of diverse types of healthcare data, including structured, semi-structured, and unstructured data. During the third step, *knowledge extraction*, key data entities and their interconnections are mapped using named entity recognition and relationship extraction techniques. Lastly, *knowledge representation* involves defining a schema for entities and relationships, setting constraints, and encoding domain knowledge to transform raw data into a coherent, interconnected graph reflecting patient data complexity. In the following subsections, we will explore these steps in detail.

4.1.1 Ontology

Ontologies provide a structured framework for representing knowledge, integrating diverse data sources, and facilitating sophisticated access to health information—critical components in patient-centered healthcare (Puustjarvi and Puustjarvi, 2010). Designing and formalizing ontologies involves creating a comprehensive schema that unifies information from various *structured* and *unstructured* data sources (Iannacone et al., 2015). Unlike knowledge bases, which are collections of domain-specific knowledge often built upon ontologies or schemas, which primarily organize data as a blueprint for databases, ontologies formally represent concepts within a domain and the relationships between them, offering a deeper contextual understanding. In healthcare, the application of ontologies has been well-documented. For instance, Sun et al. (2014) introduced an ontology-enabled healthcare service model that supports joint referral decisions between patients and general practitioners, enhancing patient-centered care.

One practical implementation of ontology principles is the Fast Healthcare Interoperability Resources (FHIR) standard. FHIR plays a crucial role in structuring patient-centric data, enabling the representation of complex healthcare information—such as patient records, clinical observations, and care plans—in a machine-readable and interoperable format (Rishi Kanth Saripalle, 2019). FHIR uses a resource-based model with RESTful APIs, which supports formats like JSON, XML, and RDF, and can be integrated into graph-based models, including PCKGs, to enhance data relationships. By integrating FHIR-based models into PCKGs, we can achieve enhanced interoperability and comprehensive patient data mapping. This ensures that PCKGs accurately reflect the multifaceted nature of patient health information, supporting standardized healthcare data exchange and facilitating more effective integration across diverse healthcare systems (Maxhelaku, 2019).

4.1.1.1 Design

The growing digitization of healthcare data and the proliferation of medical information need effective and structured methods of representing and managing patient-centric knowledge. Ontologies, formal representations of knowledge, play a critical role in healthcare systems' interoperability, data integration, and decision support (Taye, 2010). In the context of PCKGs, ontology design is a fundamental process that involves defining and modeling the important entities, connections, and features related to patient health and treatment. Figure 4 provides an example of a patient's ontology. Authors in Noy and McGuinness (2023) argue that the rationale behind ontology development is to aid in a shared understanding of information structure among people or software agents, facilitate the reuse of domain knowledge, make domain assumptions explicit, separate domain knowledge from operational knowledge, and analyze domain knowledge. However, there is no standard methodology (Kapoor and Sharma, 2010) or correct way of formulating ontology (Noy and McGuinness, 2023).

The initial stage in ontology design is to establish the scope and domain of the PCKG. The scope of the ontology specifies the breadth of medical knowledge included. In contrast, the domain defines the precise areas of concentration, such as general medical information, specific diseases, or healthcare procedures. Identifying the domain aids in the selection of acceptable current ontologies such as SNOMED-CT (El-Sappagh et al., 2018), RxNorm (Hanna et al., 2013), and LOINC (Uchegbu and Jing, 2017) are healthcare standards that can be utilized to maintain interoperability and consistency. While current ontologies and standards may provide a solid foundation, some patient-centric concepts may not be fully covered. In such circumstances, additional classes and features must be added to cater to specific requirements. This requires collaboration between domain experts and ontology engineers to guarantee that the ontological representations are consistent with healthcare practices and guidelines.

The second step is to define the core entities. Entities such as *Patient*, *Disease*, *Medication*, *Symptom*, *Medical Procedure*, *Healthcare Provider*, etc. are typically included. These entities serve as the building blocks for constructing the KG and are crucial for capturing patient-related information.

The next step is to extract or specify the relationships between entities. This step is essential for capturing the complex interactions within patient-centric healthcare. Relationships, such as *Diagnosis, Treats, Has Symptom*, and *Undergoes,* enable meaningful associations and support inferencing capabilities. During this stage, the ontology developer should also carefully consider the attributes and properties of the entities. For example, attributes of a *patient* could be *age*, *gender*,
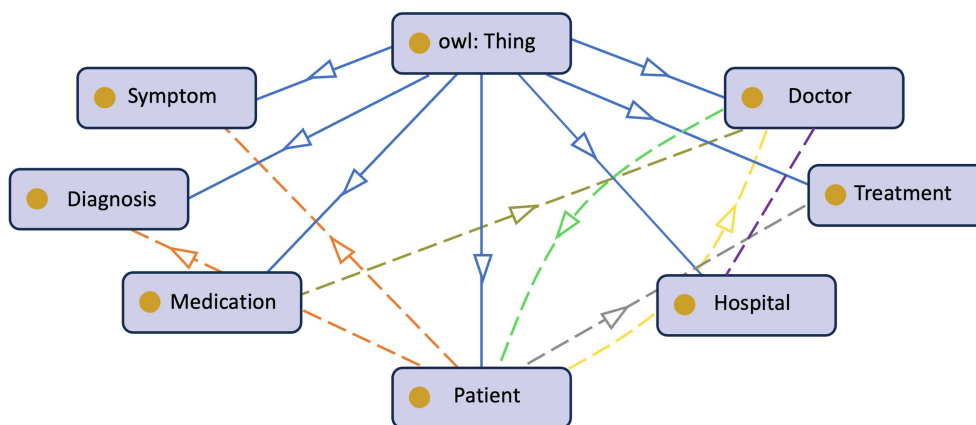

FIGURE 4
Illustration of a basic patient's ontology.

sex, *demography*, etc. These attributes can influence the diagnosis and treatment decisions (Bravo, 2019).

4.1.1.2 Formalization

Formalizing the ontology focuses on structuring and representing patient data and medical knowledge in a way that is understandable and accessible to healthcare systems. Puustjarvi and Puustjarvi (2010) leveraged the Resource Description Framework (RDF) and Web Ontology Language (OWL) to facilitate sophisticated access to health information, a cornerstone for effective patient care. This innovative approach underscores the evolving landscape of healthcare informatics, where the structuring and accessibility of patient data are critical. Similarly, Jiang et al. (2003) provided insight into the role of Formal Concept Analysis (FCA) in developing ontologies within clinical domains. FCA emerges as a key tool, offering linguistic and context- based knowledge. This knowledge is indispensable for clinical experts, aiding them in comprehending and applying ontology effectively in their practice. Integrating FCA into ontology development signifies a deeper understanding of clinical data and its nuances, enhancing the overall utility of these ontologies in real-world medical settings.

Further broadening this scope, Djedidi and Aufaure (2007) presented a methodology that underscores the integration of heterogeneous data sources with existing ontologies and standards. This approach is instrumental in constructing a medical domain ontology as a comprehensive, knowledge-centric decision support system. Such integration is crucial in ensuring the developed ontology is robust and aligns seamlessly with existing medical knowledge frameworks and data sources. Likewise, Mendes et al. (2013) contributed significantly to the Ontology for General Clinical Practice (OGCP) development. This ontology extends the Ontology for General Medical Science (OGMS) by incorporating the Clinical.

Patient Record (CPR) structure. This extension significantly enhances the representation and reasoning capabilities within clinical practice knowledge, especially in the context of natural language text. The OGCP stands as a testament to the evolving complexity and sophistication required in modern medical ontologies, catering to the nuanced needs of clinical practice.

The dynamic nature of medical knowledge and patient data necessitates methodologies for tracking and updating information. In this context, Ognyanov and Kiryakov (2002) proposed a formal model for tracking changes in RDF(s) repositories. This model is vital for maintaining the accuracy and relevance of the KG over time, addressing the ever-changing landscape of medical data and knowledge. Clarkson et al. (2018) introduced a user-centric methodology for the ontology population to address the user perspective. This approach aligns user concepts with target ontologies, proving an efficient method for building and maintaining ontologies across various domains. The user-centric approach ensures that the developed ontologies are technically sound and resonate with the needs and understandings of the end-users, be they patients or healthcare professionals.

Ferilli (2021) proposed an intermediate format that can be easily mapped onto formal ontology for complex reasoning and a graph database for efficient data handling to bridge the gap between formal ontology and practical application. This innovation represents a significant stride in harmonizing the theoretical aspects of ontology

with the practical demands of data management, ensuring that the developed ontologies are both conceptually sound and practically applicable.

Together, these diverse yet interconnected research efforts paint a comprehensive picture of ontology formalization's current state and future potential in patient-centric healthcare. They highlight a collaborative and multi-faceted approach to developing technically robust ontologies that are deeply integrated with the practical realities of healthcare delivery.

4.1.2 Knowledge sources

PCKGs are complex structures that integrate diverse data sources, as illustrated in Figure 5, to offer a comprehensive perspective on a patient's medical past, present ailments, and prospective therapies. The extent of a PCKG's depth is contingent upon the diversity and quality of the data sources that contribute to its composition. The sources can be classified into three main categories: *structured, semi-structured, and unstructured data*.

4.1.2.1 Structured data

Structured data is organized and easily searchable in relational databases. It follows a specific schema or model, making it straightforward to query and analyze. Structured healthcare data, integral to PCKGs, is exemplified by *EHRs, lab results and diagnostics*, and *genomic and molecular data*. *EHRs*, as digital records of a patient's medical history, encompass diagnoses, medications, treatment plans, and vital statistics, providing a highly structured backbone for PCKGs. Transitioning to *lab results and diagnostics*, this category includes structured, numerical, or categorical data from blood tests, imaging studies (like X-rays and MRIs), and other diagnostic tests, facilitating their integration into PCKGs. Lastly, *genomic and molecular data* derived from genetic tests offer structured insights into a patient's disease predisposition, proving essential for advancing personalized medicine. Collectively, these structured data forms are fundamental in healthcare analytics and personalized treatment planning.

4.1.2.2 Semi-structured data

Semi-structured data does not reside in a relational database but does have some organizational properties that make it easier to analyze. It often requires some level of transformation to be fully utilized. Semi- structured healthcare data encompass *clinical notes* and *patient-generated data*, each possessing unique characteristics that bridge the gap between structured and unstructured data. While fundamentally unstructured in text, *clinical notes and narratives* often adhere to templates or contain tags, rendering them semi-structured and providing contextual depth absent in purely structured datasets. Similarly, *patient- generated data*, sourced from wearables like fitness trackers and mobile health apps, presents structured data points (e.g., heart rate, steps) but lacks a consistent schema overall, contributing to the semi-structured nature of the data set. This combination of structured elements within an otherwise unstructured framework is valuable for comprehensive healthcare data analysis.

4.1.2.3 Unstructured data

Unstructured data is information that does not have a pre-defined data model, making it challenging to analyze using conventional methods. Unstructured healthcare data, characterized by its non-standardized format, includes *patient surveys*, *interviews*, *medical*

FIGURE 5
Diverse sources of medical and healthcare data with examples.
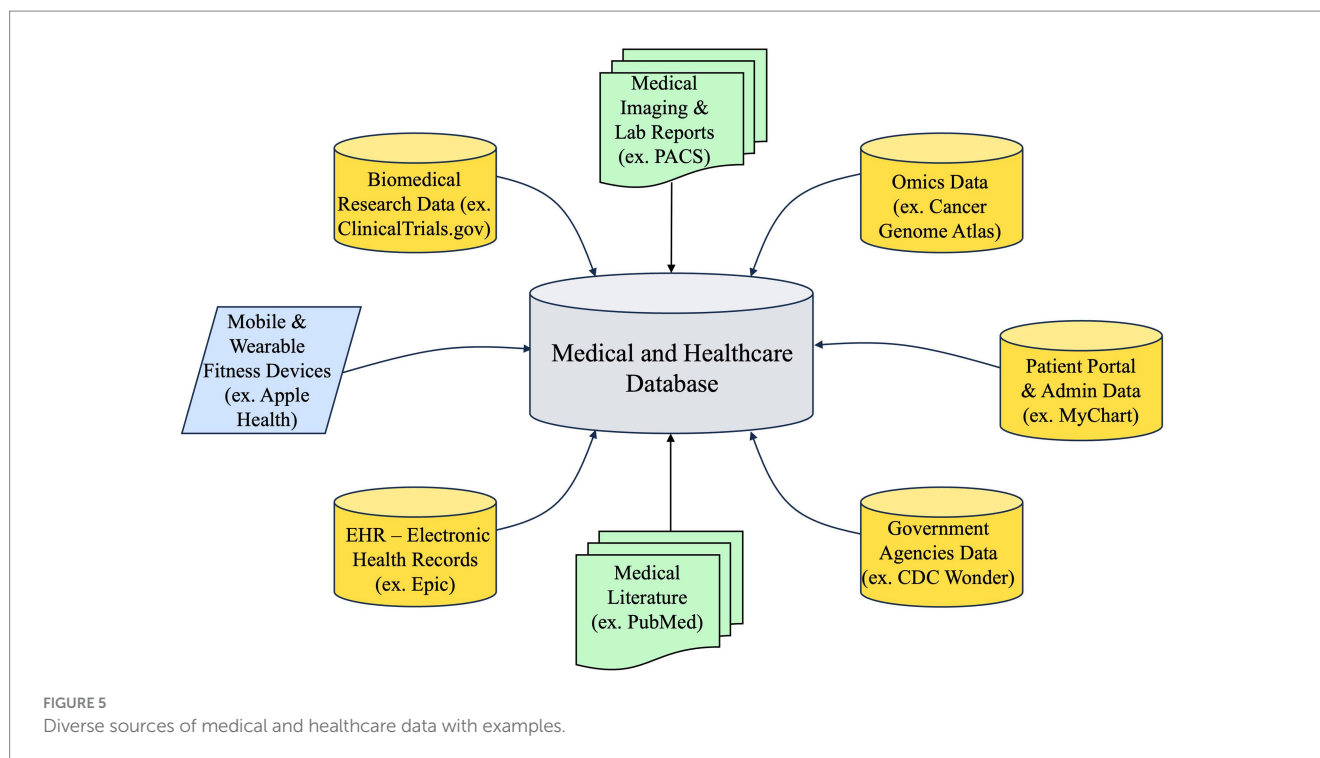
literature, and *research papers*, each offering unique challenges and insights for PCKGs. *Patient surveys and interviews*, typically in text or audio formats, yield qualitative insights into patients' conditions but are unstructured, necessitating advanced analytics for effective integration into PCKGs. Similarly, *medical literature and research papers* are a rich source of valuable insights but remain unstructured; thus, they require NLP techniques to distill and extract relevant information for PCKG incorporation. With its inherent complexity, this unstructured data is crucial in enhancing the depth and breadth of healthcare analytics and patient care understanding.

By integrating these diverse data sources, PCKGs can offer a multi-dimensional view of a patient's health, enabling more personalized and effective healthcare interventions.

4.1.3 Knowledge extraction

Knowledge extraction plays a pivotal role in constructing PCKGs. This stage leverages techniques such as Named Entity Recognition (NER) and Relationship Extraction (RE) to distill valuable insights from unstructured data. While NER focuses on identifying and categorizing key entities in the data, RE takes a step further to determine the relationships between these entities. A structured and rich KG can be constructed through the harmonious interplay of NER and RE, paving the way for more personalized and efficient healthcare solutions.

4.1.3.1 Named entity recognition

NER identifies and classifies entities such as diseases, symptoms, and medications from unstructured text. This process is fundamental in transforming raw data into structured, actionable information. Recent advancements in NER methodologies have significantly enhanced the development of KGs, particularly in the medical domain. For instance, Li et al. (2019) introduced the BiLSTM-Att-CRF model, which integrates an attention mechanism and part-of-speech

features to improve clinical NER in Chinese electronic medical records. Building on this, Keretna et al. (2015) proposed a graph-based technique that enhances medical NER performance by up to 26% in unstructured medical text, showcasing the versatility of NER applications.

Furthering these developments, Wang and Haihong (2021) developed a bi-directional joint embedding model that combines encyclopedic knowledge with original text, thereby showing improved results in Chinese medical NER. Complementing this approach, Wang et al. (2018) demonstrated that incorporating dictionaries into deep neural networks effectively addresses the challenge of rare and unseen entities in clinical NER, thus broadening the scope of NER applications. In a similar vein, Guan and Tezuka (2022) highlighted the importance of entity linking and intent recognition in medical question-answering systems, significantly improving medical KG searches' efficiency and accuracy. In a similar context, Zhou et al. (2021) developed a multi-task adversarial active learning model that enhances medical NER and normalization by considering task-specific features, further illustrating the dynamic evolution of NER methodologies.

Applying Large Language Models (LLMs) for NER is increasingly considered pivotal in healthcare- related KGs. Recent advancements in LLMs have demonstrated their effectiveness in improving NER tasks by leveraging contextual embeddings to enhance the accuracy of entity recognition, particularly in complex biomedical texts (Liu and Fang, 2023; Iscoe et al., 2023). For instance, fine-tuning LLMs on specific medical datasets has shown promising results in identifying entities related to patient care, thus enriching the KG with valuable insights that support personalized medicine (Hafsah et al., 2023). Furthermore, integrating ontological knowledge into LLM-driven NER processes can significantly enhance the interpretability and robustness of the extracted information, making it more applicable to clinical applications (Liu et al., 2022).

These diverse methodologies reflect the evolving landscape of NER in developing PCKGs. Integrating advanced machine learning techniques, attention mechanisms, LLMs, and domain-specific knowledge has led to more accurate and efficient entity recognition, which is crucial for effectively utilizing KGs in healthcare.

4.1.3.2 Relation extraction

Like NER, RE is a fundamental process in transforming unstructured data into a structured form, enabling more effective healthcare and medical research utilization. Recent advancements in RE methodologies, particularly in the context of PCKGs, have been significant. Ruan et al. (2020) demonstrated that a multi-view graph learning method could notably enhance the precision, recall, and F1 score in relation extraction from Chinese clinical records, outperforming state-of-the-art methods. Similarly, Li et al. (2019) highlighted the evolution of relationship extraction algorithms, emphasizing the role of deep learning, reinforcement learning, active learning, and transfer learning in this domain.

In the realm of therapy-disease KGs, Wang et al. (2021) constructed a Therapy-Disease KG (TDKG) using entity relationship extraction, achieving an 88.98% accuracy in extracting valid relationships from treatment-disease literature. This underscores the potential of RE in enhancing the accuracy and comprehensiveness of medical KGs. Furthermore, the integration of KGs and lexical features in biomedical relation extraction, as shown by Zhao et al. (2020), indicates significant improvements in semantic understanding. This approach exemplifies the innovative strategies adopted in recent research to enhance the effectiveness of RE in medical contexts. The development of Conditional Random Field (CRF)-powered classification models with deep learning for clinical relation extraction, as demonstrated by Lv et al. (2016), further illustrates the ongoing advancements in this field. These models have shown effectiveness in extracting clinical relation information from medical texts, thereby improving the construction of medical KGs.

In addition to NER, LLMs exhibit capabilities in processing unstructured data, enabling them to identify and extract relevant relationships between entities in clinical narratives, which is crucial for building comprehensive knowledge graphs that reflect patient information and care pathways (Liu and Fang, 2023). For instance, LLMs can automate the extraction of relationships from EHRs, thereby facilitating the integration of diverse medical data into structured formats that enhance clinical decision-making (Ge et al., 2023). Furthermore, studies have demonstrated that LLMs can achieve extraction rates as high as 87.7% in clinical contexts, outperforming some traditional NLP methods (Choi et al., 2023).

The field of NER and RE in PCKGs is rapidly evolving, incorporating cutting-edge methodologies such as multi-view graph learning, deep learning, LLMs, and knowledge-enhanced approaches. These advancements are pivotal for accurately extracting relationships from complex medical texts and enriching KGs essential for advancing patient-centered care and medical research. However, challenges like non-comprehensiveness and factfulness remain critical. Non-comprehensiveness refers to gaps where certain relevant details may be missed, especially when handling unstructured data (Nastase and Kotnis, 2019). To address this, iterative extraction techniques and the integration of multiple data sources must ensure that PCKGs provide a comprehensive view of patient information. Equally important is factfulness, as the accuracy of the extracted information

directly influences patient safety and the utility of the PCKG (Radevski, 2023). Ensuring factfulness requires rigorous validation, including cross-referencing data with authoritative medical sources and employing machine learning models designed to prioritize accuracy, thereby mitigating the risks of errors in the extraction process (Lin et al., 2018).

4.1.4 Knowledge representation

Knowledge Representation (KR) serves as a methodological backbone, enabling the systematic transformation of explicit and implicit knowledge into a structured and interpretable format. This transformation is crucial for machines to effectively process, analyze, and infer from the data, bridging the gap between raw data and actionable insights.

The process of KR encompasses four key components: schema definition, constraint establishment, domain knowledge encoding, and *axioms and rules utilization*, as explained in Section 3. Firstly, *schema definition* outlines entities' structure and relationships within the KG. This step is fundamental in organizing data in a meaningful and accessible manner. Secondly, *constraint establishment* is essential for maintaining data integrity and consistency, ensuring that the relationships and entities adhere to predefined rules and norms. Thirdly, incorporating *domain knowledge* is instrumental in embedding domain-specific intelligence into the system. This allows for a nuanced understanding of the specific area of interest. Fourthly, integrating *axioms and rules* facilitates advanced reasoning capabilities, enabling the system to infer new knowledge and insights based on the existing data.

These components collectively form a robust framework that simplifies complex data and enhances the system's capability to mirror the intricacies inherent in patient-centric data. In the following subsections, we will delve deeper into these components, exploring their roles and significance in the broader context of KR.

4.1.4.1 Schema definition

The concept of *schema definition* in PCKGs revolves around structuring and organizing patient-related data to enhance understanding and interaction with this information. Schemas in KGs serve as blueprints for organizing data, enabling more effective data integration, querying, and analysis. This is particularly crucial in healthcare, where patient data is complex and multifaceted. It aims to create a unified and comprehensive view of patient information, facilitating better healthcare outcomes through informed decision-making. Ghosh and Gilboa (2014) emphasize the role of memory schemas in coordinating knowledge structures, highlighting the importance of understanding cognitive processes in schema development. This perspective is crucial in PCKGs, where patient data must be organized to align with healthcare professionals' cognitive schemas.

Building on the concept of visual schemas by Kranjec et al. (2013), which enhance comprehension and aid in the diagnostic process, the work of Ohira et al. (2011) takes a step further by introducing the dynamic nature of graph-based data models. This flexibility is key in PCKGs, where patient data is not static but continuously evolving. The ability to dynamically add and remove relationships in these KGs ensures that they remain up-to-date and reflect the latest patient information, a critical aspect of the fast-paced healthcare environment. This dynamic nature raises an important question: should the

structure of a PCKG be acyclic or allow for cycles? Acyclic graphs, such as Directed Acyclic Graphs (DAGs) (Byeon and Lee, 2023), can be beneficial in scenarios where clear, hierarchical relationships are required, particularly in representing diagnostic pathways or decision trees (Ellison, 2023). However, in many healthcare applications, patient data often involves temporal or causal relationships that inherently form cycles. For instance, a patient's treatment may influence subsequent conditions or interventions, creating feedback loops that must be accurately represented within the graph. Therefore, allowing cycles within the PCKG may be necessary to capture healthcare data's complex, interdependent nature, enabling a more realistic and comprehensive representation of patient health journeys (Pressat Laffouilhère et al., 2022).

Transitioning from the dynamic structuring of patient data, the focus shifts to the cognitive aspects of PCKGs, as discussed by Gilboa and Marlatte (2017). Their research underscores the significance of aligning PCKGs with the mental models of healthcare professionals. This alignment is not just about efficiently structuring data; it's about ensuring that the KG resonates with the users' cognitive processes, enhancing memory recall and decision-making capabilities.

Further advancing this discussion, Ji et al. (2018) delve into the intricacies of schema induction in KGs. Their exploration into the challenges and developments in this field is particularly relevant for PCKGs. As these KGs become more complex and integral to healthcare, understanding and overcoming the challenges in schema induction is paramount for ensuring that PCKGs are robust, comprehensive, and seamlessly integrated into the healthcare workflow.

4.1.4.2 Constraints establishment

Establishing constraints in PCKGs focuses on enhancing the accuracy, privacy, and efficiency of medical data management and analysis. Various methodologies and strategies have been adopted to create and apply constraints in PCKGs, as evidenced by recent academic research. For example, Aguilar-Escobar et al. (2016) demonstrate the application of the Theory of Constraints (TOC) in improving the logistics of medical records in hospitals, highlighting the potential of TOC in enhancing service quality and reducing costs in healthcare settings. This approach underscores the importance of efficient data management in PCKGs.

Building on this notion of efficiency, the focus shifts to the critical aspect of privacy and security in healthcare, where Mathew and Obradovic (2011) illustrate how constraint graphs can secure privacy in medical transactions, preventing unauthorized attribute-based transformations in clinical data. This is crucial in PCKGs, where patient data sensitivity is paramount. The evolution of PCKGs further extends into dynamic data handling and adaptability. Matsuto and Shiina (1990) propose a model for medical knowledge representation based on constraints, which dynamically generates laboratory schedules using constraint propagation. This approach emphasizes the adaptability of PCKGs in accommodating patient-specific information.

Advancements in KG reasoning further enhance the adaptability of PCKGs. Wu et al. (2017) discuss a constraint-based embedding model for KG reasoning, focusing on semantic-type constraints in constructing corrupted triplets. This methodology significantly improves the reasoning accuracy of KGs, a key aspect in developing effective PCKGs. Alongside reasoning accuracy, the representation of knowledge and uncertainties also plays an important role. Pugh and Gillan (2020) introduce Propositional Constraint Graphs (PCG) to

represent knowledge and uncertainties in various tasks, including healthcare. This approach aids in clearly visualizing and managing complex information in PCKGs.

Applying these concepts is not limited to static scenarios but extends to dynamic and evolving medical challenges. Tu (2021) proposes a constraint propagation approach for identifying biological pathways in COVID-19 KGs. This method uses semantically extracted information to indicate the potential for PCKGs in pandemic response and other rapidly evolving medical scenarios. Finally, integrating context-aware constraints brings a new dimension to the integrity and validation of knowledge in PCKGs. Wilcke et al. (2020) discuss using context-aware constraints to improve knowledge integrity in heterogeneous KGs. This balance of complexity and utility is crucial for knowledge validation tasks in PCKGs (Wilcke et al., 2020), highlighting constraints' continuous evolution and application in developing robust and efficient PCKGs.

Establishing constraints in PCKGs is a multifaceted process involving the integration of various methodologies to enhance data accuracy, privacy, and efficiency. The strategies adopted in recent research reflect a growing emphasis on adaptability, security, and precision in managing patient-centric data in healthcare systems.

4.1.4.3 Domain knowledge encoding

Domain Knowledge encoding focuses on integrating and representing complex medical knowledge in a structured and interconnected format. This encoding process is crucial for developing PCKGs, which enhance patient care through personalized and informed decision-making. The methodologies for creating domain knowledge encoding in PCKGs are diverse and innovative. Wang et al. (2016) proposed a framework utilizing medical chart and note data, employing a bag-of-words encoding method and a model considering both global information and local correlations between diseases. This approach underscores the importance of capturing the nuances of medical data for effective KG construction.

Further expanding on these methodologies, Ji et al. (2020) highlighted the significance of representation space, scoring function, encoding models, and auxiliary information in KG creation. This comprehensive approach indicates the multi-faceted nature of KG development, where various components contribute to the robustness and accuracy of the resulting graph. Similarly, Yuan et al. (2019) presented a biomedical domain KG construction approach that includes entity recognition, unsupervised entity and relation embedding, latent relation generation, clustering, relation refinement, and relation assignment. This method demonstrates the complexity of accurately representing medical knowledge in graph form. On the other hand, Rotmensch et al. (2017) utilized rudimentary concept extraction and three probabilistic models to construct high-quality health KGs, with the Noisy OR model yielding the best results. This study exemplifies using probabilistic models to enhance the quality of KGs. Furthermore, Yu et al. (2020) proposed a relationship extraction method for domain KG construction, obtaining upper and lower relationships from structured, semi-structured, and unstructured text. This method highlights the importance of extracting relationships from diverse data sources to enrich the KG.

These methodologies contribute to the construction of comprehensive and accurate KGs. In addition, they also pave the way for innovative applications in patient-centric healthcare.

4.1.4.4 Axioms and rules utilization

PCKGs utilize axioms and rules to enhance the representation and analysis of patient data, leading to more informed healthcare decisions. Axioms in PCKGs are fundamental principles or statements accepted as accurate without proof and used to define relationships and properties within the graph. Conversely, rules are logical statements that infer new knowledge from existing data within the graph. Combining axioms and rules is essential in structuring and interpreting complex medical data, enabling more personalized and effective patient care.

The methodologies for creating axioms and rules in PCKGs vary across research works. Chu et al. (2021) demonstrated the adaptation of multi-center clinical datasets by incorporating an external KG, which enhanced patient features and improved predictions for acute kidney injury in heart failure patients. This approach signifies the importance of integrating diverse data sources and knowledge bases into PCKGs. In another study, Tomczak and Gonczarek (2012) developed a Graph-based Rules Inducer for extracting decision rules from data streams in diabetes treatment. This method supported medical interviews by tracking hidden context changes and avoiding overfitting, highlighting the importance of dynamic rule adaptation in response to evolving data. Shang et al. (2021) utilized an EHR-oriented KG system to effectively harness non-used information buried in EHRs, thereby improving healthcare quality and providing interpretable recommendations for specialist physicians.

The use of graph-based association rules for mining medical databases, as demonstrated by Alzoubi (2013), revealed the potential of discovering hidden medical knowledge, which could contribute significantly to understanding complex medical conditions like preterm birth. Similarly, RuleHub, as introduced by Ahmadi et al. (2020), is an extensible corpus of rules for public KGs, enabling users to archive and retrieve rules from popular KGs. This tool improves data understanding and reduces redundant work, illustrating the importance of rule-sharing and standardization in the field. Wright et al. (2011) developed a set of rules for inferring patient problems from clinical and billing data, which performed better than using a problem list or billing data alone. This method improved clinical decision support and care improvement, showcasing the practical application of rule-based systems in clinical settings.

Based on our review, using axioms and rules in PCKGs is a dynamic and evolving field, with methodologies ranging from integrating diverse data sources to developing specialized, disease-centric graphs. These approaches enhance the understanding of complex medical data and significantly advance personalized patient care.

4.2 PCKG evaluation

Evaluation of PCKGs is critical for ensuring their efficacy and accuracy in representing and inferring medical knowledge. These evaluations are typically conducted using two principal methodologies: *qualitative* and *quantitative* assessments. Qualitative assessment involves a detailed examination of the KG to ensure it aligns with clinical best practices and accurately reflects medical relationships. This method may utilize usability studies, content analysis, and expert panel reviews to gage the graph's relevance and correctness (Rotmensch et al., 2017; Brundage et al., 2015). On the other hand,

quantitative assessment employs statistical and computational techniques to measure performance metrics such as accuracy, recall, and precision, providing a numerical evaluation of the graph's effectiveness (Rotmensch et al., 2017).

4.2.1 Quantitative assessment

Quantitative methodologies for PCKG evaluation focus on numerical metrics to assess the graph's performance. These methods include *completeness, consistency, accuracy, and embedding techniques*.

Completeness in a KG refers to the extent to which all necessary information is represented. In healthcare, this means ensuring that a PCKG includes comprehensive data on diseases, symptoms, treatments, and patient histories. A systematic literature review by Issa et al. (2021) emphasizes the importance of assessing the completeness of KGs and identifying various methodologies and metrics used for this purpose. On the other hand, consistency involves checking for logical coherence within the graph, particularly in terms of medical terminologies and relationships. This ensures that the KG does not contain contradictory information, which is crucial for clinical decision-making. The work of Sharma and Bhatt (2022) on privacy-preserving KGs in healthcare highlights the importance of maintaining consistency in data representation.

Other methods, such as accuracy assessment and embedding techniques, are integral to developing reliable PCKGs. The accuracy of a KG, as highlighted by Li et al. (2020), is crucial for patient safety and effective treatment planning, ensuring that the information aligns with authoritative medical databases and literature. Complementing this, embedding techniques, such as node2vec or GraphSAGE, as demonstrated by Talukder et al. (2022), play a vital role in transforming complex graph data into a more accessible format by capturing and representing semantic relationships. Moreover, the results of these embedding techniques can be quantitatively evaluated through tasks like link prediction, triple classification, and clustering, which provide metrics to assess the quality and effectiveness of the embeddings in representing the underlying knowledge within the PCKG.

4.2.2 Qualitative assessment

Qualitative assessment methodologies are instrumental in evaluating the effectiveness of PCKGs in meeting the needs of their intended users, including clinicians, researchers, and patients.

In the qualitative assessment of PCKGs, usability studies and feedback loops play an essential role in the KG evaluation. Usability studies, as highlighted by Brundage et al. (2015), are crucial for evaluating the ease of use and accessibility of PCKGs, particularly in the context of interpreting patient-reported outcomes (PROs) in graphic format. These studies typically employ a blend of qualitative and quantitative methods, including surveys, interviews, and user testing to gather comprehensive feedback on the user experience. Complementing this, feedback loops, as discussed by Rospocher et al. (2016), enable users to provide direct input on their experiences with the KG. This feedback is instrumental in continuously refining and adapting the KG to ensure it aligns closely with user needs and preferences.

Furthermore, the quality of PCKGs is also gaged through comparison with established databases and expert validation. Benchmarking PCKGs against renowned medical databases like PubMed and ClinicalTrials.gov is essential for assessing their content coverage and accuracy (Huang et al., 2017). This comparison helps

identifying any gaps or discrepancies in the PCKG. Additionally, expert validation, underscored by the work of Rotmensch et al. (2017), involves domain experts who ensure the medical relevance and accuracy of the PCKG's content. Their insights into the clinical applicability of the information within the KG are vital for maintaining its reliability and trustworthiness.

The evaluation of PCKGs through qualitative and quantitative assessments is a multifaceted process that involves reviewing medical content and analyzing performance metrics. Progress in this field has benefitted medical professionals and patients by ensuring that PCKGs serve as reliable and effective tools.

4.3 PCKG processing

By processing PCKGs, healthcare providers can extract actionable insights that inform clinical decisions. The utilization of PCKGs involves different methods, including reasoning, semantic search, and inference.

4.3.1 Reasoning

KG Reasoning in healthcare extends beyond mere data aggregation, playing an essential role in deriving new insights and facilitating informed decision-making. This reasoning process involves structuring information, extracting features and relations, and performing logical deductions to uncover new knowledge from existing data within the graph (Rajabi and Kafaie, 2022). For instance, in mental healthcare, KGs have been utilized for emotion recognition from facial expressions and heart rate, demonstrating the potential of knowledge reasoning in predicting emotional states (Wenying et al., 2020). Moreover, KGs in healthcare support clinical decision-making and enhance hospital efficiency by integrating heterogeneous medical knowledge and services, thereby enabling cognitive computing and semantic reasoning (Ding et al., 2021).

Furthermore, integrating LLMs with KGs can significantly improve patient engagement by empowering patients to navigate the healthcare system more effectively. With their NLP capabilities, LLMs facilitate patient-provider communication by allowing patients to ask questions and receive personalized responses tailored to their health concerns (Kassner et al., 2023). This patient-centric approach enhances the patient experience and fosters a collaborative environment where patients are more actively involved in their care decisions (Hou et al., 2023).

4.3.2 Semantic search

It enables the retrieval of specific information by structuring searches to navigate the complex relationships within the graph. Various strategies have been adopted to enhance the precision and efficiency of information retrieval in PCKGs. For instance, using Semantic Web and Knowledge Management approaches has been pivotal in implementing patient-centric strategies with well-defined semantics (Lytras et al., 2009). Graph-based methods incorporating semantic-rich knowledge bases and lazy learning algorithms have shown promise in linking multimodal clinical data for improved diagnosis performance (Wang et al., 2018). Moreover, the retrieval of similar clinical cases has been refined by mapping text to Unified Medical Language System (UMLS) concepts and representing patient records as semantic graphs, demonstrating superiority over traditional models (Plaza and Díaz, 2010).

Traditional semantic search methods often rely on structured queries and entity-centric KGs, focusing primarily on discrete entities and their relationships. However, these approaches can be limited in capturing the complexities of patient data, which often includes temporal and contextual information. Recent advancements in knowledge graph technology, particularly in integrating event-centric data, have begun to address these limitations. For instance, highlight the importance of event-centric knowledge graphs, which can provide richer contextual information often missing in traditional entity-centric models (Gottschalk and Demidova, 2018). Similarly, it emphasizes that existing knowledge graphs frequently overlook temporal relationships, which is crucial for understanding patient histories and treatment timelines (Costa et al., 2020).

Integrating LLMs, such as GPT-3, into the semantic search framework of PCKGs presents a transformative opportunity to advance search capabilities beyond conventional methods. LLMs excel in understanding and generating human-like text, making them powerful tools for interpreting complex queries and retrieving relevant information from knowledge graphs. The combination of LLMs with PCKGs not only enhances the interpretability and contextual relevance of search results but also personalizes the information retrieved to align with patient-specific contexts, significantly improving the quality of healthcare decisions (Wilhelm et al., 2023). This advancement is particularly critical in healthcare, where the subtleties of patient data can greatly influence treatment outcomes. Furthermore, LLMs hold promise in supporting mental healthcare by providing contextually relevant information while ensuring clinicians remain central to the decision-making process (Torous and Blease, 2023).

4.3.3 Inference

Inference uses the graph's inherent structure to deduce new insights, facilitating the discovery of patterns and trends that may not be immediately apparent (Wang et al., 2018; Lytras et al., 2009). One prominent method for inference in KGs is using ontology-based reasoning. For instance, they demonstrated the effectiveness of ontology-based inference by utilizing the Elk reasoner to derive new triples from an existing knowledge graph, significantly expanding its informational scope (Alshahrani et al., 2017). This approach is particularly relevant in healthcare, where medical ontologies can encapsulate complex relationships among diseases, symptoms, and treatments. The ability to infer new knowledge from existing data allows healthcare providers to uncover insights that may not be immediately apparent from raw data alone. Moreover, the integration of statistical relational learning (SRL) techniques into knowledge graphs has been shown to enhance inference capabilities and reviewed various SRL techniques that can be applied to large-scale knowledge graphs, highlighting their potential to improve the accuracy and completeness of inferences made from these graphs (Nickel et al., 2016). This is particularly important in PCKGs, where the relationships between patient data points can be intricate and multifaceted. By leveraging SRL, PCKGs can provide more robust predictions and recommendations based on the underlying data. In addition to traditional reasoning methods, machine learning approaches, particularly those involving neural networks, have gained traction in KG inference. For example, recent advancements in link prediction algorithms based on neural embeddings have shown promise in completing sparse KGs (Agibetov and Samwald, 2020). These

algorithms can identify missing relationships between entities in a PCKG, enhancing the graph's utility for clinical applications. Furthermore, the work emphasizes the importance of effectively learning over multiple graphs to construct a unified representation, which is crucial for comprehensive patient-centric applications (Trivedi et al., 2018). Fuzzy inference systems (FIS) have also emerged as a valuable technique in PCKGs, particularly for diagnosing diseases. As highlighted, FIS can facilitate accurate and early diagnosis by utilizing fuzzy logic to interpret patient data (Chuan et al., 2022). This method is particularly beneficial in healthcare settings where uncertainty and variability in patient data are prevalent. By incorporating fuzzy inference into PCKGs, healthcare providers can better understand patient conditions and tailor interventions accordingly.

These methods underscore the advances in accuracy and efficiency in the processing and utilization of PCKGs, leading to enhanced healthcare outcomes.

After examining the methodologies for building, evaluating, and processing PCKGs, we move on to practical applications and use cases to demonstrate how PCKGs can be applied in real-world scenarios, showing their impact and significance in healthcare.

5 Applications and use cases

As the healthcare industry increasingly embraces data-driven decision-making, PCKGs have emerged as a powerful tool for personalized medicine. These KGs, which place the patient at the center of a complex network of interconnected health data, offer a holistic view of a patient's health journey. They encompass various data, from medical history and genetic information to lifestyle factors and environmental influences. The applications of these KGs are broad and innovative, offering the potential for improved disease prediction, enhanced treatment planning, and more effective preventive strategies. In the following subsections, we will explore these applications in detail, shedding light on how PCKGs revolutionize healthcare and pave the way for a more personalized, predictive, and proactive approach to patient care.

5.1 Predicting disease before onset

The healthcare sector has seen significant advancements in patient-centric solutions by integrating KGs. Using a KG to proactively forecast the likelihood of a disease developing in an individual before any symptoms appear involves synthesizing various data points related to patient history, genetic information, lifestyle factors, and broader medical knowledge to identify patterns and risk factors associated with diseases. This predictive model allows for early intervention strategies, significantly altering the disease trajectory and improving the patient's long-term health outcomes.

The methodologies for creating PCKGs for disease prediction are diverse and innovative. Chen et al. (2019) discuss the robust extraction of medical knowledge from EHRs to build graphs that evaluate accuracy across different diseases and patient demographics using non-linear functions for causal relationships. Similarly, Talukder et al. (2022) describe an AI-based approach integrating disease-related knowledge bodies with Node2VEC for link prediction in disease-symptom networks. Furthermore, Liang et al. (2022) highlight using

multi-hop reasoning over KGs, which provides interpretability and is superior to single-hop methods. Similarly, refining of medical KGs using latent representations aids in prediction and maintains the explainability of diagnoses (Heilig et al., 2022). Expanding on these advancements, Li et al. (2020) introduce a Graph Neural Network-Based Diagnosis Prediction (GNDP) model that uses spatial–temporal graph convolutional networks for diagnosis predictions. In a complementary manner, Zheng et al. (2021) proposed a Multi-modal Graph Learning framework (MMGL) that exploits multi-modality information for disease prediction.

Different strategies across research works include the use of random walk along KG (Sun et al., 2020), Graph Neural Networks (GNN) for embedding medical concepts (Sun et al., 2020), and hybrid systems combining KGs with clinical experience for pediatric disease prediction (Liu et al., 2017). Zhang (2021) built an automatic question-answering system based on medical KGs, while Saha et al. (2020) predicted missing and noisy links in clinical KGs using neighborhood-based embeddings. Key findings from our literature review indicate significant advances in disease prediction accuracy, efficiency, and outcomes. For instance, the integration of GNNs with EMR data has led to highly representative node embeddings that improve prediction accuracy (Sun et al., 2020).

The use of tensor factorization on biological KGs for predicting co-morbid disease pairs (Biswas et al., 2019) and the construction of KGs based on evidence-based medicine for diabetes complications (Wang et al., 2020) is a notable advance.

PCKGs represent a significant step forward in the field of disease prediction. They offer the potential to transform patient outcomes by enabling early detection and personalized treatment plans. Future implications include the continued refinement of these models to enhance their predictive power and the integration of even more diverse data sources to capture the full spectrum of patient health and disease progression. In Table 2 we summarize a list of selected disease prediction applications in terms of their impact and limitations in the field of PCKGs.

5.2 Recommending individualized interventions

Intervention recommendation aims to use KGs to suggest tailored medications and treatments. By mapping patients' unique health profiles, these graphs enable physicians to pinpoint optimal therapies that enhance the precision of medical decisions, leading to improved patient outcomes. The application of KGs in personalized treatments is predicated on integrating diverse data sources, including EHRs, clinical notes, and patient-generated data, to construct a comprehensive, interconnected data structure that reflects individual patient profiles. This patient-centric approach is crucial as it allows for treatments to be tailored based on each patient's unique medical history, genetic information, lifestyle, and preferences, which is a departure from the one-size-fits-all healthcare model.

The approaches to developing and implementing PCKGs differ among research studies, but they all aim to achieve robustness, accuracy, and personalization. For instance, Gyrard et al. (2018) aggregate knowledge from IoT devices, clinical notes, and EMRs to manage chronic diseases, showcasing the integration of AI and machine learning in constructing Personalized Healthcare

TABLE 2 A summary of selected literature on PCKGs for *"Predicting Disease Before Onset."*

Paper title	Focus/objective	Contribution(s)	Limitation(s)
Predicting disease before onset			
A novel link prediction approach on clinical knowledge graphs utilizing graph structures (Dörpinghaus et al., 2022)	The goal is to create a proof of concept showcasing the efficacy of graph structures in AI methodologies	• Development of a graph-based method merging Conditional Random Fields (CRFs) and graph embedding for knowledge discovery • This method successfully predicts labels for graph nodes with high precision and recall	• The method is time-intensive when querying features from the graph, particularly with large datasets • Increased runtime and memory demands for multi-node paths in the graph
Deep knowledge reasoning guided disease prediction (Chuan et al., 2022)	The paper aims to enhance model interpretability by integrating knowledge entities with single-hop and multi-hop relationships	• Introduction of HitaNet, a hierarchical time-aware attention network • HitaNet uses a self-attention based transformer model for enhanced disease prediction • The paper presents a unique token wrapping method to merge knowledge graph insights with EHR data	• Lower time efficiency compared to baseline methods as a limitation • The proposed method's effectiveness diminishes with the availability of sufficient data
Predicting missing and noisy links via neighborhood preserving graph embeddings in a clinical knowledgebase (Saha et al., 2020)	The paper proposes a model that combines support vector classification (SVC) and neural network-based probabilistic embedding (NPE) to predict the links between clinical entities in the KG	• The development of a novel approach that integrates SVC and NPE for link prediction in KGs • The model achieved promising results in terms of predicting missing associations	• The reliance on existing clinical databases, which may contain noisy or incomplete data • The potential mismatch between model predictions and clinical recommendations
Relation prediction of co-morbid diseases using knowledge graph completion (Biswas et al., 2019)	The objective is to predict relationships between co-morbid diseases using knowledge graph completion techniques	• Proposing a method that combines distributed representation learning and graph embedding to predict disease-disease relationships • Development of a novel approach that leverages the structure and semantics of a KG to predict disease relationships	• The proposed method relies on the availability of a comprehensive and accurate knowledge graph • The evaluation of the method is limited to a specific dataset
Disease prediction via graph neural networks (Sun et al., 2020)	The paper addresses the challenges in predicting both common and rare diseases by integrating expert knowledge with machine learning techniques	• Proposing a systematic solution that combines expert knowledge with machine learning • Introducing a novel graph embedding-based model for disease prediction, which learns embeddings from medical concept graphs and patient record graphs	• Struggling to adapt to rare diseases due to data scarcity and complex symptom-diagnosis relationships • Reliance on historical patient records for model training limits the ability to serve new patients

Knowledge Graphs (PHKGs). Transitioning from a general approach to a more specialized application, Individualized Knowledge Graphs (IKGs) in cardiovascular medicine are one strategy for developing these KGs, which combine biological knowledge with medical histories and health outcomes to create personalized treatment strategies (Ping et al., 2017). Further refining the methodological framework, Rotmensch et al. (2017) utilized concept extraction and probabilistic models, finding the Noisy OR model particularly effective for constructing high-quality health KGs. Building on these methodologies, Shirai et al. (2021) applied Personal Knowledge Graphs (PKGs) to integrate patient-specific information into decision-making tools for personalized healthcare. In a similar vein of enhancing disease treatment strategies through personalized data, Zhu et al. (2022) introduced a KG that enhances rare disease (RD) treatment recommendations by systematically compiling and semantically annotating RD-related scientific articles, aggregating essential research findings and therapeutic insights with a sophisticated data model.

Literature has demonstrated significant advances in treatment accuracy, efficiency, and outcomes. For instance, the Four-Tuple Path Matrix in Traditional Chinese Medicine has been proposed to create personalized KGs, enhancing diagnostic modalities (Xie et al., 2018). Li et al. (2014) demonstrated how personal KGs could be automatically constructed from user utterances in conversational dialogs, indicating the potential for real-time, dynamic treatment adjustments. PCKGs provide multiple advantages. By analyzing patient profiles using KGs, treatments are more accurate and efficient than traditional methods. For example, Zhang et al. (2022) developed an intuitive graph representation of knowledge for nonpharmacological treatment of psychotic symptoms in dementia, potentially transforming care strategies for such complex conditions.

There are many potential implications for future personalized treatment and patient outcomes. A more informed and dynamic approach to treatment has the potential to enhance medical precision, improve patient engagement and optimize health outcomes by integrating PCKGs. These methods are expected to become more sophisticated as the field evolves, allowing for even greater personalization and efficacy in treatment. In Table 3, we summarize a list of selected treatment decision applications in terms of their impact and limitations in the field of PCKGs.

TABLE 3 A summary of selected literature on PCKGs for *"Recommending Individualized Interventions."*

Paper title	Focus/objective	Contribution(s)	Limitation(s)
Recommending individualized interventions			
Individualized knowledge graph (Ping et al., 2017)	• Envisioning individualized Knowledge Graphs (iKGs) in cardiovascular medicine • Proposing a modern informatics platform for transforming clinical and scientific discovery	• Introducing the concept of iKGs for aggregating and presenting individualized cardiovascular health data • Highlighting the role of iKGs in linking biological and clinical knowledge of individual patients	• Acknowledging challenges in data fragmentation, noncommensurability, and semantic inference within cardiovascular data
Personalized health knowledge graph (Gyrard et al., 2018)	Aims to manage chronic diseases more effectively using IoT data analytics and explicit knowledge	• Proposes a methodology to build PHKG, integrating heterogeneous data sources • Offers a solution for contextualizing and personalizing healthcare information	• The paper acknowledges the complexity in semantic integration of diverse data • It highlights the challenges in tailoring generic knowledge to individual patients
Developing an intuitive graph representation of knowledge for nonpharmacological treatment of psychotic symptoms in dementia (Zhang et al., 2022)	• Develop a knowledge graph for nonpharmacological treatment of psychotic symptoms in dementia • Enhance understanding and management of dementia-related psychotic symptoms through nonpharmacological methods	• Creation of the Dementia-Related Psychotic Symptom Nonpharmacological Treatment Ontology (DRPSNPTO) • Improvement in visualization and computerization of gerontological knowledge	
Learning a health knowledge graph from electronic medical records (Finlayson et al., 2014)	Automatically learn a health KG from EMRs to link diseases and symptoms and improve clinical decision-support systems	• A methodology for deriving health KG from EMR using probabilistic models • Demonstration that the noisy OR model significantly outperforms other tested models	• Inherent difficulties in interpreting EMR data, especially the presence of complex patient conditions • The reliance on rudimentary concept extraction pipelines • Limitations related to the automatic inference of causal relationships from observational data
Applying personal knowledge graphs to health (Shirai et al., 2021)	The paper focuses on leveraging PHKGs to enhance healthcare decision-making by integrating personal health information with broader knowledge graphs	• Proposing a conceptual framework for PHKGs, highlighting how they can support personalized, knowledge-driven healthcare applications by leveraging data from EHRs, IoT devices, and other health-related data sources	• Collecting and storing personal health knowledge from heterogeneous sources • Linking personal health knowledge to external KGs enhances the PHKG with broader contextual information • Maintaining the PHKG to ensure it remains up-to-date and accurate

5.3 Enhancing clinical trials

KGs represent a paradigm shift in clinical trial patient selection by offering a structured, interconnected data framework that can encapsulate complex patient information, medical histories, and potential trial criteria. In clinical trials, patient-centric approaches are crucial because personalized medicine tailors treatments to each patient's characteristics, necessitating a comprehensive understanding of patient data.

PCKGs are created and applied using diverse methodologies. Gortzis and Nikiforidis (2008) described an N-tier system that combines KGs with human collaboration and scalable knowledge engineering tactics. Expert input must be combined with scalable data structures to select patients effectively. Xiang et al. (2019) highlighted the standardization and structural integration provided by KGs, essential for auxiliary diagnosis systems in clinical trials. Strategies for developing KGs for patient selection in clinical trials include linking of multimodal data types for automatic diagnosis (Wang et al., 2018). Nicholson and Greene (2020) discussed machine learning methods for constructing low-dimensional representations of KGs, which support applications in genomics, pharmaceutical, and clinical domains.

Various studies have indicated significant improvements in the accuracy, efficiency, and outcomes of clinical trial patient selection. For instance, the Safe Medicine Recommendation (SMR) framework by Wang et al. (2017) bridges electronic medical records with medical KGs to learn patient-disease-drug embeddings, enhancing the precision of clinical trial patient selection. Compared with traditional ways of selecting patients for clinical trials, PCKGs offer a more thorough and subtle approach. Several studies have demonstrated the effectiveness of KGs in improving clinical trial design and outcomes by providing a holistic view of patient data, facilitating personalized trial matching, and providing a more holistic view of patient data (Sharma, 2015; Weng et al., 2017; Huang et al., 2017). Ultimately, integrating patient-centric KGs in selecting clinical trial participants can transform the field by improving the precision and personalization of patient care. As a result of a better understanding of patient data, future trials will likely be more adaptive, patients will be more engaged, and outcomes may be improved. In Table 4, we summarize a list of selected treatment decision applications in terms of their impact and limitations in the field of PCKGs.

TABLE 4 A summary of selected literature on PCKGs for "*Enhancing Clinical Trials.*"

Paper title	Focus/objective	Contribution(s)	Limitation(s)
Recommending Individualized Interventions			
Knowledge graph-based clinical decision support system reasoning (Xiang et al., 2019)	The focus of the paper is to highlight the benefits of using knowledge graphs over traditional hand-crafted rule databases in CDSSs	The introduction of the Path-Ranking Algorithm (PRA) as a method for automatically discovering symptoms without human intervention	• The probability of certain paths in the KG may not be accurate • The lack of details on the classification model used
Constructing knowledge graphs and their biomedical applications (Nicholson and Greene, 2020)	• Examining the construction and application of biomedical knowledge graphs • Emphasizing how machine learning is transforming these processes	• Discussion of knowledge graph construction, including manual curation and text mining • Review of representational learning techniques and their applications in biomedical fields	• Need for advanced techniques to handle complex sentence structures • Limitations in current methods to represent diverse relationships in KGs • Scalability and memory limitations in matrix factorization techniques
SMR: Medical knowledge graph embedding for safe medicine recommendation (Sharma, 2015)	Developing a framework to recommend safe medicines by leveraging a heterogeneous graph that integrates patient data, diseases, and medicines	• Development of graph-based embedding models enabling the recommendation of newly emerged medicines effectively • A novel method to recommend safe medicines for new patients and minimizing potential adverse drug reactions • Introduction of the SMR framework as a new approach to the link prediction problem	• Dealing with the challenge of recommending safe medicines, especially new ones, to patients • Minimizing potential adverse drug reactions in medicine recommendations is critical to patient safety
Patient centric approach for clinical trials: Current trend and new opportunities (Weng et al., 2017)	Exploring the shifting paradigm in clinical trials toward a more patient-centric model	Identifying new opportunities for the clinical research industry to adopt patient-centric approaches to accelerate drug development and improve trial outcomes	• The complexity and rising costs of clinical research • Ensuring data transparency and building trust with patients participating in clinical trials
Automatic diagnosis with efficient medical case searching based on evolving graphs (Lytras et al., 2009)	Developing a method for automatic diagnosis by improving medical case searching using evolving graphs, which dynamically incorporate new medical cases and knowledge	• Introduction of an evolving graph framework that integrates new medical cases and knowledge • A novel method for medical case searching that leverages the evolving graph structure • An optimization strategy for embedding learning in the heterogeneous graph	• Handling the dynamic nature of medical knowledge and cases • Balancing the computational complexity of embedding learning in a continuously evolving graph structure • Scalability and maintaining high accuracy and efficiency as the graph expands

Although PCKGs have proven valuable in recommending individual interventions, predicting disease before onset, and improving clinical trials, their utility goes far beyond these. Using KGs in other innovative healthcare applications is increasingly becoming common, such as optimizing hospital workflows, tailoring patient engagement strategies, and even developing telemedicine platforms. With these innovative tools, healthcare can be revolutionized by providing a more holistic, integrated view of patient data and new opportunities for research and treatment methods. Having explored the diverse applications and use cases of PCKGs in various domains, we now focus on this field's challenges and future directions. This transition allows us to critically examine the current limitations and envision potential advancements that could further enhance the utility and effectiveness of PCKGs.

6 Research challenges and discussion

PCKGs in healthcare are designed to provide a comprehensive, unified view of patient data by integrating information from various sources, including EHRs, medical literature, and patient-generated data. These graphs aim to support better clinical decision-making and personalized patient care by representing complex medical data in an interconnected format that is more accessible and actionable for healthcare providers.

The current state of research in PCKGs is focused on overcoming several key challenges to maximize their potential in healthcare. Ji et al. (2020) discuss the difficulties in knowledge acquisition, completion, and temporal KG development, which are crucial for maintaining up-to-date and comprehensive patient profiles. Building on this foundation, Chen et al. (2019) highlight the need for robustness in PCKGs, particularly in addressing sample size limitations and unmeasured confounders, to extend models to larger patient visits. Moreover, Rastogi and Zaki (2020) emphasize the importance of designing, building, and operationalizing PCKGs tailored to individual patients. However, a significant hurdle remains in the actual construction of PCKGs. As noted by Gyrard et al. (2018) and Cong et al. (2018), the complexities in constructing PCKGs are often time-consuming and heavily reliant on source data quality. A significant challenge in applying traditional KG embedding methods to patient-centric healthcare applications is their struggle with structural sparsity. Hu et al. (2021) argue that conventional techniques, such as TransE

and ConvE, while adept at mapping entities and relationships into a vector space, falter because they rely solely on KG triplets, neglecting the rich auxiliary texts that describe entities. This limitation significantly limits the comprehensiveness and utility of KGs in capturing detailed patient information and medical knowledge, indicating a key challenge in leveraging KGs for complex healthcare applications.

Data quality and standardization remain significant challenges, as heterogeneous data structures, poor data quality, and varying medical standards complicate data integration into a coherent KG (Zhang et al., 2020). Integrating PCKGs into clinical workflows also presents challenges, as it requires the development of systems that complement healthcare providers' routines without causing disruptions (Gortzis and Nikiforidis, 2008). Additionally, scalability is another concern, as PKGs must be able to incorporate an ever-increasing amount of data from diverse sources, including genomic information and patient lifestyle data (Wang et al., 2018). Patient data privacy is a critical concern, particularly in utilizing PCKGs, which involve handling sensitive personal health information and require strict privacy controls to safeguard patient confidentiality. Using patients' data for various purposes, such as consultations, research, and emergencies, poses a significant challenge for authorization systems, emphasizing the need for robust privacy protection (Al-Zubaidie et al., 2019; Sharma and Bhatt, 2022). Furthermore, real-time data analysis within PCKGs is technologically demanding, requiring advanced computational methods to process and analyze data promptly for it to be clinically relevant (Hooshafza et al., 2021).

Given these challenges, it becomes clear that continued technological and methodological innovations are necessary to enhance PCKGs' predictive capabilities. Advanced analytics, machine learning, and semantic web technologies could be key. Additionally, as PCKGs become more integrated into healthcare delivery, addressing regulatory and ethical considerations becomes increasingly essential. This requires collaboration among computer scientists, healthcare professionals, and policymakers to align the development of PCKGs with broader healthcare objectives.

7 Conclusion and future directions

This literature review of PCKGs explores their development, evaluation, processing techniques, applications, challenges, and prospects. PCKGs represent a field in healthcare informatics that aims to revolutionize personalized patient care by integrating and synthesizing diverse healthcare data sources. The review highlights the current state-of-the-art methodologies for constructing and evaluating PCKGs, emphasizing the importance of qualitative and quantitative approaches to assess their effectiveness in healthcare settings. In addition to construction and evaluation, the review delves into innovative processing techniques such as reasoning, semantic search, and inference. These techniques significantly enhance the accuracy and efficiency of PCKGs, ultimately improving patient-centered care. Furthermore, exploring different applications of PCKGs in healthcare—including disease prediction, personalized treatment recommendations, and advancements in clinical trials—reveals their potential to transform healthcare through personalized and predictive medicine.

Future directions in PCKG research include leveraging advanced analytics and machine learning to improve predictive capabilities, which could lead to more accurate and timely interventions (Wu, 2021). In addition, semantic web technologies are also predicted to play a significant role in enhancing the accessibility and utility of PCKGs (Huang et al., 2017). Building on this momentum, personalized medicine emerges as a promising area where PCKGs can substantially impact. By linking genomic data with clinical outcomes, treatments can be tailored to individual patients, offering a more personalized approach to healthcare (Chandak et al., 2023). To further this advancement, methodological innovations, such as new algorithms for data harmonization and user interface design, are crucial. These innovations are needed to address current challenges and facilitate the broader adoption of PCKGs in clinical practice (Peng et al., 2023). Simultaneously, as PCKGs become more integrated into healthcare delivery, regulatory and ethical considerations gain prominence. These aspects are critical to ensuring that the deployment of PCKGs adheres to the highest standards of patient care and data management. Therefore, cross-disciplinary collaboration becomes essential for advancing PCKG technology. This involves computer scientists, healthcare professionals, and policy-makers working together to ensure that the development of PCKGs aligns with broader healthcare objectives and respects ethical guidelines (Solanki et al., 2022; Maxhelaku et al., 2022).

PCKGs represent a significant advancement in healthcare informatics. They can transform patient outcomes by enabling early detection and personalized treatment plans. Despite the challenges involved, the future outlook for PCKGs is promising, as they have the potential to improve patient care and healthcare delivery significantly. Ongoing research efforts and interdisciplinary collaboration will be crucial in fully realizing their novel impact on healthcare.

Author contributions

HSK: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. SN: Formal analysis, Methodology, Writing – original draft, Writing – review & editing. HK: Writing – review & editing. NG: Writing – review & editing. SM: Writing – review & editing. AA: Writing – review & editing. SR: Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted without any commercial or financial relationships that could potentially create a conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abu-Salih, B. (2021). Domain-specific knowledge graphs: a survey. *J. Netw. Comput. Appl.* 185:103076. doi: 10.1016/j.jnca.2021.103076

Agibetov, A., and Samwald, M. (2020). Benchmarking neural embeddings for link prediction in knowledge graphs under semantic and structural changes. *J. Web Semant.* 64:100590. doi: 10.1016/j.websem.2020.100590

Aguilar-Escobar, V. G., Garrido-Vega, P., and González-Zamora, M. D. M. (2016). Applying the theory of constraints to the logistics service of medical records of a hospital. *Eur. Res. Manag. Bus. Econ.* 22, 139–146. doi: 10.1016/j.iedee.2015.07.001

Ahmadi, N., Truong, T.-T.-D., Dao, L.-H.-M., Ortona, S., and Papotti, P. (2020). Rulehub. *J. Data Informat. Qual.* 12, 1–22. doi: 10.1145/3409384

Alagar, V., Periyasamy, K., and Wan, K.. Privacy and security for patient-centric elderly health care. In 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 1–6 (2017).

Albannai, N., Gultepe, Y., and Najih, A. M. A. (2019). A review of machine learning applications in semantic web. *Int. J. Sci. Technol. Res.* doi: 10.7176/JSTR/5-8-01

Almunawar, M., and Anshari, M. (2012). Improving customer service in healthcare with crm 2.0. *ArXiv abs/1204.3685*. doi: 10.48550/arXiv.1204.3685

Alshahrani, M., Khan, M. A., Maddouri, O., Kinjo, A. R., Queralt-Rosinach, N., and Hoehndorf, R. (2017). Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics* 33, 2723–2730. doi: 10.1093/bioinformatics/btx275

Alzoubi, W. (2013). Mining medical databases using graph-based association rules. *Int. J. Mach. Learn. Comput.* 3, 294–296. doi: 10.7763/IJMLC.2013.V3.324

Al-Zubaidie, M., Zhang, Z., and Zhang, J. (2019). Pax: using pseudonymization and anonymization to protect patients' identities and data in the healthcare system. *Int. J. Environ. Res. Public Health* 16:1490. doi: 10.3390/ijerph16091490

Balog, K., and Kenter, T.. Personal knowledge graphs: a research agenda. In Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, 217–220 (2019).

Bauer-Mehren, A., Rautschka, M., Sanz, F., and Furlong, L. I. (2010). Disgenet: a cytoscape plugin to visualize, integrate, search and analyze gene–disease networks. *Bioinformatics* 26, 2924–2926. doi: 10.1093/bioinformatics/btq538

Biswas, S., Mitra, P., and Rao, K. S. (2019). Relation prediction of co-morbid diseases using knowledge graph completion. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 708–717. doi: 10.1109/TCBB.2019.2927310

Blobel, B. (2007). Educational challenge of health information systems' interoperability. *Methods Inf. Med.* 46, 52–56. doi: 10.1055/s-0038-1628132

Blobel, B. (2011). Ontology driven health information systems architectures enable phealth for empowered patients. *Int. J. Med. Inform.* 80, e17–e25. doi: 10.1016/j.ijmedinf.2010.10.004

Bordes, A., Weston, J., Collobert, R., and Bengio, Y. (2011). Learning structured embeddings of knowledge bases. *Proceedings of the AAAI conference on artificial intelligence* 25, 301–306. doi: 10.1609/aaai.v25i1.7917

Bravo, M. (2019). Luis Fernando Hoyos Reyes, and Jose´ a Reyes Ortiz. Methodology for ontology design and construction. *Contaduría y Administración* 64:134. doi: 10.22201/fca.24488410e.2020.2368

Brundage, M., Smith, K., Little, E., Bantug, E., and Snyder, C. (2015). Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. *Qual. Life Res.* 24, 2457–2472. doi: 10.1007/s11136-015-0974-y

Byeon, S., and Lee, W. (2023). Directed acyclic graphs for clinical research: a tutorial. *J. Minimally Invasive Surg.* 26, 97–107. doi: 10.7602/jmis.2023.26.3.97

Chandak, P., Huang, K., and Zitnik, M. (2023). Building a knowledge graph to enable precision medicine. *Sci. Data* 10:67. doi: 10.1038/s41597-023-01960-3

Chen, I., Agrawal, M., Horng, S., and Sontag, D. (2019). Robustly extracting medical knowledge from ehrs: a case study of learning a health knowledge graph. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 25, 19–30.

Choi, H. S., Song, J. Y., Shin, K. H., Chang, J. H., and Jang, B.-S. (2023). Developing prompts from large language model for extracting clinical information from pathology and ultrasound reports in breast cancer. *Radiat. Oncol. J.* 41, 209–216. doi: 10.3857/roj.2023.00633

Chu, J., Chen, J., Chen, X., Dong, W., Shi, J., and Huang, Z. (2021). Knowledge-aware multi-center clinical dataset adaptation: problem, method, and application. *J. Biomed. Inform.* 115:103710. doi: 10.1016/j.jbi.2021.103710

Chuan, P. M., Lan, L. T. H., Tuan, T. M., Tan, N. H., Long, C. K., Van Hai, P., et al. (2022). Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics* 33, 83–88. doi: 10.15439/2022R35

Clarkson, K., Gentile, A. L., Gruhl, D., Ristoski, P., Terdiman, J., and Welch, S.. User-centric ontology population. In A. Gangemi, Navigli, R., Vidal, M. E., Hitzler, P., Troncy, R., Hollink, L., et al., editors, The semantic web. ESWC 2018 lecture notes in computer science. *10843*, 112–127. Springer, Cham, (2018).

Cong, Q., Feng, Z., Li, F., Zhang, L., Rao, G., and Tao, C.. Constructing biomedical knowledge graph based on semmeddb and linked open data. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Madrid, Spain: 1628–1631. IEEE (2018). doi: 10.1109/BIBM.2018.8621568

Costa, T. S., Gottschalk, S., and Demidova, E.. Event-qa: a dataset for event-centric question answering over knowledge graphs. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 3157–3164. Association for Computing Machinery, New York, NY, United States. (2020). doi: 10.1145/3340531.3412760

Cyganiak, R., Wood, D., and Lanthaler, M.. Rdf 1.1 concepts and abstract syntax. World Wide Web Consortium (W3C) (2014). (Accessed August 29, 2023).

Ding, Y., Arsintescu, B., Chen, C. H., Feng, H., Scharffe, F., Seneviratne, O., et al. International workshop on knowledge graph: Heterogenous graph deep learning and applications. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 4121–4122 (2021).

Djedidi, R., and Aufaure, M.-A. (2007). "Medical domain ontology construction approach: a basis for medical decision support" in Twentieth IEEE international symposium on computer-based medical systems (CBMS'07) (Maribor, Slovenia: IEEE), 509–511. doi: 10.1109/CBMS.2007.69

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 601–610 (2014).

Dörpinghaus, J., Hübenthal, T., and Faber, J.. A novel link prediction approach on clinical knowledge graphs utilising graph structures. In 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS), 43–52 (2022).

Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q., and Lin, Z.. Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. In 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 327–332 (2017).

Eggert, M., and Alberts, J. (2020). Frontiers of business intelligence and analytics 3.0: a taxonomy-based literature review and research agenda. *Bus. Res.* 13, 685–739. doi: 10.1007/s40685-020-00108-y

Ehrlinger, L., and Wolfram, W. Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS), 48:2 (2016).

Ellison, G. T. H. (2023). Using directed acyclic graphs (DAGs) to represent the data generating mechanisms of disease and healthcare pathways: A guide for educators, students, practitioners and researchers: Springer International Publishing, 61–101. doi: 10.1007/978-3-031-26010-0_6

El-Sappagh, S., Franda, F., Ali, F., and Kwak, K. (2018). Snomed ct standard ontology based on the ontology for general medical science. *BMC Med. Inform. Decis. Mak.* 18:76. doi: 10.1186/s12911-018-0651-5

Ferilli, S. (2021). Integration strategy and tool between formal ontology and graph database technology. *Electronics* 10:2616. doi: 10.3390/electronics10212616

Finlayson, S. G., LePendu, P., and Shah, N. H. (2014). Building the graph of medicine from millions of clinical narratives. *Sci. Data* 1, 1–9. doi: 10.1038/sdata.2014.32

Ge, J., Li, M., Delk, M. B., and Lai, J. C. (2023). A comparison of large language model versus manual chart review for extraction of data elements from the electronic health record. *medRxiv.* doi: 10.1101/2023.08.31.23294924

Ghosh, V. E., and Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia* 53, 104–114. doi: 10.1016/j.neuropsychologia.2013.11.010

Gilboa, A., and Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* 21, 618–631. doi: 10.1016/j.tics.2017.04.013

Goniewicz, K., Carlström, E., Hertelendy, A. J., Burkle, F. M., Goniewicz, M., Lasota, D., et al. (2021). Integrated healthcare and the dilemma of public health emergencies. *Sustain. For.* 13:4517. doi: 10.3390/su13084517

Gortzis, L., and Nikiforidis, G. (2008). Tracing and cataloguing knowledge in an e-health cardiology environment. *J. Biomed. Inform.* 41, 217–223. doi: 10.1016/j.jbi.2007.09.001

Gottschalk, S., and Demidova, E. (2018). "Eventkg: a multilingual event-centric temporal knowledge graph" in The semantic web (New York, NY, United States: Springer International Publishing), 272–287.

Guan, F., and Tezuka, T. (2022). "A medical q&a system with entity linking and intent recognition" in 2022 IEEE symposium series on computational intelligence (SSCI), 820–829.

Gyrard, A., Gaur, M., Thirunarayan, K., Sheth, A., and Shekarpour, S.. Personalized health knowledge graph. In: CEUR Workshop Proceedings, 2317 (2018).

Hafsah, S. S., Zakaria, L. Q., and Naswir, A. F. (2023). Rparallel-based corpus annotation for malay health documents. *Appl. Sci.* 13:13129. doi: 10.3390/app132413129

Hanna, J., Joseph, E., Brochhausen, M., and Hogan, W. R. (2013). Building a drug ontology based on rxnorm and other sources. *J. Biomed. Semant.* 4:44. doi: 10.1186/2041-1480-4-44

Harper, L., and Honour, M. (2015). Working together to personalize "the patient". *Newborn Infant Nurs Rev* 15, 21–23. doi: 10.1053/j.nainr.2015.01.004

He, Q., Yang, J., and Shi, B.. Constructing knowledge graph for social networks in a deep and holistic way. Companion Proceedings of the Web Conference (2020), 307–308 2020.

Heilig, N., Kirchhoff, J., Stumpe, F., Plepi, J., Flek, L., and Paulheim, H. (2022). Refining diagnosis paths for medical diagnosis based on an augmented knowledge graph. *ArXiv abs/2204.13329*. doi: 10.48550/arXiv.2204.13329

Hooshafza, S., Orlandi, F., Flynn, R., McQuaid, L., Stephens, G., and O'Connor, L. (2021). Modelling temporal data in knowledge graphs: a systematic review protocol. *HRB Open Res.* 4:101. doi: 10.12688/hrbopenres.13403.1

Hou, Y., Yeung, J., Hua, X., Chang, S., Wang, F., and Zhang, R. (2023). From answers to insights: unveiling the strengths and limitations of ChatGPT and biomedical knowledge graphs. *J. Healthc. Inform. Res.* doi: 10.1101/2023.06.09.23291208

Hu, L., Zhang, M., Li, S., Shi, J., Shi, C., Yang, C., et al. (2021). Text-graph enhanced knowledge graph representation learning. *Front. Artif. Intell.* 4:697856. doi: 10.3389/frai.2021.697856

Huang, Z., Yang, J., Harmelen, F. V., and Hu, Q. (2017). "Constructing disease-centric knowledge graphs: a case study for depression (short version)" in Artificial Intelligence in Medicine (Springer: Springer International Publishing), 48–52. doi: 10.1007/978-3-319-59758-4_5

Iannacone, M. D., Bohn, S., Nakamura, G. C., Gerth, J., Huffer, K. M. T., Bridges, R., et al. (2015). "Developing an ontology for cyber security knowledge graphs" in Proceedings of the 10th annual cyber and information security research conference (New York, NY, United States: ACM), 1–4. doi: 10.1145/2746266.2746278

Iscoe, M., Socrates, V., Gilson, A., Chi, L., Li, H., Huang, T., et al. (2023). Identifying signs and symptoms of urinary tract infection from emergency department clinical notes using large language models. *medRxiv*. doi: 10.1101/2023.10.20.23297156

Issa, S., Adekunle, O., Hamdi, F., Cherfi, S., Dumontier, M., and Zaveri, A. (2021). Knowledge graph completeness: a systematic literature review. *IEEE Access* 9, 31322–31339. doi: 10.1109/ACCESS.2021.3056622

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S. (2020). A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 494–514. doi: 10.1109/TNNLS.2021.3070843

Ji, Q., Qi, G., Gao, H., and Wu, T. (2018). "Survey on schema induction from knowledge graphs" in Communications in Computer and Information Science, 136–142.

Jiang, G., Ogasawara, K., Endoh, A., and Sakurai, T. (2003). Context-based ontology building support in clinical domains using formal concept analysis. *Int. J. Med. Inform.* 71, 71–81. doi: 10.1016/S1386-5056(03)00092-3

Johnson, K. B., Wei, W.-Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., et al. (2021). Precision medicine, AI, and the future of personalized health care. *Clin. Trans. Sci.* 14, 86–93. doi: 10.1111/cts.12884

Kapoor, B., and Sharma, S. (2010). A comparative study ontology building tools for semantic web applications. *Int. J. Web Semant. Technol.* 1, 1–13. doi: 10.5121/ijwest.2010.1301

Kassner, N., Tafjord, O., Sabharwal, A., Richardson, K., Schuetze, H., and Clark, P.. Language models with rationality. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 14190–14201. (Singapore: Association for Computational Linguistics) (2023). doi: 10.18653/v1/2023.emnlp-main.877

Keretna, S., Lim, C. P., and Creighton, D.. Enhancement of medical named entity recognition using graph-based features. In 2015 IEEE International Conference on Systems, Man, and Cybernetics, 1895–1900 (2015).

Kranjec, A., Ianni, G., and Chatterjee, A. (2013). Schemas reveal spatial relations to a patient with simultanagnosia. *Cortex* 49, 1983–1988. doi: 10.1016/j.cortex.2013.03.005

Li, Y., Qian, B., Zhang, X., and Liu, H. (2020). Graph neural network-based diagnosis prediction. *Big Data* 8, 379–390. doi: 10.1089/big.2020.0070

Li, X., Tur, G., Hakkani-Tür, D., and Li, Q. (2014). "Personal knowledge graph population from user utterances in conversational understanding" in 2014 IEEE Spoken Language Technology Workshop (SLT), 224–229.

Li, A., Wang, X., Wang, W., Zhang, A., and Li, B. (2019). "A survey of relation extraction of knowledge graphs" in Web and Big Data. APWeb-WAIM 2019. Lecture Notes in Computer Science, 52–66.

Li, L., Wang, P., Yan, J., Wang, Y., Li, S., Jiang, J., et al. (2020). Real-world data medical knowledge graph: construction and applications. *Artif. Intell. Med.* 103:101817. doi: 10.1016/j.artmed.2020.101817

Li, S., Wong, K. K. W., Fung, L. C. C., and Zhu, D.. Improving question answering over knowledge graphs using graph summarization. International Conference on Neural Information Processing (2022).

Li, L., Zhao, J., Hou, L., Zhai, Y. K., Shi, J., and Cui, F. (2019). An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records. *BMC Med. Inform. Decis. Mak.* 19:235. doi: 10.1186/s12911-019-0933-6

Liang, Y., Wang, H., and Zhang, W.. Deep knowledge reasoning guided disease prediction. In 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2695–2702 (2022).

Lin, Y., Keming, L., Sheng, Y., Cai, T., and Zitnik, M. (2023). Multimodal learning on graphs for disease relation extraction. *J. Biomed. Inform.* 143:104415. doi: 10.1016/j.jbi.2023.104415

Lin, P., Song, Q., and Yinghui, W. (2018). Fact checking in knowledge graphs with ontological subgraph patterns. *Data Sci. Eng.* 3, 341–358. doi: 10.1007/s41019-018-0082-4

Liu, S., and Fang, Y. Use large language models for named entity disambiguation in academic knowledge graphs. In Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023), 681–691. Atlantis Press (2023). doi: 10.2991/978-94-6463-264-4_79

Liu, K., Wang, F., Ding, Z., Liang, S., Zhengfei, Y., and Zhou, Y. (2022). A review of knowledge graph application scenarios in cyber security. *ArXiv abs/2204.04769*. doi: 10.48550/arXiv.2204.04769

Liu, X., Wang, H., He, T., Liao, Y., and Jian, C. (2022). Recent advances in representation learning for electronic health records: a systematic review. *J. Phys. Conf. Ser.* 2188:012007. doi: 10.1088/1742-6596/2188/1/012007

Liu, P., Wang, X., Sun, X., Shen, X., Chen, X., Sun, Y., et al. (2017). "Hkdp: a hybrid knowledge graph based pediatric disease prediction system" in Smart health (Springer International Publishing), 78–90. doi: 10.1007/978-3-319-59858-1_8

Lv, X., Guan, Y., Yang, J., and Wu, J. (2016). Clinical relation extraction with deep learning. *Int. J. Hybrid Informat. Technol.* 9, 237–248. doi: 10.14257/ijhit.2016.9.7.22

Lytras, M. D., Sakkopoulos, E., and Pablos, P. O. (2009). Semantic web and knowledge management for the health domain: state of the art and challenges for the seventh framework programme (fp7) of the european union (2007-2013). *Int. J. Technol. Manag.* 47, 239–249. doi: 10.1504/IJTM.2009.024124

MacLean, F. (2021). Knowledge graphs and their applications in drug discovery. *Expert Opin. Drug Discov.* 16, 1057–1069. doi: 10.1080/17460441.2021.1910673

Mathew, G., and Obradovic, Z. (2011). "Constraint graphs as security filters for privacy assurance in medical transactions" in Proceedings of the 2011 ACM symposium on applied computing (ACM), 502–504.

Matsuto, T., and Shiina, S. Laboratory test scheduling based on constraint propagation. In Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1334–1335. Philadelphia, PA, USA: IEEE (1990). doi: 10.1109/IEMBS.1990.691775

Maxhelaku, A. K. S.. Improving interoperability in healthcare using hl7 fhir. In: IISES International Academic Conference (2019).

Maxhelaku, S., Xhumari, E., and Xhina, E. (2022). "Exploiting knowledge graphs in healthcare" in Proceedings of the 16th Economics and Finance, 111–118.

Mendes, D., Rodrigues, I., and Baeta, C. F. Development and population of an elaborate formal ontology for clinical practice knowledge representation. In Proceedings of the International Conference on Knowledge Engineering and Ontology Development (IC3K 2013) - KEOD, 286–292. Vilamoura, Portugal: SciTePress (2013) doi: 10.5220/0004548602860292

Mesko, B., and Debronkart, D. (2022). Patient design: the importance of including patients in designing health care. *J. Med. Internet Res.* 24:e39178. doi: 10.2196/39178

Nastase, V.and Kotnis, B.. Abstract graphs and abstract paths for knowledge graph completion. Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019) (2019).

Newell, A., Shaw, J. C., and Simon, H. A. (1959). "Report on a general problem solving program" in IFIP Congress, vol. *256* (Pittsburgh, PA), 64.

Nicholson, D. N., and Greene, C. (2020). Constructing knowledge graphs and their biomedical applications. *Comput. Struct. Biotechnol. J.* 18, 1414–1428. doi: 10.1016/j.csbj.2020.05.017

Nickel, M., Murphy, K., Tresp, V., and Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proc. IEEE* 104, 11–33. doi: 10.1109/JPROC.2015.2483592

Noy, N. F., and McGuinness, D. L.. Ontology development 101: A guide to creating your first ontology. Available at: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (2023) (Accessed July 77, 2023).

Ognyanov, D., and Kiryakov, A. (2002). "Tracking changes in rdf(s) repositories" in Knowledge engineering and knowledge management: Ontologies and the semantic web. EKAW 2002 lecture notes in computer science. eds. A. Go'mez Pe'rez and V. R. Benjamins, vol. *2473* (Berlin, Heidelberg: Springer), 373–378.

Ohira, Y., Hochin, T., and Nomiya, H. (2011). "Introducing specialization and generalization to a graph-based data model" in Knowledge-based and intelligent information and engineering systems. KES 2011 lecture notes in computer science (Berlin, Heidelberg: Springer), 1–13.

Paulheim, H. (2017). Knowledge graph refinement: a survey of approaches and evaluation methods. *Semant. Web* 8, 489–508. doi: 10.3233/SW-160218

Peng, C., Xia, F., Naseriparsa, M., and Osborne, F. (2023). Knowledge graphs: opportunities and challenges. *Artif. Intell. Rev.* 56, 13071–13102. doi: 10.1007/s10462-023-10465-9

Ping, P., Watson, K., Han, J., and Bui, A. (2017). Individualized knowledge graph: a viable informatics path to precision medicine. *Circ. Res.* 120, 1078–1080. doi: 10.1161/CIRCRESAHA.116.310024

Plaza, L., and Díaz, A. (2010). "Retrieval of similar electronic health records using UMLS concept graphs" in Natural language processing and information systems, 296–303.

Pressat Laffouilhère, T., Grosjean, J., Bénichou, J., Darmoni, S. J., and Soualmia, L. F. (2022). Ontobiostat: supporting causal diagram design and analysis. *Stud. Health Technol. Inform.* 294, 302–306. doi: 10.3233/SHTI220463

Pugh, Z. H., and Gillan, D. (2020). "Propositional constraint graphs: an intuitive, domain-general tool for diagramming knowledge, assumptions, and uncertainties" in Proceedings of the human factors and ergonomics society annual meeting, vol. *64* (SAGE Publications), 254–258. doi: 10.1177/1071181320641060

Puustjarvi, J., and Puustjarvi, L.. Providing relevant health information to patient-centered healthcare. In The 12th IEEE International Conference on e-Health Networking, Applications and Services, 215–220. IEEE (2010). doi: 10.1109/health.2010.5556569

Radevski, G. (2023). "Linking surface facts to large-scale knowledge graphs" in Proceedings of the 2023 conference on empirical methods in natural language processing.

Raghupathi, W., and Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Informat. Sci. Syst.* 2:2. doi: 10.1186/2047-2501-2-3

Rajabi, E., and Kafaie, S. (2022). Knowledge graphs and explainable ai in healthcare. *Information* 13:459. doi: 10.3390/info13100459

Rastogi, N., and Zaki, M. J. (2020). Personal health knowledge graphs for patients. *ArXiv abs/2004.00071*. doi: 10.48550/arXiv.2004.00071

Richens, R. H. (1956). Preprogramming for mechanical translation. *Mech. Transl. Comput. Linguist.* 3, 20–25.

Rishi Kanth Saripalle (2019). Fast health interoperability resources (fhir): current status in the healthcare system. *IJEHMC* 10:18.

Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., et al. (2016). Building event-centric knowledge graphs from news. *J. Web Semant.* 37-38, 132–151. doi: 10.1016/j.websem.2015.12.004

Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., and Sontag, D. (2017). Learning a health knowledge graph from electronic medical records. *Sci. Rep.* 7:5994. doi: 10.1038/s41598-017-05778-z

Ruan, C., Wu, Y., Luo, G. S., Yang, Y., and Ma, P. (2020). Relation extraction for Chinese clinical records using multi-view graph learning. *IEEE Access* 8, 215613–215622. doi: 10.1109/ACCESS.2020.3037086

Saha, B., Mandloi, V., and Ghosh, S.. Predicting missing and noisy links via neighbourhood preserving graph embeddings in a clinical knowlegebase. In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 1124–1131 (2020).

Sarwar, T., Seifollahi, S., Chan, J., Zhang, X., Aksakalli, V., Hudson, I., et al. (2022). The secondary use of electronic health records for data mining: data characteristics and challenges. *ACM Comput. Surv.* 55, 1–40.

Shang, Y., Tian, Y., Zhou, M., Zhou, T., Lyu, K., Wang, Z., et al. (2021). EHR-oriented knowledge graph system: toward efficient utilization of non-used information buried in routine clinical practice. *IEEE J. Biomed. Health Informat.* 25, 2463–2475. doi: 10.1109/JBHI.2021.3085003

Sharma, N. (2015). Patient-centric approach for clinical trials: current trend and new opportunities. *Perspect. Clin. Res.* 6, 134–138. doi: 10.4103/2229-3485.159936

Sharma, N., and Bhatt, R. (2022). Privacy preserving knowledge graph for healthcare applications. *J. Phys. Conf. Ser.* 2339:012013. doi: 10.1088/1742-6596/2339/1/012013

Shirai, S., Seneviratne, O., and McGuinness, D. (2021). Applying personal knowledge graphs to health. *ArXiv abs/2104.07587*. doi: 10.48550/arXiv.2104.07587

Shortliffe, E. (2012). Computer-based medical consultations: MYCIN, vol. *2*: Elsevier.

Smirnova, A., Audiffren, J., and Mauroux, P. C. (2018). "Distant supervision from knowledge graphs" in Encyclopedia of big data technologies, 1–7.

Solanki, P., Grundy, J., and Hussain, W. (2022). Operationalising ethics in artificial intelligence for healthcare: a framework for ai developers. *AI Ethics* 3, 223–240.

Sun, Z., Dong, W., Shi, J., and Huang, Z. (2020). Interpretable disease prediction based on reinforcement path reasoning over knowledge graphs. *ArXiv abs/2010.08300*. doi: 10.48550/arXiv.2010.08300

Sun, L., Yamin, M., Mushi, C., Liu, K., Alsaigh, M., and Chen, F. (2014). Information analytics for healthcare service discovery. *J. Healthc. Eng.* 5, 457–478. doi: 10.1260/2040-2295.5.4.457

Sun, Z., Yin, H., Chen, H., Chen, T., Cui, L., and Yang, F. (2020). Disease prediction via graph neural networks. *IEEE J. Biomed. Health Inform.* 25, 818–826. doi: 10.1109/JBHI.2020.3004143

Talukder, A., Schriml, L., Ghosh, A., Biswas, R., Chakrabarti, P., and Haas, R. E. (2022). Diseasomics: actionable machine interpretable disease knowledge at the point-of-care. *PLoS Digit. Health* 1:e0000128. doi: 10.1371/journal.pdig.0000128

Taye, M. M. (2010). Understanding semantic web and ontologies: theory and applications. *J. Comput.* 2-6:6.

Tomczak, J. M., and Gonczarek, A. (2012). Decision rules extraction from data stream in the presence of changing context for diabetes treatment. *Knowl. Inf. Syst.* 34, 521–546. doi: 10.1007/s10115-012-0488-7

Torous, J., and Blease, C. (2023). Chatgpt and mental healthcare: balancing benefits with risks of harms. *BMJ Mental Health* 26:e300884. doi: 10.1136/bmjment-2023-300884

Trivedi, R., Sisman, B., Dong, X. L., Faloutsos, C., Ma, J., and Zha, H. (2018). "Linknbed: multi-graph representation learning with entity linkage" in Proceedings of the 56th annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Association for Computational Linguistics), 252–262. doi: 10.18653/v1/P18-1024

Tu, T. K.. A constraint propagation approach for identifying biological pathways in covid-19 knowledge graphs. In 2021 IEEE International Conference on Big Data (Big Data), 2678–2684. Orlando, FL, USA: IEEE (2021). doi: 10.1109/BigData52589.2021.9671290

Uchegbu, C., and Jing, X. (2017). The potential adoption benefits and challenges of LOINC codes in a laboratory department: a case study. *Health Informat. Sci. Syst.* 5:6. doi: 10.1007/s13755-017-0027-8

Wang, S., Chang, X., Li, X., Long, G., Yao, L., and Sheng, Q. Z. (2016). Diagnosis code assignment using sparsity-based disease correlation embedding. *IEEE Trans. Knowl. Data Eng.* 28, 3191–3202. doi: 10.1109/TKDE.2016.2605687

Wang, Q., and Haihong, E. Bi-directional joint embedding of encyclopedic knowledge and original text for Chinese medical named entity recognition. In 2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT), 304–309. IEEE (2021).

Wang, M., Liu, M., Liu, J., Wang, S., Long, G., and Qian, B. (2017). Safe medicine recommendation via medical knowledge graph embedding. *ArXiv abs/1710.05980*. doi: 10.48550/arXiv.1710.05980

Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* 29, 2724–2743. doi: 10.1109/TKDE.2017.2754499

Wang, X., Wang, Y., Gao, C., Lin, K., and Li, Y. (2018). Automatic diagnosis with efficient medical case searching based on evolving graphs. *IEEE Access* 6, 53307–53318. doi: 10.1109/ACCESS.2018.2871769

Wang, H., Wang, A., Su, F., Feng, H., and Chen, Y.. Construction of therapy-disease knowledge graph (tdkg) based on entity relationship extraction. In 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), 851–855. Sanya, China: IEEE (2021). doi: 10.1109/CECIT53797.2021.00060

Wang, Q., Xia, Y., Zhou, Y., Ruan, T., Gao, D., and He, P. (2018). Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition. *J. Biomed. Inform.* 92:103133. doi: 10.1016/j.jbi.2019.103133

Wang, L., Xie, H., Han, W., Yang, X., Shi, L., Dong, J., et al. (2020). Construction of a knowledge graph for diabetes complications from expert-reviewed clinical evidences. *Comput. Assist. Surg.* 25, 29–35. doi: 10.1080/24699322.2020.1850866

Weng, H., Liu, Z., Yan, S., Fan, M., Ou, A., Chen, D., et al. (2017). "A framework for automated knowledge graph construction towards traditional Chinese medicine" in HealthInformation science, vol. *10594* (Springer International Publishing), 170–181. doi: 10.1007/978-3-319-69182-4_18

Wenying, Y., Ding, S., Yue, Z., and Yang, S.. Emotion recognition from facial expressions and contactless heart rate using knowledge graph. In 2020 IEEE International Conference on Knowledge Graph (ICKG), 64–69. Nanjing, China: IEEE (2020). doi: 10.1109/ICBK50248.2020.00019

Wilcke, X., Kleijn, M. D., Boer, V. D., Scholten, H., and Harmelen, F. Bottom-up discovery of context-aware quality constraints for heterogeneous knowledge graphs. *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020) - KDIR.81--92* (2020). doi: 10.5220/0010113500810092

Wilhelm, T. I., Roos, J., and Kaczmarczyk, R. (2023). Large language models for therapy recommendations across 3 clinical specialties: comparative study. *J. Med. Internet Res.* 25:e49324. doi: 10.2196/49324

Wright, A., Pang, J., Feblowitz, J., Maloney, F., Wilcox, A., Ramelson, H., et al. (2011). A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. *J. Am. Med. Informat. Assoc.* 18, 859–867. doi: 10.1136/amiajnl-2011-000121

Wu, J. (2021). Construct a knowledge graph for China coronavirus (covid-19) patient information tracking. *Risk Manag. Healthc. Policy* 14, 4321–4337. doi: 10.2147/RMHP.S309732

Wu, Y., Pan, J., Lu, P., Lin, K., and Yu, X. (2017). Knowledge graph embedding translation based on constraints. *J. Inf. Hiding Multim. Signal Process* 8, 1119–1131.

Xiang, X., Wang, Z., Jia, Y., and Fang, B.. Knowledge graph-based clinical decision support system reasoning: a survey. In 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), 373–380 (2019).

Xie, Y., Yan, C., and Zhang, D.. Personalized diagnostic modal discovery of traditional Chinese medicine knowledge graph. In 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 1096–1103 (2018).

Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., and Achan, K.. Product knowledge graph embedding for e-commerce. Proceedings of the 13th International Conference on Web Search and Data Mining (2020).

Yu, H., Li, H., Mao, D., and Cai, Q. (2020). A relationship extraction method for domain knowledge graph construction. *World Wide Web* 23, 735–753. doi: 10.1007/s11280-019-00765-y

Yuan, J., Jin, Z., Guo, H., Jin, H., Zhang, X., Smith, T. H., et al. (2019). Constructing biomedical domain-specific knowledge graph with minimum supervision. *Knowl. Inf. Syst.* 62, 317–336. doi: 10.1007/s10115-019-01351-4

Yuan, Y., Savage, R., and Markowetz, F. (2011). Patient-specific data fusion defines prognostic cancer subtypes. *PLoS Comput. Biol.* 7:e1002227. doi: 10.1371/journal.pcbi.1002227

Zhang, K.. A knowledge graph based medical intelligent question answering system. In 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), 460–465 (2021).

Zhang, Y., Sheng, M., Zhou, R., Wang, Y., Han, G., Zhang, H., et al. (2020). Hkgb: an inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians' expertise incorporated. *Inf. Process. Manag.* 57:102324. doi: 10.1016/j.ipm.2020.102324

Zhang, Z., Yu, P., Pai, N., Chang, H., Chen, S., Yin, M., et al. (2022). Developing an intuitive graph representation of knowledge for nonpharmacological treatment of psychotic symptoms in dementia. *J. Gerontol. Nurs.* 48, 49–55. doi: 10.3928/00989134-20220308-02

Zhao, Q., Li, J., Xu, C., Yang, J., and Zhao, L. (2020). Knowledge-enhanced relation extraction for Chinese EMRS. *IT Professional* 22, 57–62. doi: 10.1109/MITP.2020.2984598

Zheng, S., Zhu, Z., Liu, Z., Guo, Z., Liu, Y., and Zhao, Y. (2021). Multi-modal graph learning for disease prediction. *IEEE Trans. Med. Imaging* 41, 2207–2216. doi: 10.1109/TMI.2022.3159264

Zhou, B., Cai, X., Zhang, Y., Wu, G., and Yuan, X. (2021). Mtaal: multi-task adversarial active learning for medical named entity recognition and normalization. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 14586–14593. doi: 10.1609/aaai.v35i16.17714

Zhu, Q., Qu, C., Liu, R., Vatas, G., Clough, A., Nguyen, D., et al. (2022). Rare disease-based scientific annotation knowledge graph. *Front. Artif. Intell.* 5:932665. doi: 10.3389/frai.2022.932665