BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 07/10 pp201–217
Volume 6, Number 2, June 2023
DOI: 10.26599/BDMA.2022.9020021

Medical Knowledge Graph: Data Sources, Construction, Reasoning, and Applications

Xuehong Wu†, Junwen Duan†, Yi Pan, and Min Li*

Abstract: Medical knowledge graphs (MKGs) are the basis for intelligent health care, and they have been in use in a variety of intelligent medical applications. Thus, understanding the research and application development of MKGs will be crucial for future relevant research in the biomedical field. To this end, we offer an in-depth review of MKG in this work. Our research begins with the examination of four types of medical information sources, knowledge graph creation methodologies, and six major themes for MKG development. Furthermore, three popular models of reasoning from the viewpoint of knowledge reasoning are discussed. A reasoning implementation path (RIP) is proposed as a means of expressing the reasoning procedures for MKG. In addition, we explore intelligent medical applications based on RIP and MKG and classify them into nine major types. Finally, we summarize the current state of MKG research based on more than 130 publications and future challenges and opportunities.

Key words: medical knowledge graph; knowledge graph construction; knowledge reasoning; intelligent medical applications; intelligent healthcare

1 Introduction

A knowledge graph (KG)[1,2] is a semantic network composed of entities and their relations in the real world. KGs represent one of the benchmarks in artificial intelligence research and offer an ideal way to integrate heterogeneous data resources and enhance knowledge-based applications. In particular, medical KGs (MKGs) attract the attention of academics and the healthcare industry for their potential in intelligent healthcare applications[3–10].

MKG research has achieved significant advancements in most areas. MKGs are constantly built from the collection and extraction of structured knowledge from unstructured or semi-structured heterogeneous medical information resources, such as electronic medical record (EMR)[11], electronic health record (EHR)[12], clinical trials[13] and other clinical data[14,15], medical literature, textbooks, internet medical resources, shared standard medical terminology, and open-and-shared medical knowledge bases[16–20]. MKG construction is steadily progressing from manual[21] to semi-automatic[22] to automatic construction[23]. Furthermore, an increasing number of frameworks and platforms for KG creation have been proposed[21,24,25]. The construction of an MKG for all fields is very challenging. Therefore, the main research is currently focused on the direction led by particular application scenarios[26–30].

MKGs provide computer systems with the cognitive capacity necessary to support all types of intelligent applications by combining techniques such as knowledge

• Xuehong Wu is with School of Computer Science and Engineering, Central South University, Changsha 410083, China and School of Computer Science, Hunan First Normal University, Changsha 410006, China. E-mail: wuxuehong214@163.com.
• Junwen Duan and Min Li are with School of Computer Science and Engineering, Central South University, Changsha 410083, China. E-mail: jwduan@csu.edu.cn; limin@mail.csu.edu.cn.
• Yi Pan is with Faculty of Computer Science and Control Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518000, China. E-mail: yipan@gsu.edu.
† Xuehong Wu and Junwen Duan contribute equally to this work.
* To whom correspondence should be addressed.
Manuscript received: 2022-07-03; accepted: 2022-07-13

representation, logic rules, machine learning, and deep learning. MKGs have also been applied in various intelligent medical scenarios, including intelligent and semantic medical knowledge retrieval[31, 32], auxiliary diagnosis of diseases[33, 34], clinical education[35, 36], drug analysis[37, 38], diagnosis and treatment plan recommendation[39, 40], intelligent question answer and chatbots[41, 42], intelligent nursing[43], and smart health management[44–46].

Given their significance in research and a wide range of applications, we aim to provide a comprehensive description of MKGs. We begin by reviewing the creation mechanism of MKGs, including an overview of medical information sources, construction techniques, and themes of MKGs. Second, we summarize and introduce three reasoning models for the implementation of MKG-based reasoning and provide a reasoning implementation path (RIP) to define the reasoning procedures. Afterward, we describe applications based on MKGs and categorize findings based on application scenarios. Finally, to address the growing need for MKGs in intelligent applications, we discuss the present challenges and opportunities to stimulate further research in this field.

2 Construction of MKG

2.1 Data sources of MKG

Data resources are crucial for the development of a trustworthy MKG. The vast amount of accessible medical data enables the construction of a large-scale MKG with rich and dependable medical entities and relations. We classify data sources utilized in related works into four groups, namely, real-world data, scientific publications, standard libraries, and open-and-shared medical knowledge databases.

Real-world data[47] mainly refer to the data from clinical diagnosis and treatment processes, such as EMR, EHR, clinical trials, and other clinical data. These data sources provide essential clinical knowledge and are thus tapped early on. For example, Wang et al.[48] created a large and high-quality heterogeneous graph connecting patients, diseases, and drugs (PDD) in EMRs, and Malik et al.[23] created an MKG from the electronic health records of 1025 patients with intracranial aneurysms.

Scientific publications, such as literature, textbooks, and guidelines, are published by authoritative institutions, publishers, and researchers. As these data sources are significantly trustworthy and widely available, they have been used to construct huge MKGs or MKGs for specific diseases. For example, Zhang and Che[49] collected Parkinson's disease-related connections from medical literature and constructed a medical literature KG. Sun et al.[45] collected 185 796 drug labels from the China Food and Drug Administration, 3390 types of disease information from medical textbooks, and information from 5272 examinations as knowledge sources. Ernst et al.[50] introduced a well-configured KG, KnowLife, which is a harvested text from a variety of genres, including scientific journals, health portals, and online communities.

Standard libraries are a kind of standard and shared medical terminology and standards, which include the Medical Subject Headings (MeSH)[16], Unified Medical Language System (UMLS)[18], International Classification of Diseases (ICD), SNOMED-CT[17], and so on; Zhang et al.[29] created an obstetric KG based on the hierarchical structure of MeSH[16] as the ontology prototype. Patil et al.[51] developed a concept graph engine (CG-Engine) and used the UMLS database as its medical knowledge base.

Open-and-shared medical knowledge databases, such as FreeBase[52], RepoDB[53], DurgBank[19], SemMedDB[54], etc., refer to a collection of open and freely accessible medical knowledge created by researchers. Teng et al.[55] constructed a KG with five entity classes (i.e., disease, symptom, medicine, surgery, and examination) by extracting entities related to ICD-9 from Freebase. Malas et al.[56] used the semantic information between drugs and diseases from the existing KG RepoDB[53], which is a standard drug repurposing database. Korn et al.[27] built COVID-KOP, a new knowledge base combining the ROBOKOP[57] biomedical KG with information from contemporary biomedical literature on COVID-19 annotated in the CORD-19 collection.

Several of the data sources mentioned above, such as the standard library and shared medical knowledge databases, can be utilized directly to build MKGs. However, some of them, such as real-world data and scientific articles, require additional structural and semantic examination. The proper transformation and integration of medical information into MKG are more critical than the acquisition of additional medical knowledge. Table 1 provides a comprehensive summary and analysis of medical knowledge sources.

2.2 Methods for constructing MKGs

The most frequently used strategy for MKG creation

Table 1 Data sources of MKG and related research.

Source type	Source	Related research
Real-word data	EMR	[21, 30, 47, 48, 55, 58–71]
	EHR	[23, 25, 72–76]
	Chinese clinical dataset Levis hypertension	[77]
	Cancer registry data	[15]
	Clinically pediatric cases	[33]
	Clinical trial reports	[13, 78, 79]
	NHANES data set	[14]
Scientific publications	Literature	[4, 10, 13, 25, 27, 37, 46, 49, 50, 58, 78–91]
	Textbooks	[4, 29, 33, 45]
	Internet resource	[6, 21, 29, 41, 45, 50, 79, 85, 87, 92, 93]
	Guidelines	[13, 43, 45, 78, 79]
Standard library	Mesh	[16, 29, 46, 50, 58, 79, 85, 90, 94–96]
	UMLs	[7,18,31,46,49–51,58,59,64,71,79,81–83,86,94–99]
	ICD9/10	[45, 46, 48, 55, 59, 66, 67, 70–72, 74, 85]
	SNOMED-CT	[13, 17, 70, 71, 74, 79, 85, 100, 101]
Open and shared medical knowledge database	FreeBase	[52, 55]
	SemMedDB	[7, 54, 86, 87, 98, 99]
	RepoDB	[53, 56]
	ROBOKOP	[27, 57]
	DrugBank	[13, 19, 48, 49, 66, 67, 79, 102–105]
	KEGG	[20, 49, 95, 102, 105]
	Google health knowledge graph	[59]
	SIDER	[13, 49, 79, 103, 106]
	Cancer/tumor/case dataset	[15]
	InterPro	[102, 107]
	UniProt	[87, 102, 104, 108]
	Gene Ontology	[18, 109]
	OpenKG.CN	[26, 104]
	Linked open data	[110]
	Linked life data	[110]
	Therapeutic target database	[88, 104]
	BioGRID	[104]
	DBpedia/CN-DBpedia	[34, 97, 111]

is the extraction of entities and relations from structured and unstructured resources. Nevertheless, approaches such as medical concept normalization[112–114], knowledge fusion[115–117], knowledge completion[99, 118, 119], and complex knowledge representation[3, 120, 121] are crucial to ensure the completeness and quality of MKG.

2.2.1 Entity and relation extraction

Earlier studies relied on expert knowledge to generate feature sets for entity and relation extraction. Zhao et al.[30] manually collected medical entities and their modifiers from two EMR datasets. These entities included but were not limited to the patient's basic information, primary complaints, tests, test findings, diagnosis, and treatment plans. Song et al.[85] constructed a pediatric KG using manual annotation, knowledge fusion, and other technologies; they expanded the triplet form of knowledge to a sextuplet form. Cheng et al.[60] used data mining to mine medical laws, transformed them into medical knowledge with the assistance of specialists and then constructed the KG appropriately. Zhang et al.[29] used a combination of bootstrapping and support vector machine (SVM) methods to extract relations between entities to build an obstetric KG. Sang et al.[80] presented SemaTyP, a technique for drug discovery based on biomedical KGs. It trains a logistic regression model by learning the semantic types of pathways of known medication treatments in the KG and then uses the model to identify novel disease-specific pharmacological therapies. Rotmensch et al.[59] employed maximum likelihood estimation to automatically generate KGs using three probabilistic models: logistic regression, a naive Bayes classifier, and a Bayesian network with nosy OR gates. A recent study focused on automated entity and relation extraction using deep learning techniques[76]. Sun et al.[45] extracted entities and relations from knowledge sources and then linked them using a multilevel similarity matching strategy to ensure the MKG's quality.

2.2.2 Medical concept normalization (MCN)

MCN aims to map informal medical terms to formal medical concepts, which is crucial in ensuring an MKG's quality. Li et al.[112] studied the efficacy of BERT-based models in the biological and clinical domains for entity normalization. They demonstrated that the methods based on BERT outperformed many state-of-the-art techniques. Li et al.[47] defined nine types of entities and utilized ICD-10 as the de facto standard for disease terminology and ICD-9 for surgical terminology. They also standardized words for other medical concepts. Yue et al.[122] developed a disease-centric and physician-guided annotation method and specification for named entities and relations. Luo et al.[114] developed a new manually annotated large-coverage corpus for clinical concept normalization, and the MCN corpus was shared with the scientific community as part of a collaborative effort. Pattisapu et al.[123] trained MCN models using automatically labeled instances retrieved from patient discussion forums and then utilized pretrained sentence encoding models to determine the k-nearest words for each medical topic.

2.2.3 Knowledge graph completion

KG completion is the process of predicting new or missing facts based on existing facts, ensuring that the KG is complete. Yin et al.[124] created and constructed a diabetic KG from electronic medical information, and they suggested a paradigm for KG completion via translation. Biswas et al.[125] completed and predicted edges in a KG using the ComplEx embedding approach, which considered all binary relational features (reflexive, symmetric, and transitive) in the graph. Zhang et al.[99] predicted drug-repurposing candidates for COVID-19 using five state-of-the-art neural KG completion algorithms (i.e., translating embedding (TransE), RotatE, DistMult, ComplEx, and STELP). Moreover, Zhang and Che[49] used five KG completion approaches to identify treatment candidates for Parkinson's disease: DistMult and ComplEx for semantic matching models, ConvE and ConvTransE for neural network models, and TransE for translational distance models.

2.2.4 Knowledge fusion

Knowledge fusion is used to increase the number of entities in the final KG and to assure their objective uniqueness. MKGs combine knowledge from disparate data sources, whereas knowledge fusion approaches combine descriptive data of approximately the same entity or concept from numerous sources. Zhang et al.[29] collected obstetric disease characteristics using heterogeneous data from medical specifications, classic textbooks, and medical internet websites and then integrated the information using the Simhash–TF-IDF method. Gong et al.[66] created a very heterogeneous graph by creating patient–disease and patient–medicine bipartite graphs using EMRs and connecting them to ICD-9 and DrugBank MKGs. Zhang and Che[49] ensured that the same entity had the same name by using the UMLS identification rather than a specific entity in the local medical knowledge base. Li et al.[98] fused the KG in two stages: entity mapping, which utilizes the standard name as the entity name, and entity alignment, which calculates the similarity of entity names using the Jaccard similarity algorithm[126]. Fang et al.[71] suggested a head-and-tail entity fusion model, which obtained 97% accuracy while fusing data from diverse sources. Yan et al.[104] created a COVID-19 KG by integrating 14 publicly available bioinformatic databases comprising information on medications, genes, proteins, viruses, illnesses, and symptoms and their associations. They utilized the DrugBank ID for each drug, the National Center for Biotechnology Information gene ID for each gene, and the MeSH ID for each disease because they are all standardized. Chen et al.[117] introduced the MUFFIN multiscale feature fusion deep-learning model for learning drug representation using drug-self structural information and KGs with rich biomedical knowledge.

2.2.5 KG tools

Research has been consistently focused on the development of scalable, adaptable, automated, and easy-to-use domain-based tools or frameworks. Duan et al.[43] and Zhang et al.[29] constructed and visualized the entity graph using protégé[24]. Xie et al.[21] developed an incremental expansion approach for constructing expandable MKGs based on an EMR. Their architecture enables the integration of external knowledge gleaned from medical information websites and the mining of prospective knowledge associated with new EMRs.

Alobaidi et al.[25] suggested a novel framework for automated ontology generation comprising five primary modules: (1) text processing via compute-on-demand; (2) medical semantic annotation via n-gram, ontology linking, and classification algorithms; (3) relation extraction via graph method and syntactic patterns; (4) semantic enrichment via rdf mining; (5) a domain inference engine that generates the formal ontology. Malik et al.[23] suggested a system for domain-specific automated knowledge curation; this system enables the extraction of terms, relations, individual and cohort graphs, and predictive information. The system has an accuracy of 78% and a recall of 71%. Weng et al.[77] investigated an automated MKG creation framework based on semantic analysis. The framework consists of a medical ontology constructor, a knowledge component generator, a constructed knowledge dataset generator, and a graph model (GM) constructor, all of which significantly enhance accuracy. DEKGB[61] is also an expandable framework for MKG; it was used to generate KGs for specific diseases using pre-existing medical information and doctor-involved electronic medical records. Doctors may easily and quickly develop highly specialized health KGs with the assistance of DEKGB. Other additional and similar frameworks have been observed [22,47,50,127–129].

According to the studies above, the KG building technology has advanced significantly in the biomedical field, and with the growth of big data, natural language processing, and deep learning technology, MKG construction techniques have also expanded. A certain number of construction tools are emerging to assist researchers in performing similar and repeated labor-intensive tasks.

2.3 Constructed MKGs

A variety of subject-oriented MKGs have been developed for specific objectives using a range of medical resources and MKG creation methodologies. In this article, we gather and evaluate the created MKGs and classify them based on their subjects (Table 2).

We classify the MKGs developed by academics in recent years into six distinct categories based on their domain and application scenarios (Table 2). (1) Integration-oriented MKGs incorporate information from all medical fields and encompass a large number of fundamental concepts; they are often expandable. (2) Disease-oriented MKGs use specific diseases as core concepts and include medical facts, such as disease–drug and disease–symptom relations. (3) Drug-oriented MKGs use specific drugs or drug analysis application scenarios to generate drug KGs. (4) Department-based MKGs are formed from department-specific knowledge, such as disease, symptoms, examinations, and tests from certain departments, such as obstetrics and pediatrics. (5) Biomedical-oriented MKGs are constructed from biological and medical knowledge. (6) Other MKGs comprise a collection of MKGs that support a variety of application services.

As shown in Table 2, we additionally gather information on the scale of MKGs using four indicators: the number of entity types, relation types, entities, and facts. The researchers refer to the four indicators using different names. The term “entity types” refers to a variety of medical concepts, including diseases, symptoms, medicine, surgery, therapies, genes, and tests. Xiu et al.[130] used the term “class” to refer to entity classes; “class” is a meta term from the ontology language (Web Ontology Language) and refers to the concept of objective existence. A “fact” is a fundamental unit of information expressed in KGs, and it is often described using the RDF framework, which is also known as a triplet. However, Li et al.[47] represented medical knowledge using a quadruplet structure rather than a triplet structure, and Song et al.[85] expanded the triplet form of a KG to a sextuplet form. Varied terminologies have different meanings in various description situations. Ernst et al.[50] used the term “relations” to refer to relation types, but Zhu et al.[94] and Zhang et al.[29] applied the term “relations” to refer to medical facts. Despite their distinct phrasing, their objectives were the same. In general, the numbers of entities, relations, and entity and relation types are highly correlated with the expression capability of MKGs. Table 2 also includes the scale of MKGs, which should inspire the construction and fusion of MKGs for follow-up studies.

3 Reasoning over MKG

The reasoning over MKG allows the discovery of new prospective knowledge and relevance prediction of existing knowledge entities, which are critical for a wide range of intelligent applications. To facilitate the description and comparison of various reasoning approaches, we propose RIP, a novel hierarchical-structured framework (Fig.1). The RIP can be expressed as below.

Table 2 Subjects of MKGs from studies.

Subjects	MKG	Scale of MKGs			
		# entity types	# relation types	# entities	# facts
Integrative oriented	Knowlife KG[50]	7	13	50 000	609 322
	MKG[47]	9	-	22 508	579 094
	HG[31]	–	14	400 000	200 000
	SMR KG[66]	4	4	367 201	1 707 609
Disease oriented	CLKG[127]	–	–	200 000	1 000 000
	Geriatric KG[26]	6	7	–	–
	Depression KG[79]	–	–	–	8 892 722
	Knee osteoarthritis KG[65]	8	10	2518	29 972
	SemKG[92]	5	4	–	–
	DSTKG[130]	7	16	9868	11 005
	Cancer KG[15]	–	–	–	90 673 527
	KGHC[98]	10	22	5028	13 296
	Rare disease KG[94]	10	42-	3 819 623	84 223 681
	COVID-KOP[27]	–	–	45 300	5 532 000
	DRKF KG[49]	–	43	12 497	165 901
	LT-D DB[51]	3	11	–	–
	KDKG [13]	–	–	–	10 146 311
	StrokeKG[91]	9	30	46 000	157 000
	KGPA[71]	11	10	–	–
Drug oriented	WATRIMed KG[131]	472	75	–	–
	Drug KG[6]	4	3	5828	70 382
	ADRs KG[28]	4	3	12 473	154 239
	KEGG_MED[102]	2	1	5229	12 112
	TBKG[37]	4	6	–	–
	MCKG[96]	2	6	8014	123 890
	TCM-KG[60] [4]	127	58	10 000	1 000 000
Department oriented	Obstetric-KG[29]	4	22	625	3863
	PMKG[85]	–	–	22 023	23 434
Biomedical oriented	PharmKG[132]	3	29	–	50 000
	Bio KG[81]	3	3	6827	16 912
	ROBOKOP KG[57]	43	157	9 399 969	254 940 828
	Integrated KG[82]	–	171	3 527 423	68 413 238
	BioKGLM[87]	5	11	502 100	96 500 000
Others	EMKN[30]	5	3	–	–
	EBDPKG [72]	–	–	1989	10 380
	EMR-based KG[58]	4	6	634 000	14 000 000 000
	SHKG[74]	8	3	28 518	5591
	Symptom KG[100]	3	2	5080	1521
	PDD[48]	3	3	58 030	4 244 856
	DeepPS KG[67]	7	7	42 613	1 888 950
	G_Coder KG[55]	6	9	1560	20 000
	FWA-KG[45]	7	–	1 616 549	5 963 444

$$[\text{RM}]: [\text{ALG}_1] \rightarrow [\text{ALG}_2] \rightarrow [\text{ALG}_n] : [\text{TARGET}]$$

where "RM" denotes the reasoning model, "ALG_n" refers to the n-step algorithms or processing method of the path, and "TARGET" indicates the objective of the reasoning methods, which can be "retrieval", "entity prediction", "relation prediction", "ranking", etc. For example, Chai[92] embedded the KG and transformed each element in it into a vector representation using TransE; then, they used stochastic gradient descent (SGD) to obtain the final graph embedding representation. Next, they predicted the connection between pathology and disease using dual

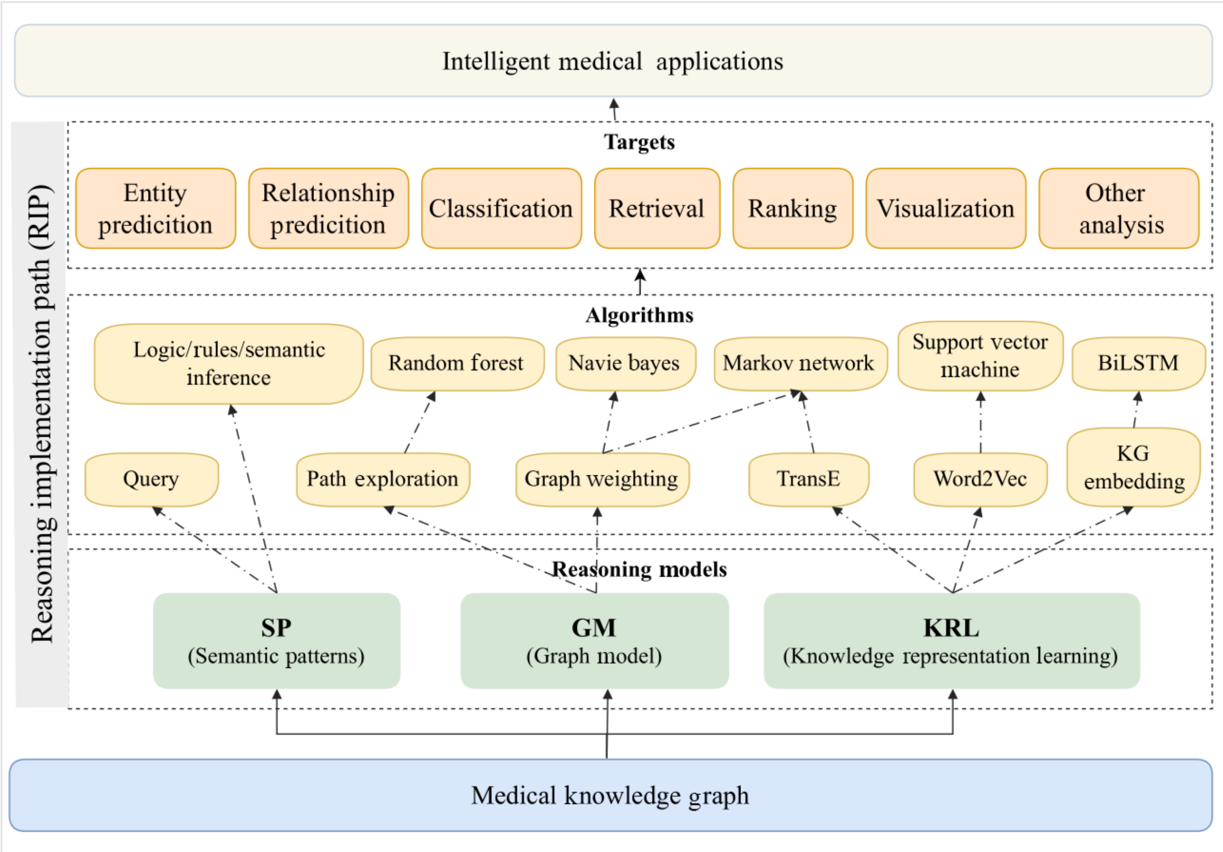

Fig. 1 Reasoning framework based on MKG.

BiLSTM. We may characterize its RIP as follows based on the foregoing explanation.

KRL: TransE → SGD → BiLSTM: relation prediction

We classify reasoning models into three broad categories: reasoning based on semantic patterns (SP), reasoning based on graph model (GM), and reasoning based on knowledge representation learning (KRL). The following sections are used to examine the intricacies of each reasoning model and outline a partial implementation path for each model (Fig. 1).

3.1 Reasoning based on SP

The KG is modeled as a semantic relation network. The links between symbols are its core. We can perform KG utilization analysis using semantic relations and semantic rules. Hasan et al.[15] constructed a KG prototype using the Louisiana Tumor Registry dataset. The dataset was utilized to efficiently query and explore data from a variety of perspectives (surgery, chemotherapy, radiation therapy, and hormone therapy). As a result, the query runtime performance increased significantly by up to 76%. Liu et al.[33] structured the expertise of domain experts as (If…Then…) rules in a hybrid KG and proposed two types of rules to check the available disease labels, namely, the sufficient and required condition rules. Then, they employed medical book knowledge to detect disease labels in predicted label sets, where the differential diagnosis tag captures the conditions for a disease's ultimate diagnosis. Bakal and Kavuluru[86] predicted therapy relations using semantic graph patterns on biological knowledge networks. Shi et al.[74] investigated the complicated semantics between objects by utilizing the linkages between medical terminology and chain inference techniques. They performed reasoning on semantic health knowledge graph (SHKG) using first-order predicate logic, followed by forward and reverse chaining over the KG. Papageorgiou et al.[3] created EYE, a general-purpose reasoning engine, to tackle the challenge of formalizing medical information for decision support. EYE takes advantage of probabilistic and fuzzy impact processes embedded in the semantic web. Malik et al.[23] suggested predictive criteria for subarachnoid hemorrhage prediction using a healthcare domain-specific knowledge network. Sun et al.[45] created the rules for identifying suspected claims using

MKG reasoning.

SP-based reasoning is popular in early research. On the premise of accurate rules, this method has high reasoning accuracy and strong interpretability. However, with the gradual improvement of the scale and complexity of MKG, the efficiency and complexity of reasoning based on this pattern have also gradually improved, which makes this method less computationally efficient. In addition, when noise exists in data, misleading reasoning easily occurs.

3.2 Reasoning based on GM

The KG is a way of representing and organizing graph-based knowledge. Each vertex represents an entity in this specific graph, and each edge indicates the direct link between two entities. Given these properties, graph theory-based algorithms may be used effortlessly in knowledge reasoning[133]. Patil et al.[51] developed a CG-Engine that treats MKGs as graphs. The graph's edges are weighted, and the disease's risk value is calculated by computing the comprehensive weight value of the edge connected to each disease node. Finally, the diseases are classified based on their risk value. Liu et al.[93] created an MKG as a collection of vertices and edges and independently calculated the distance between two vertices and the weights of entity characteristics. To enhance classification performance, we adjust the noise labels in training examples using a combination of weight modification and polishing. Goodwin and Harabagiu[58] modeled the MKG as a factorized Markov network, a probabilistic graphical model that enabled them to compute the probability distribution across all possible clinical scenarios and treatments for patients. Bean et al.[6] employed a matrix to represent the scaled features for each drug node in the drug KG and a binary classifier to produce a score for each drug of each predictor type.

GM-based reasoning also has the problem of the rapid increases in computational complexity with the scale growth of KGs. In addition, multiple semantic relations, which generally have directionality, may exist between two entity KGs. In this case, using graph expressions to support the semantics of a KG is often difficult, resulting in the loss of accuracy during reasoning.

3.3 Reasoning based on KRL

Although symbolic representation enables quantitative reasoning based on statistical probability, its inclusion in machine learning models that execute numerical operations is challenging. KRL aims to convert objects of interest (entities and relations in KG) into a continuous low-dimensional vector space[75, 134] to efficiently measure the semantic correlations between entities and relations and to significantly improve knowledge acquisition, fusion, and inference performance. The KRL model of TransE has demonstrated remarkable outcomes in KG reasoning research. Zhao et al.[30] identified four distinct types of medical entities from records and built a medical knowledge network based on EMRs (EMKN). They developed KRL methods to capture a certain degree of similarity between entities by embedding them in a low-dimensional dense vector space using the latent factor models (LFMs) and TransE models. Chai[92] employed the TransE to embed each element into a vector representation and the SGD to obtain the final graph embedding representation. Then, they predicted the connection between pathology and diseases using dual BiLSTM. Finally, the data associated with recognized pathological diseases were utilized for training the BiLSTM-based illness diagnostic model. Li et al.[47] employed PrTransH to learn embedding vectors from the generated quadruplet-based MKG because it can embed the probability of a single fact into the embedding vectors. Finally, the graph embedding technique was used on a neural network challenge for disease-specific prescription prediction. Biswas et al.[125] used a technique known as tensor factorization. Dai et al.[8] established a new framework for KG embedding by incorporating adversarial autoencoders (AAEs) for drug-drug interaction (DDI) tasks based on Wasserstein distances and the Gumbel-Softmax. They added AAEs to KG representation learning and used Gumbel-Softmax and Wasserstein distance to tackle the problem of vanishing gradients on discrete data.

KRL-based reasoning is a popular method at present. This method can effectively transform the entities and relations in a KG into multi-dimensional vectors, which is convenient for computer calculation and significantly improves reasoning performance. Neural networks and deep learning algorithms can be applied effectively to learn to express objects in a KG. However, this method lacks deep expression capability and interpretability for the semantics of KGs.

However, combining various reasoning models is also a successful strategy; Yan et al.[104] employed motif-based graph analysis (GM-based) and KG embedding (KRL-based) to compute the scores for candidate drugs independently and then combine them using a linear function.

4 Intelligent Applications of MKG

MKG research and application have successfully demonstrated that these graphs can provide a wide variety of decision analysis services and applications for intelligent healthcare. The application cases include intelligent medical information retrieval, disease diagnosis, intelligent clinical education, drug analysis, intelligent question and answering and chatbots, intelligent nursing, and health management. In Table 3, we assess and categorize the major applications of MKGs.

As shown in Table 3, we categorize and highlight recent achievements in MKGs based on application scenarios. In addition, for each application, the accompanying MKG and RIP information are included to aid in comprehension of the implementation techniques. The "-" in the third column indicates that the associated MKG was not stated nor discussed in the research. Notably, some researchers did not name their developed MKGs specifically. Thus, we name them after their application scenario or system framework, followed by a "KG" suffix, for example, "FWA KG," "DeepPS KG," or "DSQA KG."

As the data indicate, disease diagnosis and drug analysis are research hotspots. The use of disease diagnosis is mostly focused on the identification of particular diseases and offering prediction results for them. For example, the applications based on MKGs can provide diagnostic services for pediatric[33] and geriatric[26] diseases. Certain prevalent specialist diseases, such as type 2 diabetes, thyroid disease, subarachnoid hemorrhage, and sepsis, have garnered significant study attention. Diagnoses of uncommon and common disorders are equally critical for clinical outcomes. The primary study directions in drug analysis include the prediction of adverse drug reactions, prediction of drug interactions, novel drug development, drug reuse, medication recommendation, and drug safety. The development of drugs for COVID-19 is also an important research topic [104, 143, 144]. The RIPs in Table 3 demonstrate the application mechanisms of MKGs for reasoning, utilizing both innovative (HKDP, MedSim, KGETM, TriModel, etc.) and traditional (SVM, Word2Vec, logistic regression, TransE, CNN, Bayesian, etc.) approaches for the three reasoning models.

5 Discussion

In 2012, Google introduced the notion of a KG to strengthen its search engine and other applications. The core concept is to use ontologies to model entities and relations in the actual world to assist machines in intelligently comprehending them. We gather over 130 publications on KGs in the biomedical sector from Web of Science, PubMed, Elsevier/ScienceDirect, IEEE/IET Electronic Library, SpringerLink, and many others. Fig. 2 illustrates the research trends and directions for MKG based on literature analysis. As illustrated in Fig. 2a, scholars have embraced and implemented the notion of KGs in the medical field since its introduction. The amount of related literature produced annually is rising, and the idea of MKG is well recognized and continually evolves as its usefulness is steadily demonstrated via real-world applications. In Fig. 2b, we investigate and statistically analyze three major research directions, namely, (1) MKG construction, which focuses on the source and benchmark of medical knowledge, and construction techniques, systems, and tools; (2) reasoning techniques, which primarily consist of a range of reasoning methods based on KGs and deep learning, big data, machine learning, graph theory algorithms, logic and rule inference, and so on; (3) intelligent applications, which focus on a variety of smart application scenarios, including disease detection, drug analysis, cdss, and health management. A recent study statistically indicated the three directions as research hotspots.

6 Challenge and Outlook

Although MKGs have made significant advancements, the constant expansion of medical data and the rising need for intelligence have introduced a number of new challenges.

(1) Large-scale heterogeneous medical data[145]. Fusing information from disparate data sources into a shared, actual, and unified medical ontology is an ongoing research area.

(2) Medical domain complexity. Given the complicated medical information, conveying objective medical facts using specific SPs, particularly the sequential MKG, which is critical for clinical diagnosis and therapy, is difficult. However, a limited number of studies have been conducted on this subject.

(3) Accurate KRL. KRL is commonly used to describe the semantic properties of MKGs, but it is insufficiently accurate, particularly for complicated KGs with numerous relations, attributes, and entity kinds.

(4) Diversification of MKGs. Numerous MKGs have

Table 3 Intelligent applications based on MKG and RIP.

Scenario	Application	Related MKG	Related RIP
Intelligent retrieval	COVID-19 related literature retrieval[32]	CORD-19 NEKG, CORD-19 AKG	SP: query: retrieve
	TCM retrieval[4]	TCM KG	SP: query: retrieve
	Medical text retrieval[31]	HG	KRL: KGE: retrieve
Disease diagnosis	Pediatric disease prediction[33]	Hybrid-KG	GM: HKDP: classification
	Geriatric disease reasoning[26]	Geriatric KG	SP: logic&rule inference: relation prediction
	Rare disease diagnosis[34]	–	SP: entity feature→ SVM: classification
	Thyroid disease diagnosis[92]	SemKG	KRL: KGE→ BLSTM: Classification
	Predicate type 2 diabetes (T2D)[63]	T2D KG	GM: logistic regression→ graph weighting: relation prediction & visualization
	Predicate the severity of sepsis[135]	–	KRL: self-attentation → BiLSTM → KGE → attentation: classification
	Identify potential migraine biomarkers[82]	–	SP: extract compounds → filter&rank: ranking
	Multi-disease diagnosis[136]	CEMRs KG	KRL: RNKN: classification
	Decision support for UTI[3]	–	SP: EYE(BBNs&FCMs): ranking
	Provide personalized disease ranking[51]	LT-D DB	GM: CG-Engine: ranking
	Predicate subarachnoid hemorrhage[23]	–	SP: rule&logic: relation prediction
Clinical education	Predict outcomes to questions on the SIDES platform[35]	OntoSIDES	KRL: KGE: ranking
	Disease diagnosis in a medical training system[36]	–	KRL: KGE: classification
Drug analysis	Prediction of unknown ADRs[6]	ADRs	GM: graph weighting → machine learning algorithm: classification
	ADR discovery[37]	TBKG	GM: graph weighting → Navie Bayes: relation prediction
	Prediction of ADRs[28]	ADRs KG	KRL: Word2Vec → logistic regression: classification
	Therapeutic substitution of antibiotics[103]	DrugBank	KRL: MedSim: similarity analysis
	Combined drug therapies[78]	–	SP: rules → filter → automated algorithm: relation prediction
	Herb recommendation[137]	TCM KG	KRL: KGETM&HC-KGETM: entity prediction
	Prediction of DDIs[38]	DrugBank	GM: similarity measures → logistic regression: relation prediction
	Drug-drug interaction prediction[8]	DeepDDI, Decagon	KRL: AAEs: relation prediction & classification
	Analysis of neglected influencing factors of statin-induced myopathy[110]	LOD, LLD	SP: SPARQL query: other analysis
	Drug discovery[80]	SemKG	GM: path exploration → logistic regression: ranking
	Drug efficacy screening[138]	Guney, EMC	GM: path exploration → random forest: classification
	Drug repurposing[139]	DTINet	KRL: DDTE: entity prediction
	Drug repurposing[56]	RepoDB	GM: extract paths → random forest: classification
	Drug repurposing against Parkinson's disease[49]	DRKF KG	KRL: KGE → SVM: classification
	Drug repurposing for COVID-19[99]	SemMedDB	KRL: KGE: entity prediction
	Drug-drug similarity(DDS)[140]	KGDDS KG	KRL: KGE → similarity compute: similarity analysis
	Safe medicine recommendation[66]	SMR KG	KRL: KGE: ranking
	Prediction of drug target proteins[102]	KGEE_MED	KRL: KGE (TriModel)→relation prediction

(to be continued)

Table 4 Intelligent applications based on MKG and RIP. **(continued)**

Scenario	Application	Related MKG	Related RIP
CDSS	A data-intensive CDSS platform[39]	IDS KG	SP: rule inference: other analysis
	Intelligent diagnose assistant system[74]	SHKG	SP: semantic inference: other analysis
	Medical aided diagnosis system[40]	–	KRL: → KGE (TransR) → LSTM: other analysis
Medical Q&A and chatbot	Medical question answering[58]	–	SP: probabilistic inference: ranking
	An online medical chatbot system[41]	–	SP: extract entity → query: retrieve
	A QA system for smart health[64]	DSQA KG	SP: generate QA pairs → query: retrieve
	Medical question answering system[42]	–	SP: query: retrieve
Intelligent nursing	Eldercare[43]	–	SP: query: other analysis
	Balance reactive care and proactive care[72]	EBDPKG	SP: Bayesian: classification
Health management	Dietary supplements (DSs)[44]	iDISK	SP: Web App → query: visualization& KBQA
	Health risk prediction[14]	HKG	GM: optimisation algorithm: classification
	Predicting the status of health risks[46]	KB-HIG	KRL: graph weighting → Word2Vec: classification
	Estimating personalized risk ranking[51]	LT-D DB	KRL: CG-Engine: ranking
Others	Fraud, waste, and abuse (FWA) detection[45]	FWA KG	SP: rule inference: relation prediction
	Deep patient similarity[67]	DeepPS KG	KRL: KGE → CNN: similarity
	Automated ICD coding[55]	G_Coder KG	KRL: SDNE → attentation → FGM: classification
	Semantic enhancement[11, 87, 141, 142]	–	–

been synthesized. However, their description framework, ontology model, medical terminologies, semantic identities, and storage technologies are all distinct, which results in extremely difficult further reuse, linking, and sharing.

(5) Graph of common sense knowledge. MKGs should consider not only medical knowledge but also common sense. Knowledge of common sense is critical because it serves as the foundation for cognitive competence.

The advancement of KG approaches in the medical industry has resulted in the creation of an increasing number of application scenarios and intelligent services. Students, doctors, patients, clinical administrators, and researchers in the medical profession will progressively profit from the application scenarios offered. We hope that MKG research will result in the development of a human-inspired artificial intelligence system capable of integrating generic, common-sense, and domain-specific information with societal values and norms and individual cognitive models.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (No. 2021YFF1201200), the National Natural Science Foundation of China (No. 62006251), and the Science and Technology Innovation Program of Hunan Province (No. 2021RC4008).

References

[1] A. Singhal, Introducing the knowledge graph: Things, not strings, https://blog.google/products/search/introducing-knowledge-graph-things-not/, 2012.

[2] J. Wang, X. Wang, C. Ma, and L. Kou, A survey on the development status and application prospects of knowledge graph in smart grids, *IET Gener. Transm. Distrib.*, vol. 15, no. 3, pp. 383–407, 2021.

[3] E. I. Papageorgiou, C. Huszka, J. De Roo, N. Douali, M. C. Jaulent, and D. Colaert, Application of probabilistic and fuzzy cognitive approaches in semantic web framework for medical decision support, *Comput. Methods Programs Biomed.*, vol. 112, no. 3, pp. 580–598, 2013.

[4] T. Yu, J. Li, Q. Yu, Y. Tian, X. Shun, L. Xu, L. Zhu, and H. Gao, Knowledge graph for TCM health preservation: Design, construction, and applications, *Artif. Intell. Med.*, vol. 77, pp. 48–52, 2017.

[5] P. Ping, K. Watson, J. Han, and A. Bui, Individualized knowledge graph: A viable informatics path to precision medicine, *Circ. Res.*, vol. 120, no. 7, pp. 1078–1080, 2017.

[6] D. M. Bean, H. Wu, E. Iqbal, O. Dzahini, Z. M. Ibrahim, M. Broadbent, R. Stewart, and R. J. B. Dobson, Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records, *Sci. Rep.*, vol. 7, no. 1, p. 16416, 2017.

[7] G. Bakal, P. Talari, E. V. Kakani, and R. Kavuluru, Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations, *J. Biomed. Inform.*, vol. 82, pp. 189–199, 2018.

[8] Y. Dai, C. Guo, W. Guo, and C. Eickhoff, Drug-drug interaction prediction with Wasserstein Adversarial Autoencoder-based knowledge graph embeddings, *Brief. Bioinform.*, vol. 22, no. 4, p. bbaa256, 2020.

[9] H. Xu, J. Li, L. Nie, and X. Xu, CKGECS: A Chinese knowledge graph for elderly care service, *Int.J. Serv. Technol. Manag.*, vol. 26, nos. 2&3, pp. 131–148, 2020.

[10] C. Chen, K. E. Ross, S. Gavali, J. E. Cowart, and C. H. Wu, COVID-19 knowledge graph from semantic integration of biomedical literature and databases, *Bioinformatics*, vol. 37, no. 23, pp. 4597–4598, 2021.

[11] Q. Zhao, J. Li, C. Xu, J. Yang, and L. Zhao, Knowledge-enhanced relation extraction for Chinese EMRs, *IT Prof.*, vol. 22, no. 4, pp. 57–62, 2020.

[12] T. D. Gunter and N. P. Terry, The emergence of national electronic health record architectures in the United States and Australia: Models, costs, and questions, *J. Med. Internet Res.*, vol. 7, no. 1, p. e3, 2005.

[13] Z. Huang, Q. Hu, M. Liao, C. Miao, C. Wang, and G. Liu, Knowledge graphs of Kawasaki disease, *Health Inf. Sci. Syst.*, vol. 9, no. 1, p. 11, 2021.

[14] X. Tao, T. Pham, J. Zhang, J. Yong, W. P. Goh, W. Zhang, and Y. Cai, Mining health knowledge graph for health risk prediction, *World Wide Web*, vol. 23, no. 4, pp. 2341–2362, 2020.

[15] S. M. S. Hasan, D. Rivera, X. C. Wu, E. B. Durbin, J. B. Christian, and G. Tourassi, Knowledge graph-enabled cancer data analytics, *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 1952–1967, 2020.

[16] C. E. Lipscomb, Medical Subject Headings (MeSH), *Bull. Med. Libr. Assoc.*, vol. 88, no. 3, pp. 265–266, 2000.

[17] K. Donnelly, SNOMED-CT: The advanced terminology and coding system for eHealth, *Stud. Health Technol. Inform.*, vol. 121, pp. 279–290, 2006.

[18] O. Bodenreider, The Unified Medical Language System (UMLS): Integrating biomedical terminology, *Nucl. Acids Res.*, vol. 32, no. S1, pp. D267-D270, 2004.

[19] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, DrugBank: A comprehensive resource for in silico drug discovery and exploration, *Nucl. Acids Res.*, vol. 34, pp. D668–D672, 2006.

[20] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, KEGG: New perspectives on genomes, pathways, diseases and drugs, *Nucl. Acids Res.*, vol. 45, no. D1, pp. D353–D361, 2017.

[21] J. Xie, J. Jiang, Y. Wang, Y. Guan, and X. Guo, Learning an expandable EMR-based medical knowledge network to enhance clinical diagnosis, *Artif. Intell. Med.*, vol. 107, p. 101927, 2020.

[22] Y. Zhang, M. Sheng, R. Zhou, Y. Wang, G. Han, H. Zhang, C. Xing, and J. Dong, HKGB: An inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians' expertise incorporated, *Inform. Proc. Manag.*, vol. 57, no. 6, p. 1023246, 2020.

[23] K. M. Malik, M. Krishnamurthy, M. Alobaidi, M. Hussain, F. Alam, and G. Malik, Automated domain-specific healthcare knowledge graph curation framework: Subarachnoid hemorrhage as phenotype, *Expert Syst. Appl.*, vol. 145, p. 113120, 2020.

[24] M. A. Musen, The protégé project: A look back and a look forward, *AI Matt.*, vol. 1, no. 4, pp. 4–12, 2015.

[25] M. Alobaidi, K. M. Malik, and M. Hussain, Automated ontology generation framework powered by linked biomedical ontologies for disease-drug domain, *Comput. Methods Programs Biomed.*, vol. 165, pp. 117–128, 2018.

[26] S. Feng, H. Ning, S. Yang, and D. Zhao, Geriatric disease reasoning based on knowledge graph, in *Int. 2019 Cyberspace Congress, CyberDI and CyberLife on Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health*, Beijing, China, 2019, pp. 452–465.

[27] D. Korn, T. Bobrowski, M. Li, Y. Kebede, P. Wang, P. Owen, G. Vaidya, E. Muratov, R. Chirkova, C. Bizon, et al., COVID-KOP: Integrating Emerging COVID-19 Data with the ROBOKOP Database, *Bioinformatics*, vol. 37, no. 4, pp. 586–587, 2021.

[28] F. Zhang, B. Sun, X. Diao, W. Zhao, and T. Shu, Prediction of adverse drug reactions based on knowledge graph embedding, *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, p. 38, 2021.

[29] K. Zhang, K. Li, H. Ma, D. Yue, and L. Zhuang, Construction of MeSH-like obstetric knowledge graph, in *Proc. 2018 Int. Conf. Cyber-Enabled Distributed Computing and Knowledge Discovery*, New York, NY, USA, 2018, pp. 160–168.

[30] C. Zhao, J. Jiang, Y. Guan, X. Guo, and B. He, EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning, *Artif. Intell. Med.*, vol. 87, pp. 49–59, 2018.

[31] M. R. Kamdar, C. E. Stanley Jr, M. Carroll, L. Wogulis, W. Dowling, H. F. Deus, and M. Samarasinghe, Text snippets to corroborate medical relations: An unsupervised approach using a knowledge graph and embeddings, *AMIA Jt. Summits Transl. Sci. Proc.*, vol. 2020, pp. 288–297, 2020.

[32] F. Michel, F. Gandon, V. Ah-Kane, A. Bobasheva, E. Cabrio, O. Corby, R. Gazzotti, A. Giboin, S. Marro, T. Mayer, et al., Covid-on-the-Web: Knowledge graph and services to advance COVID-19 research, in *19th Int. Semantic Web Conf. on the Semantic Web*, Athens, Greece, 2020, pp. 294–310.

[33] P. Liu, X. Wang, X. Sun, X. Shen, X. Chen, Y. Sun, and Y. Pan, HKDP: A hybrid knowledge graph based pediatric disease prediction system, in *Int. Conf. on Smart Health*, Haikou, China, 2017, pp. 78–90.

[34] X. Li, Y. Wang, D. Wang, W. Yuan, D. Peng, and Q. Mei, Improving rare disease classification using imperfect knowledge graph, *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 5, p. 238, 2019.

[35] A. Ettorre, O. R. Rodríguez, C. Faron, F. Michel, and F. Gandon, A knowledge graph enhanced learner model to predict outcomes to questions in the medical field, in *22nd Int. Conf. on Knowledge Engineering and Knowledge Management*, Bolzano, Italy, 2020, pp. 237–251.

[36] S. Ansong, K. F. Eteffa, C. Li, M. Sheng, Y. Zhang, and C. Xing, How to empower disease diagnosis in a medical education system using knowledge graph, in *16th Int. Conf. on Web Information Systems and Applications*, Qingdao, China, 2019, pp. 518–523.

[37] M. Wang, X. Ma, J. Si, H. Tang, H. Wang, T. Li, W. Ouyang, L. Gong, Y. Tang, and X. He, et al., Adverse drug reaction discovery using a tumor-biomarker knowledge graph, *Front. Genet.*, vol. 11, p. 625659, 2021.

[38] I. Abdelaziz, A. Fokoue, O. Hassanzadeh, P. Zhang, and M. Sadoghi, Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions, *J. Web Semant.*, vol. 44, pp. 104–117, 2017.

[39] M. Sheng, Q. Hu, Y. Zhang, C. Xing, and T. Zhang, A data-intensive CDSS platform based on knowledge graph, in *7th Int. Conf. on Health Information Science*, Cairns, Australia, 2018, pp. 146–155.

[40] F. Song, B. Wang, Y. Tang, and J. Sun, Research of medical aided diagnosis system based on temporal knowledge graph, in *16th Int. Conf. on Advanced Data Mining and Applications*, Foshan, China, 2020, pp. 236–250.

[41] Q. Bao, L. Ni, and J. Liu, HHH: An online medical chatbot system based on knowledge graph and hierarchical bi-directional attention, in *Proc. Australasian Computer Science Week Multiconference*, Melbourne, Australia, 2020, pp. 1–10.

[42] Z. Jiang, C. Chi, and Y. Zhan, Research on medical question answering system based on knowledge graph, *IEEE Access*, vol. 9, pp. 21094–21101, 2021.

[43] Y. Duan, P. Ji, L. Jin, A. Zou, J. Yang, H. Xie, and N. An, A knowledge graph for eldercare: Constructing a domain entity graph with guidelines, in *4th Int. Conf. on Human Aspects of IT for the Aged Population. Applications in Health, Assistance, and Entertainment*, Vegas, NV, USA, 2018, pp. 25–35.

[44] X. He, R. Zhang, R. Rizvi, J. Vasilakes, X. Yang, Y. Guo, Z. He, M. Prosperi, J. Huo, J. Alpert, et al., ALOHA: Developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design, *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 4, p. 150, 2019.

[45] H. Sun, J. Xiao, W. Zhu, Y. He, S. Zhang, X. Xu, L. Hou, J. Li, Y. Ni, and G. Xie, Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: Model development and performance evaluation, *JMIR Med. Inform.*, vol. 8, no. 7, p. e17653, 2020.

[46] T. Pham, X. Tao, J. Zhang, and J. Yong, Constructing a knowledge-based heterogeneous information graph for medical health status classification, *Health Inf. Sci. Syst.*, vol. 8, no. 1, p. 10, 2020.

[47] L. Li, P. Wang, J. Yan, Y. Wang, S. Li, J. Jiang, Z. Sun, B. Tang, T. H. Chang, S. Wang, et al., Real-world data medical knowledge graph: Construction and applications, *Artif. Intell. Med.*, vol. 103, p. 101817, 2020.

[48] M. Wang, J. Zhang, J. Liu, W. Hu, S. Wang, X. Li, and W. Liu, PDD Graph: Bridging electronic medical records and biomedical knowledge graphs via entity linking, in *16th Int. Semantic Web Conf. on the Semantic Web*, Vienna, Austria, 2017, pp. 219–227.

[49] X. Zhang and C. Che, Drug repurposing for parkinson's disease by integrating knowledge graph completion model and knowledge fusion of medical literature, *Future Internet*, vol. 13, no. 1, p. 14, 2021.

[50] P. Ernst, A. Siu, and G. Weikum, KnowLife: A versatile approach for constructing a large knowledge graph for biomedical sciences, *BMC Bioinform.*, vol. 16, no. 1, p. 157, 2015.

[51] M. A. Patil, S. Bhaumik, S. Paul, S. Bissoyi, R. Roy, and S. Ryu, Estimating personalized risk ranking using laboratory test and medical knowledge (UMLS), in *2013 35th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, 2013, pp. 1274–1277.

[52] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, Freebase: A collaboratively created graph database for structuring human knowledge, in *Proc. 2008 ACM SIGMOD Int. Conf. Management of Data*, Vancouver, Canada, 2008, pp. 1247–1250.

[53] A. S. Brown and C. J. Patel, A standard database for drug repositioning, *Sci. Data*, vol. 4, no. 1, p. 170029, 2017.

[54] H. Kilicoglu, D. Shin, M. Fiszman, G. Rosemblat, and T. C. Rindflesch, SemMedDB: A PubMed-scale repository of biomedical semantic predications, *Bioinformatics*, vol. 28, no. 23, pp. 3158–3160, 2012.

[55] F. Teng, W. Yang, L. Chen, L. Huang, and Q. Xu, Explainable prediction of medical codes with knowledge graphs, *Front. Bioeng. Biotechnol.*, vol. 8, p. 867, 2020.

[56] T. B. Malas, W. J. Vlietstra, R. Kudrin, S. Starikov, M. Charrout, M. Roos, D. J. M. Peters, J. A. Kors, R. Vos, P. A. C. 't Hoen, et al., Drug prioritization using the semantic properties of a knowledge graph, *Sci. Rep.*, vol. 9, no. 1, p. 6281, 2019.

[57] C. Bizon, S. Cox, J. Balhoff, Y. Kebede, P. Wang, K. Morton, K. Fecho, and A. Tropsha, ROBOKOP KG and KGB: Integrated knowledge graphs from federated sources, *J. Chem. Inf. Model.*, vol. 59, no. 12, pp. 4968–4973, 2019.

[58] T. R. Goodwin and S. M. Harabagiu, Medical question answering for clinical decision support, *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, vol. 2016, pp. 297–306, 2016.

[59] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, and D. Sontag, Learning a health knowledge graph from electronic medical records, *Sci. Rep.*, vol. 7, no. 1, p. 5994, 2017.

[60] B. Cheng, Y. Zhang, D. Cai, W. Qiu, and D. Shi, Construction of traditional Chinese medicine knowledge graph using data mining and expert knowledge, in *Proc. 2018 IEEE Int. Conf. Network Infrastructure and Digital Content*, Guiyang, China, 2018, pp. 209–213.

[61] M. Sheng, Y. Shao, Y. Zhang, C. Li, C. Xing, H. Zhang, J. Wang, and F. Gao, DEKGB: An extensible framework for health knowledge graph, in *International Conference on Smart Health*, Shenzhen, China, 2019, pp. 27–38.

[62] L. Li, P. Wang, Y. Wang, S. Wang, J. Yan, J. Jiang, B. Tang, C. Wang, and Y. Liu, A method to learn embedding of a probabilistic medical knowledge graph: Algorithm development, *JMIR Med. Inform.*, vol. 8, no. 5, p. e17645, 2020.

[63] L. Wang, H. Xie, W. Han, X. Yang, L. Shi, J. Dong, K. Jiang, and H. Wu, Construction of a knowledge graph

for diabetes complications from expert-reviewed clinical evidences, *Comput. Assist. Surg.*, vol. 25, no. 1, pp. 29–35, 2020.

[64] M. Sheng, A. Li, Y. Bu, J. Dong, Y. Zhang, X. Li, C. Li, and C. Xing, DSQA: A domain specific QA system for smart health based on knowledge graph, in *17th Int. Conf. on Web Information Systems and Applications*, Guangzhou, China, 2020, pp. 215–222.

[65] X. Li, H. Liu, X. Zhao, G. Zhang, and C. Xing, Automatic approach for constructing a knowledge graph of knee osteoarthritis in Chinese, *Health Inf. Sci. Syst.*, vol. 8, no. 1, p. 12, 2020.

[66] F. Gong, M. Wang, H. Wang, S. Wang, and M. Liu, SMR: Medical knowledge graph embedding for safe medicine recommendation, *Big Data Res.*, vol. 23, p. 100174, 2021.

[67] Z. Lin, D. Yang, and X. Yin, Patient similarity via joint embeddings of medical knowledge graph and medical entity descriptions, *IEEE Access*, vol. 8, pp. 156663–156676, 2020.

[68] W. Deng, P. Guo, and J. Yang, Medical entity extraction and knowledge graph construction, in *2019 16th Int. Computer Conf. Wavelet Active Media Technology and Information Processing*, Chengdu, China, 2019, pp. 41–44.

[69] T. Ruan, Y. Huang, X. Liu, Y. Xia, and J. Gao, QAnalysis: A question-answer driven analytic tool on knowledge graphs for leveraging electronic medical records for clinical research, *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 82, 2019.

[70] W. Van Woensel, C. Armstrong, M. Rajaratnam, V. Gupta, and S. S. R. Abidi, Using knowledge graphs to plausibly infer missing associations in EMR data, *Stud. Health Technol. Inform.*, vol. 281, pp. 417–421, 2021.

[71] A. Fang, P. Lou, J. Hu, W. Zhao, M. Feng, H. Ren, and X. Chen, Head and tail entity fusion model in medical knowledge graph construction: Case study for pituitary adenoma, *JMIR Med. Inform.*, vol. 9, no. 7, p. e28218, 2021.

[72] A. K. Talukder, J. B. Sanz, and J. Samajpati, 'Precision Health': Balancing reactive care and proactive care through the evidence based knowledge graph constructed from real-world electronic health records, disease trajectories, diseasome, and patholome, in *8th Int. Conf. on Big Data Analytic*s, Sonepat, India, 2020, pp. 113–133.

[73] I. Y. Chen, M. Agrawal, S. Horng, and D. Sontag, Robustly extracting medical knowledge from EHRs: A case study of learning a health knowledge graph, *Pac. Symp. Biocomput.*, vol. 25, pp. 19–30, 2020.

[74] L. Shi, S. Li, X. Yang, J. Qi, G. Pan, and B. Zhou, Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services, *Biomed. Res. Int.*, vol. 2017, p. 2858423, 2017.

[75] J. Lee, C. Liu, J. H. Kim, A. Butler, N. Shang, C. Pang, K. Natarajan, P. Ryan, C. Ta, and C. Weng, Comparative effectiveness of medical concept embedding for feature engineering in phenotyping, *JAMIA Open*, vol. 4, no. 2, p. ooab028, 2021.

[76] K. He, L. Yao, J. Zhang, Y. Li, and C. Li, Construction of genealogical knowledge graphs from obituaries: Multitask neural network extraction system, *J. Med. Internet. Res.*, vol. 23, no. 8, p. e25670, 2021.

[77] H. Weng, Z. Liu, S. Yan, M. Fan, A. Ou, D. Chen, and T. Hao, A framework for automated knowledge graph construction towards traditional Chinese medicine, in *6th Int. Conf. on Health Information Science*, Moscow, Russia, 2017, pp. 170–181.

[78] J. Du and X. Li, A knowledge graph of combined drug therapies using semantic predications from biomedical literature: Algorithm development, *JMIR Med. Inform.*, vol. 8, no. 4, p. e18323, 2020.

[79] Z. Huang, J. Yang, F. van Harmelen, and Q. Hu, Constructing knowledge graphs of depression, in lecture notes in computer science, in *6th Int. Conf. on Health Information Science*, Moscow, Russia, 2017, pp. 149–161.

[80] S. Sang, Z. Yang, L. Wang, X. Liu, H. Lin, and J. Wang, SemaTyP: A knowledge graph based literature mining method for drug discovery, *BMC Bioinform.*, vol. 19, no. 1, p. 193, 2018.

[81] J. Yuan, Z. Jin, H. Guo, H. Jin, X. Zhang, T. Smith, and J. Luo, Constructing biomedical domain-specific knowledge graph with minimum supervision, *Knowl. Inf. Syst.*, vol. 62, no. 1, pp. 317–336, 2020.

[82] W. J. Vlietstra, R. Zielman, R. M. van Dongen, E. A. Schultes, F. Wiesman, R. Vos, E. M. van Mulligen, and J. A. Kors, Automated extraction of potential migraine biomarkers using a semantic graph, *J. Biomed. Inform.*, vol. 71, pp. 178–189, 2017.

[83] A. Rossanez, J. C. dos Reis, R. da Silva Torres, and H. de Ribaupierre, KGen: A knowledge graph generator from biomedical scientific literature, *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 4, p. 314, 2020.

[84] D. N. Sosa, A. Derry, M. Guo, E. Wei, C. Brinton, and R. B. Altman, A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases, *Pac. Symp. Biocomput.*, vol. 25, pp. 463–474, 2020.

[85] Y. Song, L. Cai, K. Zhang, H. Zan, T. Liu, and X. Ren, Construction of Chinese pediatric medical knowledge graph, in *9thJoint Int. Conf. on Semantic Technology*, Hangzhou, China, 2020, pp. 213–220.

[86] G. Bakal and R. Kavuluru, Predicting treatment relations with semantic patterns over biomedical knowledge graphs, in 3rd *Int. Conf. on Mining Intelligence and Knowledge Exploration*, Hyderabad, India, 2015, pp. 586–596.

[87] H. Fei, Y. Ren, Y. Zhang, D. Ji, and X. Liang, Enriching contextualized language model from knowledge graph for biomedical information extraction, *Brief. Bioinform.*, vol. 22, no. 3, p. bbaa110, 2021.

[88] D. Zhao, J. Wang, S. Sang, H. Lin, J. Wen, and C. Yang, Relation path feature embedding based convolutional neural network method for drug discovery, *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 2, p. 59, 2019.

[89] C. T. Hoyt, D. Domingo-Fernández, R. Aldisi, L. Xu, K. Kolpeja, S. Spalek, E. Wollert, J. Bachman, B. M. Gyori,

P. Greene, et al., Re-curation and rational enrichment of knowledge graphs in biological expression language, *Database*, vol. 2019, p. baz068, 2019.

[90] T. Pham, X. Tao, J. Zhang, J. Yong, X. Zhou, and R. Gururajan, MeKG: Building a medical knowledge graph by data mining from MEDLINE, in *12th Int. Conf. on Brain Informatics*, Haikou, China, 2019, pp. 159–168.

[91] X. Yang, C. Wu, G. Nenadic, W. Wang, and K. Lu, Mining a stroke knowledge graph from literature, *BMC Bioinform.*, vol. 22, no. 10, p. 387, 2021.

[92] X. Chai, Diagnosis method of thyroid disease combining knowledge graph and deep learning, *IEEE Access*, vol. 8, pp. 149787–149795, 2020.

[93] Y. Liu, H. Li, and Y. Chen, Using knowledge graph to handle label imperfection, in *PAKDD 2014 Int. Workshops*: *DANTH*, *BDM*, *MobiSocial*, *BigEC*, *CloudSD*, *MSMV-MBI*, *SDA*, *DMDA-Health*, *ALSIP*, *SocNet*, *DMBIH*, *BigPMA on Trends and Applications in Knowledge Discovery and Data Mining*, Tainan, China, 2014, pp. 345–356.

[94] Q. Zhu, D. T. Nguyen, I. Grishagin, N. Southall, E. Sid, and A. Pariser, An integrative knowledge graph for rare diseases, derived from the Genetic and Rare Diseases Information Center (GARD), *J. Biomed. Semant.*, vol. 11, no. 1, p. 13, 2020.

[95] T. Liu, X. Pan, X. Wang, K. A. Feenstra, J. Heringa, and Z. Huang, Predicting the relationships between gut microbiota and mental disorders with knowledge graphs, *Health Inf. Sci. Syst.*, vol. 9, no. 1, p. 3, 2021.

[96] H. Gao, Y. Ni, X. Mo, D. Li, S. Teng, Q. Huang, S. Huang, G. Liu, S. Zhang, Y. Tang, et al., Drug repositioning based on network-specific core genes identifies potential drugs for the treatment of autism spectrum disorder in children, *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 3908–3921, 2021.

[97] W. Li, L. Chai, C. Yang, and X. Wang, An evolutionary analysis of DBpedia datasets, in *15th Int. Conf. on Web Information Systems and Applications*, Taiyuan, China, 2018, pp. 317–329.

[98] N. Li, Z. Yang, L. Luo, L. Wang, Y. Zhang, H. Lin, and J. Wang, KGHC: A knowledge graph for hepatocellular carcinoma, *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 3, p. 135, 2020.

[99] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman, and H. Kilicoglu, Drug repurposing for COVID-19 via knowledge graph completion, *J. Biomed. Inform.*, vol. 115, p. 103696, 2021.

[100] Q. Wang, T. Wang, and C. Xu, Using a knowledge graph for hypernymy detection between Chinese symptoms, in *2018 Tenth Int. Conf. Advanced Computational Intelligence* (*ICACI*), Xiamen, China, 2018, pp. 601–606.

[101] S. Ganesh and A. K. Talukder, Formal methods, artificial intelligence, big-data analytics, and knowledge engineering in medical care to reduce disease burden and health disparities, in *6th Int. Conf. on Big Data Analytics*, Warangal, 2018, pp. 307–321.

[102] S. K. Mohamed, V. Nováček, and A. Nounu, Discovering protein drug targets using knowledge graph embeddings, *Bioinformatics*, vol. 36, no. 2, pp. 603–610, 2020.

[103] K. Lei, K. Yuan, Q. Zhang, and Y. Shen, MedSim: A novel semantic similarity measure in bio-medical knowledge graphs, in *11th Int. Conf. on Knowledge Science*, *Engineering and Management*, Changchun, China, 2018, pp. 479–490.

[104] V. K. C. Yan, X. Li, X. Ye, M. Ou, R. Luo, Q. Zhang, B. Tang, B. J. Cowling, I. Hung, C. W. Siu, et al., Drug repurposing for the treatment of COVID-19: A knowledge graph approach, *Adv. Ther.*, vol. 4, no. 7, p. 2100055, 2021.

[105] Y. Pan, X. Lei, and Y. Zhang, Association predictions of genomics, proteinomics, transcriptomics, microbiome, metabolomics, pathomics, radiomics, drug, symptoms, environment factor, and disease networks: A comprehensive approach, *Med. Res. Rev.*, vol. 42, no. 1, pp. 441–461, 2022.

[106] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, The SIDER database of drugs and side effects, *Nucl. Acids Res.*, vol. 44, no. D1, pp. D1075–D1079, 2016.

[107] A. L. Mitchell, T. K. Attwood, P. C. Babbitt, M. Blum, P. Bork, A. Bridge, S. D. Brown, H. Y. Chang, S. El-Gebali, M. I. Fraser, et al., InterPro in 2019: Improving coverage, classification and access to protein sequence annotations, *Nucl. Acids Res.*, vol. 47, no. D1, pp. D351–D360, 2019.

[108] The UniProt Consortium, UniProt: The universal protein knowledgebase, *Nucl. Acids Res.*, vol. 45, no. D1, pp. D158–D169, 2017.

[109] R. T. Sousa, S. Silva, and C. Pesquita, Evolving knowledge graph similarity for supervised learning in complex biomedical domains, *BMC Bioinform.*, vol. 21, no. 1, p. 6, 2020.

[110] Y. Yang, Z. Huang, Y. Han, X. Hua, and W. Tang, Using knowledge graph for analysis of neglected influencing factors of statin-induced myopathy, in *Int. Conf. on Brain Informatics*, Beijing, China, 2017, pp. 304–311.

[111] B. Xu, Y. Xu, J. Liang, C. Xie, B. Liang, W. Cui, and Y. Xiao, CN-DBpedia: A never-ending Chinese knowledge extraction system, in *30th Int. Conf. on Industrial Engineering and Other Applications of Applied Intelligent Systems on Advances in Artificial Intelligence*: *From Theory to Practice*, Arras, France, 2017, pp. 428–438.

[112] F. Li, Y. Jin, W. Liu, B. P. S. Rawat, P. Cai, and H. Yu, Fine-tuning bidirectional encoder representations from transformers (BERT)-based models on large-scale electronic health record notes: An empirical study, *JMIR Med. Inform.*, vol. 7, no. 3, p. e14830, 2019.

[113] E. Tutubalina, Z. Miftahutdinov, S. Nikolenko, and V. Malykh, Medical concept normalization in social media posts with recurrent neural networks, *J. Biomed. Inform.*, vol. 84, pp. 93–102, 2018.

[114] Y. F. Luo, W. Sun, and A. Rumshisky, MCN: A comprehensive corpus for medical concept normalization, *J. Biomed. Inform.*, vol. 92, p. 103132, 2019.

[115] H. L. Nguyen, D. T. Vu, and J. J. Jung, Knowledge graph fusion for smart systems: A Survey, *Inform. Fusion*, vol. 61, pp. 56–70, 2020.

[116] Z. Lei, Y. Sun, Y. A. Nanehkaran, S. Yang, M. S. Islam, H.

Lei, and D. Zhang, A novel data-driven robust framework based on machine learning and knowledge graph for disease classification, *Future Gener. Comp. Syst.*, vol. 102, pp. 534–548, 2020.

[117] Y. Chen, T. Ma, X. Yang, J. Wang, B. Song, and X. Zeng, MUFFIN: Multi-scale feature fusion for drug-drug interaction prediction, *Bioinformatics*, vol. 37, no. 17, pp. 2651–2658, 2021.

[118] Q. Wang, Y. Ji, Y. Hao, and J. Cao, GRL: Knowledge graph completion with GAN-based reinforcement learning, *Knowl-Based. Syst.*, vol. 209, p. 106421, 2020.

[119] R. Zhang, Y. Mao, and W. Zhao, Knowledge graphs completion via probabilistic reasoning, *Inform. Sci.*, vol. 521, pp. 144–159, 2020.

[120] X. Chen, S. Jia, and Y. Xiang, A review: Knowledge reasoning over knowledge graph, *Expert Syst. Appl.*, vol. 141, p. 112948, 2020.

[121] J. Chu, J. Chen, X. Chen, W. Dong, J. Shi, and Z. Huang, Knowledge-aware multi-center clinical dataset adaptation: Problem, method, and application, *J. Biomed. Inform.*, vol. 115, p. 103710, 2021.

[122] D. Yue, K. Zhang, L. Zhuang, X. Zhao, O. Byambasuren, and H. Zan, Annotation scheme and specification for named entities and relations on Chinese medical knowledge graph, in *20th Workshop on Chinese Lexical Semantics*, Beijing, China, 2020, pp. 563–574.

[123] N. Pattisapu, V. Anand, S. Patil, G. Palshikar, and V. Varma, Distant supervision for medical concept normalization, *J. Biomed. Inform.*, vol. 109, p. 103522, 2020.

[124] S. Yin, D. Chen, and J. Le, Deep neural network based on translation model for diabetes knowledge graph, in *2017 5th Int. Conf. Advanced Cloud and Big Data*, Shanghai, China, 2017, pp. 318–323.

[125] S. Biswas, P. Mitra, and K. S. Rao, Relation prediction of Co-morbid diseases using knowledge graph completion, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 2, pp. 708–717, 2019.

[126] J. Santisteban and J. L. Tejada Carcamo, Unilateral Jaccard similarity coefficient, in *SIGIR Workshop on Graph Search and Beyond'15*, Santiago, Chile, 2015, pp. 23–27.

[127] M. Sheng, H. Zhang, Y. Zhang, C. Li, C. Xing, J. Wang, Y. Shao, and F. Gao, CLMed: A cross-lingual knowledge graph framework for cardiovascular diseases, in *16th Int. Conf. on Web Information Systems and Applications*, Qingdao, China, 2019, pp. 512–517.

[128] M. Sheng, J. Wang, Y. Zhang, X. Li, C. Li, C. Xing, Q. Li, Y. Shao, and H. Zhang, DocKG: A knowledge graph framework for health with doctor-in-the-loopp, in *8th Int. Conf. on Health Information Science*, Xi'an, China, 2019, pp. 3–14.

[129] J. T. Reese, D. Unni, T. J. Callahan, L. Cappelletti, V. Ravanmehr, S. Carbon, K. A. Shefchek, B. M. Good, J. P. Balhoff, T. Fontana, et al., KG-COVID-19: A framework to produce customized knowledge graphs for COVID-19 response, *Patterns*, vol. 2, no. 1, p. 100155, 2021.

[130] X. Xiu, Q. Qian, and S. Wu, Construction of a digestive system tumor knowledge graph based on Chinese electronic medical records: Development and usability study, *JMIR Med. Inform.*, vol. 8, no. 10, p. e18287, 2020.

[131] B. M. J. Some, G. Bordea G, F. Thiessard, S. Schulz, and G. Diallo, Design considerations for a knowledge graph: The WATRIMed use case, *Stud. Health Technol. Inform.*, vol. 259, pp. 59–64, 2019.

[132] S. Zheng, J. Rao, Y. Song, J. Zhang, X. Xiao, E. F. Fang, Y. Yang, and Z. Niu, PharmKG: A dedicated knowledge graph benchmark for bomedical data mining, *Brief. Bioinform.*, vol. 22, no. 4, p. bbaa344, 2021.

[133] Y. Zhou, T. Zhou, T. Zhou, H. Fu, J. Liu, and L. Shao, Contrast-attentive thoracic disease recognition with dual-weighting graph reasoning, *IEEE Trans. Med. Imaging.*, vol. 40, no. 4, pp. 1196–1206, 2021.

[134] X. Sun, Y. Man, Y. Zhao, J. He, and N. Liu, Incorporating description embeddings into medical knowledge graphs representation learning, in *4th Int. Conf. on Human Centered Computing*, Mérida, Mexico, 2019, pp. 188–194.

[135] Q. Li, L. Li, J. Zhong, and L. F. Huang, Real-time sepsis severity prediction on knowledge graph deep learning networks for the intensive care unit, *J. Vis. Commun. Image Represent.*, vol. 72, p. 102901, 2020.

[136] J. Jiang, H. Wang, J. Xie, X. Guo, Y. Guan, and Q. Yu, Medical knowledge embedding based on recursive neural network for multi-disease diagnosis, *Artif. Intell. Med.*, vol. 103, p. 101772, 2020.

[137] X. Wang, Y. Zhang, X. Wang, and J. Chen, A knowledge graph enhanced topic modeling approach for herb recommendation, in *24th Int. Conf. on Database Systems for Advanced Applications*, Chiang Mai, Thailand, 2019, pp. 709–724.

[138] W. J. Vlietstra, R. Vos, A. M. Sijbers, E. M. van Mulligen, and J. A. Kors, Using predicate and provenance information from a knowledge graph for drug efficacy screening, *J. Biomed. Semant.*, vol. 9, no. 1, p. 23, 2018.

[139] C. Moon, C. Jin, X. Dong, S. Abrar, W. Zheng, R. Y. Chirkova, and A. Tropsha, Learning Drug-Disease-Target Embedding (DDTE) from knowledge graphs to inform drug repurposing hypotheses, *J. Biomed. Inform.*, vol. 119, p. 103838, 2021.

[140] Y. Shen, K. Yuan, J. Dai, B. Tang, M. Yang, and K. Lei, KGDDS: A system for drug-drug similarity measure in therapeutic substitution based on knowledge graph curation, *J. Med. Syst.*, vol. 43, no. 4, p. 92, 2019.

[141] X. Liang, D. Li, M. Song, A. Madden, Y. Ding, and Y. Bu, Predicting biomedical relationships using the knowledge and graph embedding cascade model, *PLoS One*, vol. 14, no. 6, p. e0218264, 2019.

[142] T. Qi, S. Qiu, X. Shen, H. Chen, S. Yang, H. Wen, Y. Zhang, Y. Wu, and Y. Huang, KeMRE: Knowledge-enhanced medical relation extraction for Chinese medicine instructions, *J. Biomed. Inform.*, vol. 120, p. 103834, 2021.

[143] Y. Ge, T. Tian, S. Huang, F. Wan, J. Li, S. Li, X. Wang,

H. Yang, L. Hong, N. Wu, et al., An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19, *Signal Transduct. Target. Ther.*, vol. 6, no. 1, p. 165, 2021.
[144] J. Al-Saleem, R. Granet, S. Ramakrishnan, N. A. Ciancetta, C. Saveson, C. Gessner, and Q. Q. Zhou, Knowledge graph-based approaches to drug repurposing for COVID-19, *J. Chem. Inf. Model.*, vol. 61, no. 8, pp. 4058–4067, 2021.
[145] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, Clinical big data and deep learning: Applications, challenges, and future outlooks, *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 288–305, 2019.

Xuehong Wu received the BS and MS degrees in computer engineering from Central South University, Changsha, China, in 2010 and 2013, respectively. He is currently working toward the PhD degree in the School of Computer Science and Engineering, Central South University, China. His research interests include medical big data and knowledge graphs.

Yi Pan received the BEng (computer engineering) and MEng (computer engineering) degrees from Tsinghua University, Beijing, China in 1982 and 1984, respectively, and the PhD degree (computer science) from the University of Pittsburgh, USA in 1991. He is currently working as a dean and chair professor at the Faculty of Computer Science and Control Engineering, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences Shenzhen, Guangdong, China. He has published over 250 academic papers in SCI-indexed journals, including more than 100 papers in top IEEE/ACM transactions/journals. His publications have been cited more than 19 000 times, and his current H-index is 88. He has been awarded IEEE Distinguished Achievement Award, IEEE Distinguished Service Award, IEEE Transactions Best Paper Award, IEEE and other international conference best paper awards many times, IBM Professor Award four times, Andrew Mellon Award, and other awards. His research interests include parallel and distributed processing systems, Internet technology, and bioinformatics.

Junwen Duan received the BS in information security and PhD degrees in computer science from Harbin Institute of Technology, Harbin, China, in 2013 and 2020, respectively. He is currently a lecturer at the School of Computer Science and Engineering, Central South University. His main research interests include natural language processing and text mining.

Min Li received the PhD degree in computer science from Central South University, China, in 2008. She is currently the vice dean and a professor at the School of Computer Science and Engineering, Central South University, Changsha, Hunan, China. Her research interests include computational biology, systems biology, and bioinformatics. She has published more than 100 technical papers in refereed journals such as *Genome Research*, *Genome Biology*, *Nucleic Acids Research*, *Bioinformatics*, and conference proceedings.