Journal of Biomedical Informatics 143 (2023) 104403

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Methodological Review

Towards electronic health record-based medical knowledge graph construction, completion, and applications: A literature study

Lino Murali [a,c], G. Gopakumar [b], Daleesha M. Viswanathan [c], Prema Nedungadi [a,b,*]

[a] Center for Research in Analytics and Technologies for Education (CREATE), Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, 690525, Kerala, India
[b] Department of Computer Science and Engineering, School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, 690525, Kerala, India
[c] Division of Information Technology, School of Engineering, Cochin University of Science and Technology, Kochi, 682022, Kerala, India

ARTICLE INFO

Keywords:
EHR data
Health knowledge graph
Clinical decision making

ABSTRACT

With the growth of data and intelligent technologies, the healthcare sector opened numerous technology that enabled services for patients, clinicians, and researchers. One major hurdle in achieving state-of-the-art results in health informatics is domain-specific terminologies and their semantic complexities. A knowledge graph crafted from medical concepts, events, and relationships acts as a medical semantic network to extract new links and hidden patterns from health data sources. Current medical knowledge graph construction studies are limited to generic techniques and opportunities and focus less on exploiting real-world data sources in knowledge graph construction. A knowledge graph constructed from Electronic Health Records (EHR) data obtains real-world data from healthcare records. It ensures better results in subsequent tasks like knowledge extraction and inference, knowledge graph completion, and medical knowledge graph applications such as diagnosis predictions, clinical recommendations, and clinical decision support. This review critically analyses existing works on medical knowledge graphs that used EHR data as the data source at (i) representation level, (ii) extraction level (iii) completion level. In this investigation, we found that EHR-based knowledge graph construction involves challenges such as high complexity and dimensionality of data, lack of knowledge fusion, and dynamic update of the knowledge graph. In addition, the study presents possible ways to tackle the challenges identified. Our findings conclude that future research should focus on knowledge graph integration and knowledge graph completion challenges.

1. Introduction

Graphs are a simple yet effective way of representing information in a structured manner in numerous real-life applications. Knowledge graphs describe semantic information and its relations in the form of nodes and edges of a graph. Information required for the knowledge graph is collected from experts, knowledge bases, websites, or other repositories and stored as unidirectional or bidirectional relations based on the nature of the data [1]. The expanding need to assimilate knowledge graph in the industry [2,3] contributes to many real-world applications ranging from social network analysis [4], recommendation systems [5], question-answering [6], information extraction [7], and named entity disambiguation [8].

Knowledge graphs are broadly classified as general and domain-specific based on their purpose and the type of information incorporated [9]. With the growing importance of artificial intelligence in medicine, a medical knowledge graph is the foundation of domain knowledge for several applications in medical informatics, such as diagnostics [10], treatment [11,12], and drug repurposing [13]. The emerging scope of knowledge graphs can contribute to the healthcare industry and academia by assisting in clinical decision-making, accelerating drug discovery, detecting adverse drug effects, recommending medicine, and predicting differential diagnosis. A medical knowledge graph (MKG) consists of medical concepts and events, such as diseases, symptoms, treatments, etc., as nodes and edges, establishing relevant interconnections between the nodes as relationships. It requires expert knowledge of the domain, which should be acquired either manually from domain experts or automatically from various data sources like scientific articles, web sources, textbooks, and actual patient records. Fig. 1 details various knowledge sources used for the medical knowledge graph construction. PubMed and CORD 19 [14] articles are the most common knowledge sources used in the knowledge graph created from scientific articles. To assimilate real-world medical knowledge, patient records in the form of Electronic Health Records (EHR) or

* Corresponding author at: Center for Research in Analytics and Technologies for Education (CREATE), Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, 690525, Kerala, India.
E-mail address: prema@amrita.edu (P. Nedungadi).

https://doi.org/10.1016/j.jbi.2023.104403
Received 13 February 2023; Received in revised form 16 May 2023; Accepted 19 May 2023
Available online 24 May 2023

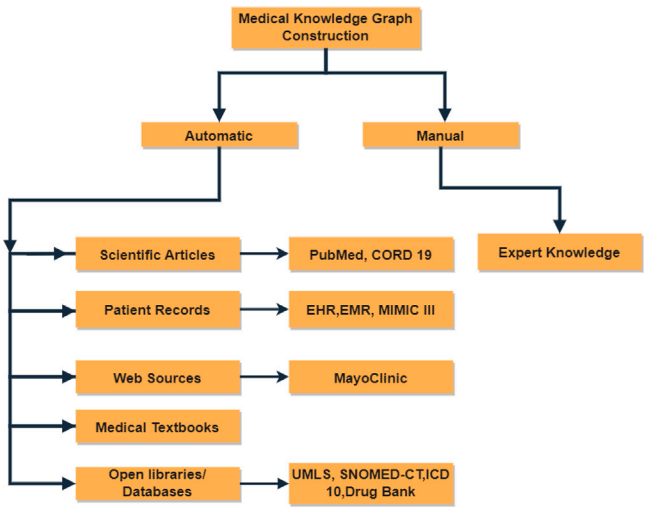

Fig. 1. Knowledge sources of medical knowledge graph.

Electronic Medical Records (EMR) such as Medical Information Mart for Intensive Care III (MIMIC III) [15] and open medical databases or libraries such as Unified Medical Language System (UMLS) [16], SNOMED-CT [17], ICD-10 [18], DrugBank [19], etc are also useful.

Recent advances in text mining and natural language processing enable medical literature, web articles, and EHR data to serve as knowledge sources of automatic knowledge graph construction [1, 20]. The critical problem with MKG curated from medical textbooks, literature, websites, and medical databases is that they provide guidelines for ideal situations and satisfies common information retrieval purpose only [21]. But a real-world knowledge source such as EHR data is crucial to provide better diagnostic support. As EHR contains heterogeneous information on the health status of patients, such knowledge graphs can better furnish the evidence required in a clinical practice [22]. Learning medical knowledge graphs from EHR data incorporates real-world data with medical knowledge and makes the process effortless compared to scientific article-based medical knowledge graphs [20].

Fig. 2 reveals that there has been a steady rise in the number of published articles in the knowledge graph domain that incorporated the use of EHR data. The number of articles shown in Fig. 2 are the articles selected for this literature survey using the search methodology mentioned in Section 3. The purpose of this paper is to review the literature on EHR-based medical knowledge graphs, which use real-world patient data for knowledge graph construction. The study set out to investigate the usefulness of longitudinal, heterogeneous, and complex EHR data compared to healthcare data obtained from disparate data sources. A comprehensive view of literature on medical knowledge graph at the representation level, extraction level, and completion level recognizes that there is much less research conducted in (i) handling complex real-world data of patients in MKG (ii) utilizing the semantic features in the descriptions of clinical concepts for knowledge alignment (iii) addressing the growing and changing nature of entities and relationships in the healthcare domain. Hence future research in medical knowledge graphs needs to focus on creating EHR-based knowledge graphs with a semantic representation of heterogeneous patient data, effective multi-source knowledge integration, and temporal knowledge graph completion.

2. Search methodology

We systematically searched the keywords ‘EHR-based medical knowledge graph construction OR Medical knowledge graph construction’, ‘Medical knowledge graph completion’, and ‘Multiview medical knowledge graph integration OR Multi-source medical knowledge graph integration’ in databases Scopus and PubMed and selected conferences for the last ten years of data. Fig. 3 shows the process of identifying and screening steps for selecting articles. The retrieved results are deduplicated and filtered by screening the title and domain. Then we again filtered the papers by reading abstracts and excluding works done on regional medical systems. After reading the complete text, we chose articles most relevant to EHR and biomedical knowledge graph construction.

The remaining part of this survey paper proceeds as follows: Section 3 describes the characteristics, challenges, and sources of EHR data. Section 4 is concerned with the subtasks of medical knowledge graph construction organized as representation, extraction, and completion, and Section 5 examines general approaches to MKG construction seen in the literature. The next part gives a brief overview of the applications of a medical knowledge graph. The sixth section presents the study’s findings with a discussion and future research scope, followed by a conclusion in Section 7.

3. EHR data: Sources, characteristics, and challenges

Patient information collected during hospital visits and the details of test results, symptoms, diagnosis, and medications are often stored in hospitals as EHR. These records contain unstructured data in the form of free notes, such as progress notes, and structured information, like patient details and lab report data [23]. In a real-time patient care system, such information embedded in EHR data helps integrate knowledge about diseases, therapies, and proteomics into clinical knowledge graphs [24]. In this section, we analyze specific features of EHR data and its challenges in using it as a primary data source in clinical knowledge graph creation.

3.1. Data sources

Primary sources of EHR data are the information stored in hospitals as patient records keeping their medical and personal details. The study and analysis of patient records require information retrieval and should not compromise patients’ privacy. Hence the biggest challenge associated with EHR-based research is that data availability heavily depends on organizational and human research policies. However, there are publicly available EHR data sets for research purposes. MIMIC-III is an extensive, centralized database of information on patients admitted to the ICUs of a large tertiary care hospital [15]. The National NLP Clinical Challenges (n2c2) datasets consist of fully de-identified clinical notes and products from challenges and n2c2 shared tasks [25]. They are freely available to the research community but are subject to a Data Use Agreement (DUA) to follow.

3.2. Multimodality in EHR

Patient records store static data like patient demographics, dynamic data like values of vital symptoms, laboratory test results, radiology reports with medical images, text data, and progress notes. Despite the complexity of including all modalities, recent EHR-based research considers most of these data in training to enrich the model’s knowledge. To integrate modalities, an encoder-based disease prediction model [26] applied fusion strategies and integrated time-invariant data, time series data, and clinical notes from EHR. They used fusion with an attention gate in each encoder for each modality and checked the significance of each modality by switching the main modality in the diagnosis prediction. There are two main phases in integrating multimodal data in medical knowledge graph building; (i) proper representation of multimodal data from structured knowledge and text from multiple sources using representation learning to overcome structural differences [27,28] (ii) adapting multimodal learning for better reasoning, classification, and prediction depending on the structure of data representation [29].

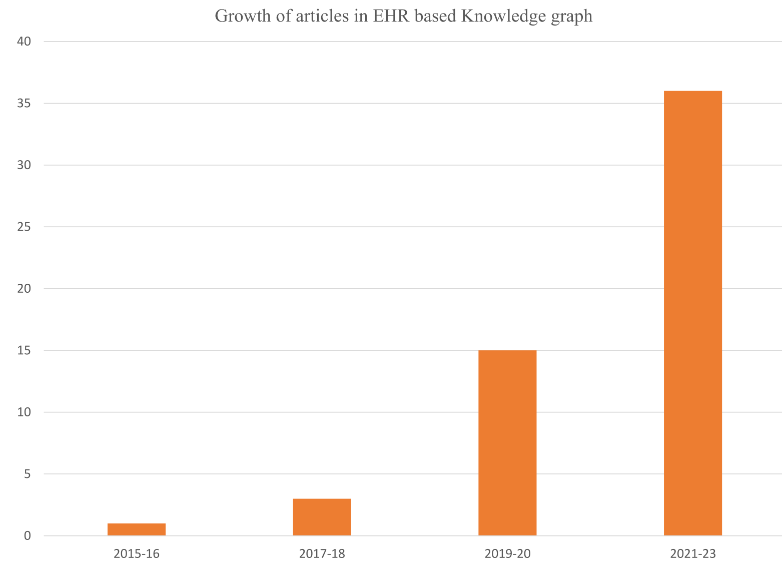

Fig. 2. Growth of knowledge graph related medical articles using EHR data: Number of articles based on the review strategy vs year.

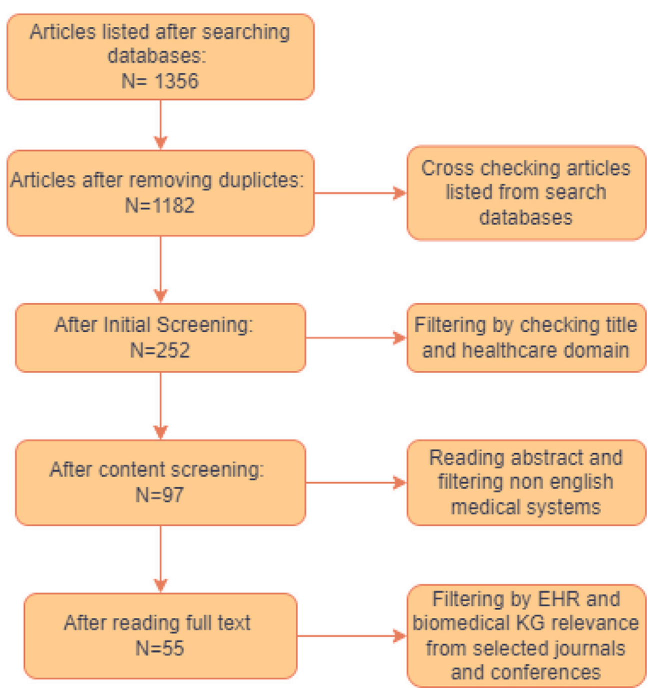

Fig. 3. Article selection criteria — PRISMA diagram.

3.3. Heterogeneity and temporality of EHR

Most EHR data stores heterogeneous medical knowledge such as demographic data, lab results, progress notes, radiology reports, procedural reports, and discharge summaries [30]. Incorporating the medical events over a period of time along with cross-departmental information in medical knowledge graph construction generates a positive impact on predictive and clinical decision support tasks. Every EHR data is a longitudinal patient record representing medical concepts of multiple sequential patient visits [31]. The associated data annotations need to handle temporal changes in medical ontology versions for predicting evolutionary relations. Constructing a historical knowledge graph combining different ontology versions into a single knowledge graph [32] assisted in maintaining semantic annotations during information retrieval from medical data. An effective strategy for improving link prediction accuracy in longitudinal data is temporal knowledge graph completion. It captures the underlying temporal dynamics of the knowledge graph from the entities and relationships for better link prediction by integrating time stamps in the model [33].

3.4. Semantic complexity of medical data

Semantic complexity in EHR data refers to the complexity underlying the clinical notes in the form of descriptions of symptoms, clinical concepts, drug mentions, and therapeutics which is complex and ambiguous. Semantic knowledge representation of the aggregated and interlinked entities facilitates enhanced utilization of contextual knowledge for better interoperability, redundancy reduction, and detection of missing entities. Complex semantic patterns, often in the form of medical conditions and terminologies in clinical notes, are the main challenge many researchers face. Shi et al. [34] introduced a framework to organize the enormous heterogeneous medical text data as a semantic health knowledge graph by integrating conceptual graphs with health data. They used textual pattern mining for knowledge retrieval and inference mechanism to prune the contextual information from complex semantic textual data. The results were not promising due to the lack of prior domain knowledge. Recently, Koshman et al. [35] proposed a method to remove repetitive semantic structures from electronic medical records quickly and language independently. They first created syntactic trees from sentences, applied word embedding algorithms to search similar subtrees, and labeled sentences into different groups.

Learning medical knowledge graphs from EHR data is complex compared to learning from structured documents like textbooks and articles. But the number of casual relations inferred by a knowledge graph learned from actual patient records is high. One advantage of generating a healthcare knowledge graph from EHR is that the post-processing required is less in general.

4. Categorical overview of EHR medical knowledge graph

To comprehend the significant aspects of EHR using MKG, we designed a complete overview that illustrates the main activities and approaches in a real-world medical knowledge graph. Fig. 4 presents the outline of the information in four main categories data sources, construction, approaches, and applications.

The primary step in any domain-specific knowledge graph construction is to gather the required knowledge from reliable sources.

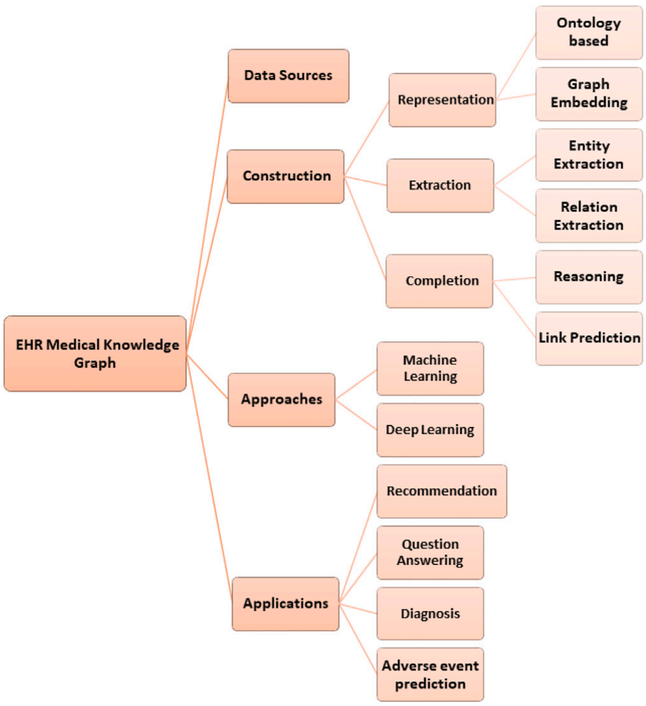

Fig. 4. Categorical overview of EHR medical knowledge graph.

In the medical knowledge graph, these reliable sources are scientific publications, medical textbooks, web articles, open databases, and real-world patient records. The data is available in structured form in the health information system and in unstructured formats, such as articles or clinical notes in EHR data. Consolidating and integrating both kinds of data build a perfect healthcare knowledge graph [36].

Real-world patient records are mainly EHR data, HIS information, and reports from laboratory and radiology [20]. For better accuracy and efficiency of medical knowledge, graph-based knowledge graph construction is the best option, along with integrating laboratory, radiology reports, and physical examination details. Applying natural language processing in EHR [37] depicts its ability to capture the syntax and semantics of clinical narratives. Utilizing freely available medical libraries like UMLS, MESH, SNOMED CT, etc., and open medical databases such as DrugBank also opens various possibilities [38].

4.1. Representation

Representation of EHR data is one major challenge in EHR-based medical informatics due to its heterogeneity, temporality, and knowledge intensity [39]. A common approach is the representation of facts in triplet format with two entities and a relation connecting these entities [40]. Extending the general triple format to a quadruple format [20] includes relevant additional fields related to a disease or patient, such as time stamps of significant events.

4.1.1. Medical ontology framework

Different formats in EHR records create data exchange issues among organizational systems. Constructing medical ontology models help in organizing and integrating knowledge from various sources to achieve interoperability. Using linked biomedical ontologies effectively incorporate syntax, semantics, relation extraction, and informative inference in a biomedical ontology framework [41]. Information exchange among health institutions needs to follow common EHR standards. Knowledge representation is a critical issue that ensures semantic health information exchange among EHR standards. Existing ontology frameworks fail to provide full semantic interoperability to achieve intelligent clinical reasoning from EHR data [42].

4.1.2. Knowledge graph embedding for healthcare

Knowledge graph embeddings are data representation techniques used in knowledge graphs to convert the graph into a low-dimensional vector format. Graph embeddings learn entities and relations in the form of distributed embeddings and preserve the semantic information between the entities in the knowledge graph to calculate entity similarity [43,44]. The embedding techniques are generally classified as translational distance models, semantic matching models, and neural network-based models [45,46]. Translational models use distance-based score functions, and semantic matching models use similarity-based score functions to calculate the similarity between entities and relationships to build the embeddings [45]. The validity of the triplets formed is measured using semantic or distance-based scoring functions. Finally, the facts representations are encoded with a suitable neural model after including additional information, such as descriptions, entity attributes, entity types, and relation paths, to enrich the knowledge of the model [3]. Translating embedding models representing entities and relations as low dimensional vectors are TransE [47], and its variants TransH [48], TransR [49], TransD [50], and more. Semantic matching models that capture the underlying semantics of triples are RESCAL [51], DistMult [52], ComplEx [53], and RotatE [54]. Several other models that used convolutional neural networks for feature analysis of triples are ConvE [55] and ConKB [56]. This review's discussion of knowledge graph embeddings will focus on clinical domain-based research only.

For biomedical natural language applications, word embedding trained on EHR and medical literature extracts semantically more relevant terms successfully [57]. Word embeddings for machine learning tasks in the clinical domain are domain-specific versions of BERT [58] such as BioBERT [59], ClinicalBERT [60] and SciBERT [61,62]. Leveraging knowledge graph embedding to clinical knowledge graph representation has shown notable results in clinical pharmacy applications such as deep learning-based Drug discovery [63,64] and drug-drug interaction prediction [65]. Also, node embeddings are used to test various link prediction methods in biomedical knowledge graphs [66]. A node2vec [67] algorithm was used for generating node embedding from a drug reaction-based knowledge graph [68]. The work conducted case studies on Covid 19 drugs and drugs causing liver injury.

A concept embedding learned from SNOMED-CT using graph-based representation learning gave promising results in link prediction and patient state prediction in healthcare analytics [69]. A joint embedding-based representation model with semantic medical entity descriptions and triples measured patient similarity leveraging deep learning [70]. The work used the MIMIC III dataset with the Siamese CNN model gave better results in patient clustering. A recursive neural knowledge network-based representation learning framework for the multi-relational graph used a Chinese EMR dataset to embed the details of patient knowledge in multi-disease diagnosis [71]. Recently, knowledge embedding technology used in multimodal reasoning for specific disease knowledge graph. The multimodal reasoning helped in improving entity prediction and evaluation [29]. Still, knowledge graph embeddings are in the early development stage in the clinical domain.

4.2. Extraction

An entity in medical texts is medical concepts like diseases, symptoms, and medications, often describing patient details. An EHR record of a patient has several associated attributes, such as demographic details, test results, lab reports, and previous history. Medical events are another entity that describes a patient's medical condition over time and is often associated with a temporal feature. Relations are facts that connect various concepts and events related to a patient. In an unstructured text like clinical notes and discharge summary in an EHR, extracting concepts, events, and relationships is the primary step in knowledge graph building.

4.2.1. Entity extraction

In knowledge graph construction identifying entities for extracting them is an early-stage action. A health knowledge graph entity is medical concepts such as diseases, symptoms, treatments, and drugs. Entity identification should consider concepts, events, diseases already listed in the vocabulary, and new entities for better accuracy. Entity extraction decides the quality of medical knowledge graph building and subsequent applications. Medical entity extraction facilitates accurate concept extraction of medical terms from health records using natural language processing, machine learning, and deep learning [72]. Concepts identified are generally diseases listed in ICD-10 diagnostic codes and symptoms mentioned in unstructured notes. Concept identification requires their common name, aliases, and acronyms for accurate extraction. Google Health Knowledge Graph (GHKG) and UMLS are the standard references utilized by researchers for this purpose. For direct concept extraction, string searching, and for extracting positive mentions, negation detection tools can be incorporated [21]. With the advent of neural networks, the Bidirectional Long Term Short Term Memory Networks (Bi-LSTM) model with Conditional Random Fields (CRF) [73] and with pattern matching rules integrated is the state-of-the-art approach for clinical named entity recognition [20].

4.2.2. Relation extraction

Relation extraction deals with finding the conceptual or semantic relationships between the extracted and disambiguated entities [36]. The connections established are unidirectional or bidirectional from triplets or quadruplets between the entities. Earlier, to identify relations from clinical records, a comparison of knowledge-based features with distance and word features is performed in BERT-based models [74]. The results identified that knowledge-based features improve limited relations, whereas sentence embedding feature improves result in many features. Recent research performs clinical relation extraction analysis with various methods of domain data integration using the BERT model [75]. UMLS medical knowledge base integration with BERT infers triplets with better accuracy compared to other clinical relation extraction techniques [75]. A comparative study of transformer architectures on clinical relation extraction used BERT, ROBERTa, and XLNET. They analyzed the performance of these models on relation extraction classification and on handling cross-sentence relations. The performance of ROBERTa and XLNET was better than BERT in relation extraction task [76]. In a real-world knowledge graph, missing relationships often play a crucial role in clinical diagnostic reasoning. Extraction of new and meaningful relations from patient medical history using real-world clinical knowledge graphs is necessary for reasoning diagnosis prediction [77]. Recently, a multimodal approach developed a disease relation extraction method to complete the missing information in a disease knowledge graph. The model fused knowledge graph embeddings with deep learning models such as SciBERT and graph neural networks for disease–disease relation extraction [28].

The advances in language models help real-world relation extraction by fine-tuning GPT-2. Data Augmented Relation Extraction (DARE) augments training data by GPT-2 and then combines the trained dataset with BERT for relation extraction [78]. MedGPT [79] showed notable performance in capturing relational associations from medical disorders with better predictions for differential diagnosis. They offer modest explainability on real-world EHR data. The work focused on patient history to predict future diagnoses using entity extraction tools. A study on depression relation extraction from tweets to analyze the association between cannabis and depression selected language models GPT3 or BERT to give state-of-the-art results in relation extraction [80]. The work efficiently extracted real-world relations when domain knowledge is infused with language models trained on large datasets. The model effectively addressed the complex relation extraction from free text clinical notes. Zeroshot information extraction for a question-answering task used the power of ChatGPT in entity extraction, relation extraction, and event extraction using a Chat Information Extraction framework [81]. Information extraction in the medical domain highly needs such models with impressive performance.

Inferring causal relations between medical concepts and symptoms from EHR data is essential for completing a medical knowledge graph. A model trained on UMLS and SemMedDB utilized semantic pattern-based features in machine learning algorithms for causal relation prediction [82]. Extracting a semantic relation expressed in clinical text as a cause–effect relation often forms the base of knowledge construction [83]. The medical events that occur together and one after another fall into this category and are highly significant in predicting diseases. A health causality extraction framework used dynamic probability analysis and created causal probability relations inside the knowledge for causal relation extraction [84]. Although the model has got some sort of explainability, it needs to be improved for adapting deep learning and graph-based algorithms. The model came as effective in predicting treatment relations from biomedical entities. Recently a causal knowledge graph based on EHR data was introduced by extracting causal relation triples. They used a graph pruning method with co-occurrence and causality ratio to obtain better accuracy [85].

4.3. Completion

A knowledge graph completion automatically fills in missing information in a knowledge graph using link prediction and knowledge graph reasoning. In this phase, a knowledge graph uses the existing entities and relationships to uncover hidden relations and infer new facts that extend the graph. A knowledge graph completion requires accuracy and data quality and uses existing triples to resolve the incompleteness of the knowledge graph [3,86]. A pre-trained model KG-BERT proposed a framework to accept entity and relationship triples in the BERT model and compute score function for relation prediction and link prediction facts [87]. A path-based reasoning approach used in knowledge graph completion utilizes a multi-hop path to extract knowledge from inner nodes, which are not extracted by direct paths. The extracted information needs noise removal and an attention-based mechanism to enable entity and relation prediction [88]. Knowledge representing multi relations is capable of inferring relations in a better manner. The feature is better utilized for acquiring a node's local information by considering the entity node's adjacency. An attention mechanism added to knowledge completion helps acquire local information from entity nodes [46]. Recently unified framework FTL-LM combined a path-based method for topology contexts learning and a variational expectation–maximization (EM) algorithm for soft logical rule distilling gave state-of-the-art results in knowledge graph completion [89]. In a medical knowledge graph, word embeddings are used to predict the RDF triples for entity completion problems. The word embeddings trained on medical big data provided prior knowledge and filtered semantic relations to improve the embedding features of the relationships from sample data [90]. Recent research uses medical knowledge graph completion for detecting medication risk by extracting medication patterns and historical prescriptions. The degree of risk level is predicted by extracting the off-label drugs and also detects the risk of new drugs using a knowledge graph completion task [91].

4.3.1. Link prediction in EHR

The information missing from health records results in missing associations between medical concepts in a medical knowledge graph. Inferring new relationships from the EMR knowledge graph realizes predicting causal relationships utilizing semantic similarity analysis and temporal metrics [92,93]. The crucial task in a medical knowledge graph is accurately capturing the relationship between diagnosis and treatment. Noise in the data set and low accuracy in capturing semantic similarity and temporal metrics often leads to incorrect link predictions.

Pretrained language models efficiently learn rich, contextual semantic representations for natural language processing tasks, including

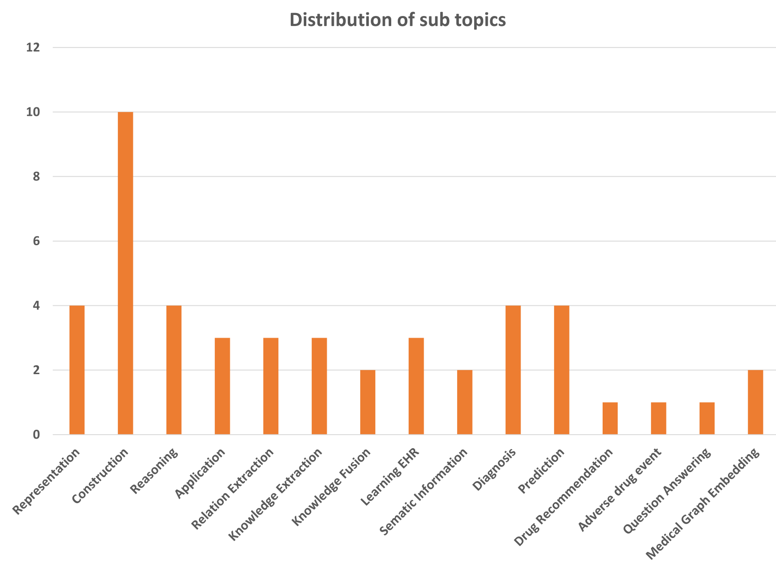

Fig. 5. Distribution of sub-topics in medical knowledge graph articles.

knowledge graph completion. BERT model with an attention mechanism efficiently captures textual semantics from entity pairs and predicts their relationship [94]. They resolved the symbolic reasoning issue by enhancing the path reasoning algorithm. The research efficiently predicts missing relationships for medical knowledge graph completion in the Chinese symbol knowledge graph. Extending the work beyond the Chinese knowledge graph can advent promising results in medical knowledge graph completion.

4.3.2. Reasoning in EHR

Domain-specific knowledge reasoning to infer missing information and filtering information improves the accuracy of clinical decision support systems. Rule-based reasoning, logic reasoning, neural networks, and embedding-based models are the most common techniques [3,86]. Knowledge graph completion uses translational and semantic matching algorithms for applications including drug repurposing for Covid 19 from a literature knowledge graph [13]. Recent advances utilize attention mechanisms in semantic and hierarchical filtering to improve transductive and inductive reasoning in complex, real-world medical knowledge graphs [95].

Fig. 5 depicts the distribution of subtasks and activities in the literature. The most seen sub-topic is the knowledge graph construction with diagnosis and prediction applications, and the least explored by the research community is medical knowledge graph completion and inference with real-world medical data.

4.4. Challenges

In healthcare, both ontology models and knowledge graph embeddings need improved methods for incorporating patient-oriented knowledge representation. The demand for an extensive data set is often a limitation for embedding models to ensure better learning and results. Dynamic graph embeddings that support online learning are a growing field gaining much attention nowadays [96]. This topic remains an unexplored area in real-world medical knowledge graphs.

4.4.1. Entity linking

Entity linking is essential as most medical entities have synonyms, and the same entity is mentioned by different practitioners using different terms. Hence linking every entity to a standard entity set is necessary [29]. Entity disambiguation occurs after the recognition of entities by classifying them into similar categories, and these results are used later in the entity resolution phase [36]. A novel approach to medical entity disambiguation from text snippets was implemented using Graph Neural Networks(GNN) and later applying optimization techniques to improve the results [97]. The work suffered multiple entity linking errors due to unstructured text snippets and multiple relationships associated with entities.

4.4.2. Relation linking

Medical knowledge graph researchers rarely discussed relation linking. But there exist synonymies in relations, too, and linking to a standard set of relations can make drastic improvements in relation extraction. However, relation normalization requires a standard relation dictionary for relation mapping, especially for commonly occurring relations [29]. Researchers must extend the study towards rarely arising or newly inferring connections applicable to link prediction.

4.4.3. Conflict resolution

The presence of noisy and uncertain data in real-world knowledge graphs leads to inconsistent statements at different points in time. Knowledge inference should use temporal inference rules and consistency-checking constraints for identifying conflicting facts [98]. One major challenge in clinical notes during relation extraction is class imbalance and identifying negative and positive instances. Addressing inter-sentence relations often needs multi-hop relation extraction. Identifying and extracting temporal instances and the presence of multiple disorders from clinical notes is more challenging. An attention mechanism-based approach with the RNN model solved contained relations from clinical notes considering the interrelationship between sentences [99,100]. Deep learning using a knowledge graph evolution algorithm combined time attributes, semantic embedding representations, and graph structure features for conflict resolution in dynamic knowledge graphs [101]. The framework successfully identifies inconsistent facts from the knowledge graph. Recently to eliminate the temporal conflict facts from the growing knowledge graph, a conflict resolution algorithm used the temporal knowledge graph embedding method [102]. After eliminating the temporal conflict facts, they applied link prediction to complete the knowledge graph.

4.4.4. Knowledge fusion

Knowledge fusion helps to combine all data of a specific entity from multiple sections and unify all the collected information into the base entity. Fusion is highly required when there are different data sources for knowledge graphs. The knowledge support provided by multiple-source knowledge fusion is highly needed in intelligent data processing

Table 1
Machine learning based approaches.

Study	Approach/Model	Subtask/Application	Data set	MKG-Phase
[103]	Logistic regression, naïve Bayes, random forest + Ontology	Diagnosis	EHR	Reasoning
[104]	Semantic rules+ Ontology Structure	Relation Extraction	EMR	Construction
[21]	(Logistic regression, naïve Bayes, Noisy OR) + MLE	Knowledge extraction	EHR	Construction+ Inference
[10]	Semi-auto constructed Framework + KG Tools	Knowledge extraction	Multi data + EHR	Construction

and extracting hidden knowledge patterns from knowledge graphs. Mapping of differently expressed entities to a single original entity is a significant task known as entity resolution [27]. Knowledge integration from web sources and medical literature helps enrich the graph with recent advances and helps to enhance the information in EHR [22]. Fusing medical knowledge database UMLS with pre-trained models such as BERT improves relation extraction [75]. Integrating clinical concepts from multiple institutes is useful for establishing cooperation and shareability of clinical knowledge. Learned embedding vectors of EHR data on clinical codes were used for effective knowledge extraction by using a regression pipeline [105]. This helped in identifying patient-specific phenotypes in a better way compared to a single institutional knowledge graph. To enable similar relation extraction from multi-institutional data, MIKIGI [106] algorithm utilized co-occurrence of clinical data and semantic embeddings. It is also observed that integrating EHR from multiple institutions often results in less quality, and the MKG constructed results with inconsistency and mismatch between diagnosis results. Including a data-cleaning framework for cleaning EHR data from multiple sources with medical knowledge graph construction ensures effectiveness and accuracy in diagnostic results [107]. Moreover, Deep collective matrix factorization (DCMF) applied in integrating multiple data source EHR with knowledge graph delivered efficient patient representations for clinical decision support [108].

Knowledge graph completion in real-world medical knowledge graphs demands higher accuracy than generic ones. The algorithms must accurately infer contextual and diagnostic relations along with semantic, causal, and temporal relations.

5. Approaches in learning medical knowledge graph

The computational approaches in medical knowledge graphs witnessed rapid growth in deep learning techniques over the past few years. Earlier ontology frameworks, and rule-based logic reasoning, were used along with machine learning algorithms. This section provides an overview of the significant machine learning and deep learning approaches researchers used to represent, learn and complete EHR-based biomedical knowledge graphs.

5.1. Machine learning approaches

As mentioned, in addition to machine learning methods, statistical analysis [109] and ontology structures [34,43] are used to extract and organize information in medical knowledge graph construction. Statistical algorithms aid in identifying synonyms, identifying the relationship between concepts, and categorizing concepts, whereas ontology structures help to organize concepts and relationships in a structured format to query and navigate. Table 1 lists machine learning-oriented studies that used EHR as the data source in medical knowledge graphs.

Early work on the biomedical knowledge graph Knowlife [40] combined pattern analysis and consistent reasoning for extracting valuable and accurate relations. They used pattern analysis to form the entity and relation triplets from seed facts and applied logical consistency to remove unwanted facts. They collected facts from different text genres like scientific publications, online web portals, and medical dictionaries, and the final information extraction pipeline showed numerous incorrect facts.

Joint learning of disease and symptoms from EHR data in a medical knowledge graph needs accurate extraction of positive mentions from the clinical descriptions [21]. The knowledge graph construction used three probabilistic models logistic regression, naïve Bayes, and a noisy OR model with maximum likelihood estimation as parameter learners. The constructed graph is evaluated against a manually curated GHKG and an expert to verify that all known relations are derived. However, the predicted causality accuracy for all associations depended on unknown confounding factors.

To analyze the disease and symptoms co-occurrence derived from EHR data while learning knowledge graph, researchers used logistic regression, naïve Bayes, and random forest algorithms. They also included demographic details of patients, such as age, gender, and nonlinear functions, in the analysis to improve accuracy [103]. The evaluation of the learned health knowledge graph using the GHKG found that demographic data and correlated diseases impact the results more positively than the sample data size and model complexity. The inferred causal relations and the size of patient data used found as limited.

An end-to-end knowledge graph framework from multiple data sources, including an EHR dataset, integrated the clinician's prior knowledge with an extensible mechanism to add new diseases [10]. Several factors, such as diversity in knowledge sources, scalability of the knowledge graph, and evaluation techniques, are essential while designing an end-to-end knowledge graph framework. The evaluation strategies used to assess clinician's efforts and methods for interpreting physician annotations in such frameworks should be error-free.

Utilization of unused and hidden information from longitudinal EHR data, especially historical details of patients and cross-departmental information, is essential in solving data fragmentation issues. A top-level ontology structure as the base of the knowledge graph's semantic representation and design addressed the problem, often unnoticed by physicians as they might focus on the current disease or department only [104]. The explainable result offered by semantic rule-based reasoning helps clinicians to accept or reject cross-departmental patient data. There are limitations in handling unstructured data and medical images. Also, adapting cross-institutional data enhancement is a solution for data fragmentation issues.

5.2. Deep learning approaches

Deep learning techniques are currently the hottest and most effective techniques used by medical knowledge graph researchers. The deep learning models used to automate the process of building and maintaining medical knowledge graphs from EHR records now include transformer-based BERT models and variants of GNN. The state-of-the-art model in named entity recognition, Bi-LSTM with CRF layer, was adopted in EMR-based medical knowledge graph construction in 2019 by Li et al. [20]. They used the Bi-LSTM model with medical KG graph embedding to predict medicines from disease information [20]. The work transferred medical knowledge to the Bi-LSTM model after learning the medical knowledge graph embedding using PrTransH and resulted in a 2 percent rise in the Jaccard score in the medicine prediction task.

The language understanding capabilities of pre-trained language models assisted in capturing rich semantic patterns from free text and hence combined with knowledge graph and knowledge graph embeddings to enhance knowledge representation. A framework for knowledge graph completion, Knowledge Graph BERT (KG-BERT), used

Table 2
Deep learning based approaches.

Study	Approach/Model	Subtask/Application	Data set	MKG-Phase
[20]	Bi-LSTM+ MedicalKGEmbeddingg	Medicine Prediction	UMLS	Representation + Reasoning
[87]	KG + BERT	entity prediction	UMLS	Representation + Reasoning
[75]	ClinicalBERT + UMLSKGE	Relation extraction	UMLS	Representation + Reasoning
[110]	BERT+CRF	Entity prediction +relation extraction	MIMIC III	Representation + Construction + Reasoning
[111]	GCN	Diagnosis prediction	MIMIC III, Private EHR	Completion
[31]	GAT +RNN	Disease prediction	MIMIC III, CLAIM	Completion
[95]	GAT+ Gated Tree	Relation prediction	EMRNet	Completion

entities and relations of UMLS to fine-tune BERT for entity prediction [87]. The link prediction results in the UMLS dataset were not promising, as the model required several text sequences for improved performance. Later, Roy et al. [75] integrated medical knowledge into BERT and ClinicalBERT models by generating UMLS knowledge graph embeddings. The fusion of UMLS knowledge graph embedding with ClinicalBERT indicated a notable performance gain in clinical relation extraction.

A generic approach to knowledge graph construction and analysis from unstructured clinical notes was proposed by A. Harnoune et al. [110] using the BERT-based model with the MIMIC III [15] data set. The knowledge graph construction used medical domain-based BERT variants with a CRF layer for entity prediction and a fully connected binary classification layer for relation extraction. The framework successfully extracted information with high accuracy for entity recognition and relation extraction tasks. However, the framework could improve relation extraction by establishing missing links between symptoms and drugs.

GNN can insert domain knowledge into the model architecture, enabling the neural model to infer new explainable relationships from the domain knowledge graph utilizing the representation scheme [112]. A graph convolutional network (GCN) based model fully exploited medical knowledge by using the spatial and temporal features underlying EHR to improve the accuracy of diagnosis prediction [111]. The work outperformed all previous RNN-based attention mechanism models that partially used medical knowledge as external integrated information in diagnosis prediction.

Graph Attention Networks (GAT) use attention mechanisms to weigh the importance of the different edges and nodes in the graph, which allow them to focus on the most relevant information when making predictions or performing other tasks. A hierarchical GAT(H-GAT) proposed an iterative self-learning strategy for knowledge graph completion and data reconstruction utilizing EHR data and medical knowledge graph [31]. The model uses node embeddings generated by the attention network with an RNN to fine-tune the disease prediction based on the learned knowledge graph and updates the graph with missing links. Gated Tree-based Graph Attention Networks (GTGAT) efficiently handle complex datasets by combining GAT with the gated-recursive mechanism [95]. GTGAT refines nearby relations using an attention mechanism and prevents information loss using gated units and the Huffman tree. Table 2 shows the most used deep learning models in the EHR-oriented medical knowledge graph research.

Overall, GATs and their variants allow the modeling of complex relationships between entities in the knowledge graph and enable the network to focus on the most relevant information from the most pertinent neighbor nodes when making predictions.

5.3. Privacy preserving approaches

While learning medical knowledge graphs from EHR data, it is important to preserve patient privacy by protecting sensitive patient information such as medical history, diagnoses, and treatments. Medical data sharing and analysis require better privacy-preserving policies to enhance collaborative research among medical institutions for better health services and for knowledge graph integration [113].

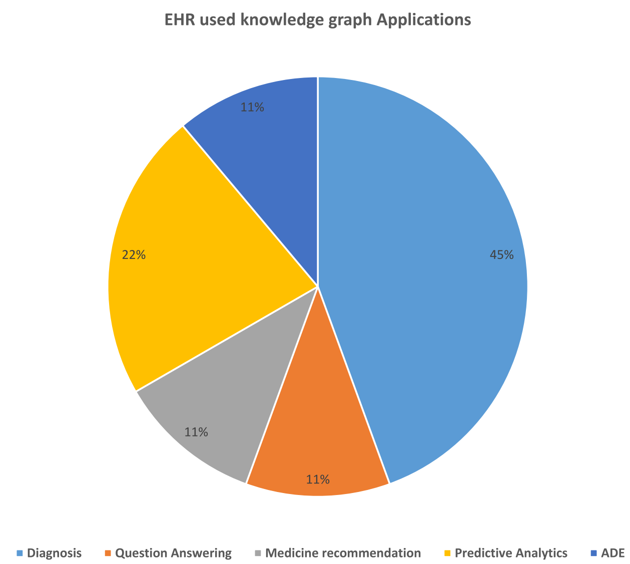

Fig. 6. Utilization of EHR-based knowledge graph in healthcare applications.

Co-occurrence counts from clinical notes and co-frequencies from annotations were applied to overcome the data sharing limitation of EHR data [109]. In addition, several other privacy-preserving approaches are used in healthcare data, including anonymization, differential privacy, federated learning, cryptographic techniques, secure multi-party computation, and homomorphic encryption [114,115]. The major obstacles to privacy preservation in clinical information are the massive growth of data, cross-institutional data interoperability, and security and privacy of patient records [113].

Medical blockchain is a concept introduced to bring together correctness and privacy while sharing EHR data. They used blockchain to preserve patient privacy and a smart contract to retain the integrity of patient data [116]

6. Major applications of EHR-based MKG

As a powerful tool for representing and organizing complex and real-world medical knowledge, an EHR-curated knowledge graph can support several clinical decision support applications. This section explores the applications where real-world medical knowledge graphs exhibited promising outputs. Fig. 6 shows the percentage distribution of applications used in the real-world medical knowledge graph.

6.1. Medication recommendation

The scope of using a knowledge graph system with EHR data in health care recommendation systems is increasing as the knowledge graph stores patient details, diseases, medications, and diagnoses details. A heterogeneous medical knowledge graph built from MIMIC III, ICD-9, and Drug Bank for safe medicine recommendations used graph embeddings learned with patients, diseases, and medicines. The framework recommends new and safe therapy to the patient by inferring new links [117].

Table 3
Applications of medical knowledge graphs.

Study	Year	Application	Purpose
[117]	2021	Medicine Recommendation	Recommends safe medicine by inferring links
[119]	2021	Diagnosis, Prediction	Predicts future patient diagnoses
[120]	2017	Adverse Drug reaction	Predicts unknown drug reaction
[22]	2020	Diagnosis	Extended diagnosis with graph completion
[31]	2021	Prediction	Future diagnosis and Multiple prediction tasks
[26]	2021	Prediction	Diagnosis prediction with knowledge fusion
[118]	2021	Question Answering	Answers patient-specific queries

6.2. Disease diagnosis

The primary and most common application of medical knowledge graphs is the classification of diseases from various data sources like medical images, EHR data, and clinical notes. Disease concepts and their interrelations represented as nodes and links help to learn semantic features from input data modeled using CNN, GCN, word embedding, pre-trained networks, and attention mechanisms [118]. A deep learning model with an attention mechanism on the MIMIC III data set used integrated multimodal input features, including patient demographics, diagnosis codes, clinical features, and disease ontology, to predict future patient diagnoses. The model obtained performance improvement compared to previous works, yet including demographic domain knowledge as prior knowledge in the knowledge graph might lead to excellent results [119] (see Table 3).

6.3. Adverse drug event detection

A machine learning-based approach used a prediction algorithm and inferred unknown drug reactions as unknown edges, taking a graph as input [120]. The model also predicts the reasons for adverse drug reactions utilizing a knowledge graph constructed with nodes as drugs, proteins, targets, and indicators of adverse reactions. A knowledge graph with an expandable framework based on EHR data facilitates the integration of new disease-specific information into the existing graph structure. An incremental expansion of the knowledge graph towards graph completion improves the diagnosis prediction results with new possibilities [22].

6.4. Predictive analytics

Predictive analytics in healthcare is becoming increasingly widespread as more healthcare organizations recognize its potential to improve patient outcomes and reduce costs. The self-supervised learning strategy for knowledge graph completion is well suited to utilize longitudinal EHR data for multiple predictive tasks, including future diagnosis prediction [31]. For improved predictions, applying fusion methods that integrate time series, time-invariant, and unstructured clinical notes resulted in superior performance [26]. Predictive analytics in an EHR-based knowledge graph aims to support healthcare providers in making more informed decisions by uncovering hidden relationships and patterns in patient data.

6.5. Question answering

Question answering in the medical domain is vital as a decision support tool. A domain-specific and EMR-based knowledge graph provides intelligent and reasonable answers to patient-specific queries compared to a traditional schema-based question-answer system [30]. Understanding the semantics and context of the data to be retrieved for information retrieval demanding domain-specific information is highly important. It applies to multiple modalities in the healthcare domain, and previous knowledge support is essential for better decision-making, diagnosis, and recommendation accuracy. Integrating a knowledge graph facilitates access to multimodal images without losing the communicated semantic features in medical image retrieval [118].

The thrust areas where medical knowledge graphs need to deploy further are the medical education sector and medical summary generation. EHR-based knowledge graphs can benefit medical education by making it interactive, allowing students to explore different aspects of the topic, and making the learning process more engaging [121]. Using knowledge graphs in medical training systems empowers disease diagnosis feature extraction and addresses insufficient data and interpretability issues [122]. A medical domain knowledge graph-based web environment offers a wide range of potential applications suitable for medical students in addressing clinical case analysis and builds EHR-based visualizations for better understanding [123]. Also, medical knowledge can reduce the need for manual summarization by mapping out the relationships between different concepts and entities in the graph and providing a compact and comprehensive overview of large amounts of medical information.

7. Discussion and future scope

A health knowledge graph constructed from patient records encompasses knowledge representation, knowledge extraction, and knowledge-based reasoning. A representation of medical entities and relations as triplets or quadruples stores the facts in a structured format in the graph. Most studies approached medical knowledge graph construction by extracting named entities and their relations from unstructured data. A detailed analysis of existing literature lacks a solution to address discrepancies in EHR data in representation. For example, handling unlabeled, missing, and fragmented data and linking cross-departmental data. Addressing these problems at the initial stage enrich the graph with robust information, improving the accuracy of all downstream activities.

The knowledge extraction phase incorporates entity extraction, relation extraction, and further disambiguation steps, including knowledge fusion from multiple data sources and resolving entities and relations. Pure entity extraction in terms of accuracy and scalability, identifying and extracting semantic complexities and interoperability, and avoiding semantic duplication are still challenges. At the same time, building a concrete knowledge foundation demands the utilization of multiple data sources. A framework supporting knowledge graph integration from various data sources is essential for large-scale medical knowledge graphs.

A knowledge graph completion is the final and the least explored in EHR-based medical knowledge graphs, which opens the door to numerous downstream activities. Knowledge graph embeddings and neural network-based models aim to infer new and vital facts from the medical knowledge graph to predict missing relationships. However, temporal dynamics of events, capturing of semantic similarity, and identification of causal relations limit link prediction accuracy.

7.1. Future research scope

7.1.1. Patient-based embeddings

Medical knowledge graphs built on EHR data demand frameworks that accept multimodal and heterogeneous EHR data and integrate patient temporal and demographic attributes in learning the model. A high-dimensional patient-oriented representation of EHR knowledge assists in retrieving the critical information required in precision medicine and clinical decision support. Learning patient-level representation by data-driven approaches from EHR and clinical notes proves patient embeddings show better results in patient future disease prediction and diagnosis prediction in patient cohorts [124] [125]. Deep learning frameworks for learning patient similarity from EHR along with temporal properties [126], and incorporating knowledge graph embedding and medical concept embedding with patient similarity learning addressed personalized patient care more accurately and precisely [70]. Recently hybrid data and knowledge-driven frameworks incorporating patient embeddings to knowledge graph embedding models [127] and temporal patient embedding with knowledge graph [128] created patient-level knowledge graph representations. Yet the mapping of clinical features from EHR data is not completely utilized by these models resulting in low-quality of patient representations. A hybrid data and knowledge graph embedding model with high-quality patient representation is indispensable for better patient outcomes.

7.1.2. Knowledge fusion with relation linking

Semantic interoperability in the health knowledge graph is associated with combining patient details from multiple institutions while integrating two or more medical knowledge graphs. Knowledge fusion that extends entity linking with contextual relation linking is an open research challenge that can improve the results of tasks like diagnosis prediction and virtual patient generation. Lack of standard mapping library of relations often leads to a manual mapping [29]. Another way is to use a hierarchical relation structure clustering similar relations forming relation clusters [129]. However, this method is not suitable for large medical knowledge graphs [24,29,130]. We suggest combining knowledge graph embedding with suitable deep learning models for relation linking in a multi-source large medical knowledge graph with millions of relations [28].

7.1.3. Dynamic update in real-world medical knowledge graph

Knowledge graph completion in the healthcare domain considers static medical knowledge graphs only. A real-world patient record-based medical knowledge graph requires adding and removing concepts and facts with time. Dynamic updates of real-world health knowledge graphs exclusively require accurate and interpretable methods of knowledge reasoning based on temporal dynamics in applications like medical report generation and recommendation systems. A machine comprehension task uses a graph neural network-based method to dynamically build and update graphs to capture the information update from an interactive agent [131]. A knowledge graph embedding update mechanism used in a real-world knowledge graph dynamically updates the graph using optimal local information and avoids full retraining of the embedding [132]. A knowledge graph embedding update can be adapted to the healthcare domain for promising results in all downstream activities.

In addition, bias in healthcare data leads to unequal access to health services and incorrect or inappropriate medical decisions. To ensure the accuracy and fairness of healthcare data, it is essential to consider and address potential sources of bias in all data collection, analysis, and interpretation stages. Hence the collaboration of physicians and AI engineers is crucial in developing AI systems like medical knowledge graphs in health informatics. Doctors bring their medical expertise and understanding of clinical processes, patient needs, and the healthcare system, while AI engineers have data analysis and machine learning skills. It ensures that the technology is grounded in medical knowledge and reflects the needs of patients and the healthcare system.

8. Conclusion

Medical knowledge graphs realize a strong domain-specific prior knowledge support for various health informatics applications like clinical decision support, diagnosis prediction, medication recommendation, precision medicine, and virtual patient generation. An EHR-derived knowledge graph augments the medical knowledge graph with real-world patient data, which is essential for improving the accuracy and efficiency of knowledge graph-driven critical care applications. We conducted a detailed study on medical knowledge graph construction focusing on EHR data, and our research contributions are following. Our study identified the challenges raised by the complex, heterogeneous, temporal, and incomplete nature of EHR in knowledge graph construction at extraction, representation, and completion level. We provided detailed insight into the need to address knowledge graph integration, knowledge alignment, and the growing nature of data in knowledge graphs. The recent trends in medical knowledge graphs focus on incorporating deep learning advances in EHR-based MKG at the representation level using knowledge graph embeddings. Future research should be devoted to solving research gaps in medical knowledge graph completion due to diseases' dynamic and extending nature. Further studies should aim to provide explainability in the medical knowledge graph leading to more accurate predictions and better outcomes for patients to increase trust in technology-assisted medical decision-making.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D.N. Nicholson, C.S. Greene, Constructing knowledge graphs and their biomedical applications, Comput. Struct. Biotechnol. J. 18 (2020) 1414–1428.
[2] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G.D. Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier, A.-C.N. Ngomo, A. Polleres, S.M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Comput. Surv. 54 (4) (2021) http://dx.doi.org/10.1145/3447772.
[3] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. Yu, A survey on knowledge graphs: Representation, Acquis. Appl. (2021).
[4] M. Nickel, K. Murphy, V. Tresp, E. Gabrilovich, A review of relational machine learning for knowledge graphs, Proc. IEEE 104 (1) (2016) 11–33, http://dx.doi.org/10.1109/JPROC.2015.2483592.
[5] F. Zhang, N.J. Yuan, D. Lian, X. Xie, W.-Y. Ma, Collaborative knowledge base embedding for recommender systems, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 353–362.
[6] X. Wang, X. He, Y. Cao, M. Liu, T.-S. Chua, Kgat: Knowledge graph attention network for recommendation, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 950–958.
[7] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, Y. Choi, COMET: Commonsense transformers for automatic knowledge graph construction, 2019, arXiv preprint arXiv:1906.05317.
[8] G. Zhu, C.A. Iglesias, Exploiting semantic similarity for named entity disambiguation in knowledge graphs, Expert Syst. Appl. 101 (2018) 8–24.
[9] J. Lin, Y. Zhao, W. Huang, C. Liu, H. Pu, Domain knowledge graph-based research progress of knowledge representation, Neural Comput. Appl. 33 (2021) 681–690.
[10] Y. Zhang, M. Sheng, R. Zhou, Y. Wang, G. Han, H. Zhang, C. Xing, J. Dong, HKGB: An inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians' expertise incorporated, Inf. Process. Manage. 57 (6) (2020) 102324.
[11] P. Richardson, I. Griffin, C. Tucker, D. Smith, O. Oechsle, A. Phelan, M. Rawling, E. Savory, J. Stebbing, Baricitinib as potential treatment for 2019-nCoV acute respiratory disease (vol 395, pg e30, 2020), Lancet 395 (10241) (2020) 1906.
[12] J. Stebbing, A. Phelan, I. Griffin, C. Tucker, O. Oechsle, D. Smith, P. Richardson, COVID-19: Combining antiviral and anti-inflammatory treatments, Lancet Infect. Dis. 20 (4) (2020) 400–402.
[13] R. Zhang, D. Hristovski, D. Schutte, A. Kastrin, M. Fiszman, H. Kilicoglu, Drug repurposing for COVID-19 via knowledge graph completion, J. Biomed. Inform. 115 (2021) 103696.
[14] L.L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Eide, K. Funk, R. Kinney, Z. Liu, W. Merrill, et al., Cord-19: The COVID-19 open research dataset, 2020, ArXiv.
[15] A. Johnson, T.J. Pollard, L. Shen, L.-w.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L.A. Celi, R.G. Mark, MIMIC-III, a freely accessible critical care database sci, Data 3 (160035) (2016) 10–1038.

[16] O. Bodenreider, The unified medical language system (UMLS): Integrating biomedical terminology, Nucleic Acids Res. 32 (suppl_1) (2004) D267–D270.
[17] K. Donnelly, et al., SNOMED-CT: The advanced terminology and coding system for ehealth, Stud. Health Technol. Inform. 121 (2006) 279.
[18] W.H. Organization, et al., The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research, World Health Organization, 1993.
[19] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: A comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. 34 (suppl_1) (2006) D668–D672.
[20] 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. 103 (2020) 101817.
[21] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, D. Sontag, Learning a health knowledge graph from electronic medical records, Sci. Rep. 7 (1) (2017) 1–11.
[22] J. Xie, J. Jiang, Y. Wang, Y. Guan, X. Guo, Learning an expandable EMR-based medical knowledge network to enhance clinical diagnosis, Artif. Intell. Med. 107 (2020) 101927.
[23] P.L. Teixeira, W.-Q. Wei, R.M. Cronin, H. Mo, J.P. VanHouten, R.J. Carroll, E. LaRose, L.A. Bastarache, S.T. Rosenbloom, T.L. Edwards, D.M. Roden, T.A. Lasko, R.A. Dart, A.M. Nikolai, P.L. Peissig, J.C. Denny, Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals, J. Am. Med. Inform. Assoc. 24 (1) (2016) 162–171, http://dx.doi.org/10.1093/jamia/ocw071.
[24] A. Santos, A.R. Colaço, A.B. Nielsen, L. Niu, P.E. Geyer, F. Coscia, N.J.W. Albrechtsen, F. Mundt, L.J. Jensen, M. Mann, Clinical knowledge graph integrates proteomics data into clinical decision-making, 2020, pp. 05–2020, BioRxiv.
[25] S. Henry, K. Buchan, M. Filannino, A. Stubbs, O. Uzuner, 2018 N2C2 shared task on adverse drug events and medication extraction in electronic health records, J. Am. Med. Inform. Assoc. 27 (1) (2020) 3–12.
[26] B. Yang, L. Wu, How to leverage multimodal EHR data for better medical predictions? 2021, arXiv preprint arXiv:2110.15763.
[27] X. Zhao, Y. Jia, A. Li, R. Jiang, Y. Song, Multi-source knowledge fusion: A survey, World Wide Web 23 (2020) 2567–2592.
[28] Y. Lin, K. Lu, S. Yu, T. Cai, M. Zitnik, Multimodal learning on graphs for disease relation extraction, 2022, arXiv preprint arXiv:2203.08893.
[29] C. Zhu, Z. Yang, X. Xia, N. Li, F. Zhong, L. Liu, Multimodal reasoning based on knowledge graph embedding for specific diseases, Bioinformatics 38 (8) (2022) 2235–2245.
[30] J. Park, Y. Cho, H. Lee, J. Choo, E. Choi, Knowledge graph-based question answering with electronic health records, in: Machine Learning for Healthcare Conference, PMLR, 2021, pp. 36–53.
[31] X. Xu, X. Xu, Y. Sun, X. Liu, X. Li, G. Xie, F. Wang, Predictive modeling of clinical events with mutual enhancement between longitudinal patient records and medical knowledge graph, in: 2021 IEEE International Conference on Data Mining, ICDM, IEEE, 2021, pp. 777–786.
[32] S.D. Cardoso, M. Da Silveira, C. Pruski, Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies, Knowl.-Based Syst. 194 (2020) 105508.
[33] B. Cai, Y. Xiang, L. Gao, H. Zhang, Y. Li, J. Li, Temporal knowledge graph completion: A survey, 2022, arXiv preprint arXiv:2201.08236.
[34] L. Shi, S. Li, X. Yang, J. Qi, G. Pan, B. Zhou, Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services, BioMed Res. Int. 2017 (2017).
[35] V. Koshman, A. Funkner, S. Kovalchuk, An unsupervised approach to structuring and analyzing repetitive semantic structures in free text of electronic medical records, J. Personalized Med. 12 (1) (2022) 25.
[36] B. Abu-Salih, M. AL-Qurishi, M. Alweshah, M. AL-Smadi, R. Alfayez, H. Saadeh, Healthcare knowledge graph construction: State-of-the-art, open issues, and opportunities, 2022, arXiv preprint arXiv:2207.03771.
[37] Y. Luo, W.K. Thompson, T.M. Herr, Z. Zeng, M.A. Berendsen, S.R. Jonnalagadda, M.B. Carson, J. Starren, Natural language processing for EHR-based pharmacovigilance: A structured review, Drug Saf. 40 (2017) 1075–1089.
[38] X. Wu, J. Duan, Y. Pan, M. Li, Medical knowledge graph: Data sources, construction, reasoning, and applications, Big Data Min. Anal. 6 (2) (2023) 201–217.
[39] X. Liu, H. Wang, T. He, Y. Liao, C. Jian, Recent advances in representation learning for electronic health records: A systematic review, J. Phys.: Conf. Ser. 2188 (1) (2022) 012007.
[40] P. Ernst, A. Siu, G. Weikum, Knowlife: A versatile approach for constructing a large knowledge graph for biomedical sciences, BMC Bioinformatics 16 (2015) 1–13.
[41] M. Alobaidi, K.M. Malik, M. Hussain, Automated ontology generation framework powered by linked biomedical ontologies for disease-drug domain, Comput. Methods Programs Biomed. 165 (2018) 117–128.
[42] S. Sachdeva, S. Bhalla, Using knowledge graph structures for semantic interoperability in electronic health records data exchanges, Information 13 (2) (2022) 52.
[43] B. Cheng, J. Zhang, H. Liu, M. Cai, Y. Wang, Research on medical knowledge graph for stroke, J. Healthc. Eng. 2021 (2021).
[44] M. Kartheek, G. Sajeev, Building semantic based recommender system using knowledge graph embedding, in: 2021 Sixth International Conference on Image Information Processing, ICIIP, 6, IEEE, 2021, pp. 25–29.
[45] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: A survey of approaches and applications, IEEE Trans. Knowl. Data Eng. 29 (12) (2017) 2724–2743.
[46] J. Huang, T. Lu, J. Zhu, W. Yu, T. Zhang, Multi-relational knowledge graph completion method with local information fusion, Appl. Intell. 52 (7) (2022) 7985–7994.
[47] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, Adv. Neural Inf. Process. Syst. 26 (2013).
[48] Z. Wang, J. Zhang, J. Feng, Z. Chen, Knowledge graph embedding by translating on hyperplanes, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 28, no. 1, 2014.
[49] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, Learning entity and relation embeddings for knowledge graph completion, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 29, no. 1, 2015.
[50] G. Ji, S. He, L. Xu, K. Liu, J. Zhao, Knowledge graph embedding via dynamic mapping matrix, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 687–696.
[51] M. Nickel, V. Tresp, H.-P. Kriegel, et al., A three-way model for collective learning on multi-relational data, in: Icml, Vol. 11, no. 10.5555, 2011, pp. 3104482–3104584.
[52] B. Yang, W.-t. Yih, X. He, J. Gao, L. Deng, Embedding entities and relations for learning and inference in knowledge bases, 2014, arXiv preprint arXiv:1412.6575.
[53] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, G. Bouchard, Complex embeddings for simple link prediction, in: International Conference on Machine Learning, PMLR, 2016, pp. 2071–2080.
[54] Z. Sun, Z.-H. Deng, J.-Y. Nie, J. Tang, Rotate: Knowledge graph embedding by relational rotation in complex space, 2019, arXiv preprint arXiv:1902.10197.
[55] T. Dettmers, P. Minervini, P. Stenetorp, S. Riedel, Convolutional 2D knowledge graph embeddings, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, no. 1, 2018.
[56] D.Q. Nguyen, T.D. Nguyen, D.Q. Nguyen, D. Phung, A novel embedding model for knowledge base completion based on convolutional neural network, 2017, arXiv preprint arXiv:1712.02121.
[57] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen, P. Kingsbury, H. Liu, A comparison of word embeddings for the biomedical natural language processing, J. Biomed. Inform. 87 (2018) 12–20.
[58] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, S. Fidler, Aligning books and movies: Towards story-like visual explanations by watching movies and reading books, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 19–27.
[59] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang, BioBERT: A pre-trained biomedical language representation model for biomedical text mining, Bioinformatics 36 (4) (2020) 1234–1240.
[60] E. Alsentzer, J.R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, M. McDermott, Publicly available clinical BERT embeddings, 2019, arXiv preprint arXiv:1904.03323.
[61] I. Beltagy, K. Lo, A. Cohan, SciBERT: A pretrained language model for scientific text, 2019, arXiv preprint arXiv:1903.10676.
[62] F.K. Khattak, S. Jeblee, C. Pou-Prom, M. Abdalla, C. Meaney, F. Rudzicz, A survey of word embeddings for clinical text, J. Biomed. Inform. 100 (2019) 100057.
[63] S. Sang, Z. Yang, X. Liu, L. Wang, H. Lin, J. Wang, M. Dumontier, GrEDeL: A knowledge graph embedding based method for drug discovery from biomedical literatures, Ieee Access 7 (2018) 8404–8415.
[64] S. Bonner, I.P. Barrett, C. Ye, R. Swiers, O. Engkvist, C.T. Hoyt, W.L. Hamilton, Understanding the performance of knowledge graph embeddings in drug discovery, Artif. Intell. Life Sci. 2 (2022) 100036.
[65] R. Celebi, H. Uyar, E. Yasar, O. Gumus, O. Dikenelli, M. Dumontier, Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings, BMC Bioinformatics 20 (1) (2019) 1–14.
[66] G. Crichton, Y. Guo, S. Pyysalo, A. Korhonen, Neural networks for link prediction in realistic biomedical graphs: A multi-dimensional evaluation of graph embedding-based approaches, BMC Bioinformatics 19 (1) (2018) 1–11.
[67] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks. KDD 2016: 855–864, 2016.
[68] P. Joshi, V. Masilamani, A. Mukherjee, A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network, J. Biomed. Inform. 132 (2022) 104122.
[69] K. Agarwal, T. Eftimov, R. Addanki, S. Choudhury, S. Tamang, R. Rallo, Snomed2Vec: Random walk and poincaré embeddings of a clinical knowledge base for healthcare analytics, 2019, arXiv preprint arXiv:1907.08650.
[70] Z. Lin, D. Yang, X. Yin, Patient similarity via joint embeddings of medical knowledge graph and medical entity descriptions, IEEE Access 8 (2020) 156663–156676.

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

[72] J. Qu, A review on the application of knowledge graph technology in the medical field, Sci. Program. 2022 (2022).

[73] Y. Qin, Y. Zeng, Research of clinical named entity recognition based on bi-LSTM-CRF, J. Shanghai Jiaotong Univ. (Sci.) 23 (2018) 392–397.

[74] I. Alimova, E. Tutubalina, Multiple features for clinical relation extraction: A machine learning approach, J. Biomed. Inform. 103 (2020) 103382.

[75] A. Roy, S. Pan, Incorporating medical knowledge in BERT for clinical relation extraction, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 5357–5366.

[76] X. Yang, Z. Yu, Y. Guo, J. Bian, Y. Wu, Clinical relation extraction using transformer-based models, 2021, arXiv preprint arXiv:2107.08957.

[77] V. Socrates, Extraction of diagnostic reasoning relations for clinical knowledge graphs, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2022, pp. 413–421.

[78] Y. Papanikolaou, A. Pierleoni, Dare: Data augmented relation extraction with GPT-2, 2020, arXiv preprint arXiv:2004.13845.

[79] Z. Kraljevic, A. Shek, D. Bean, R. Bendayan, J. Teo, R. Dobson, MedGPT: Medical concept prediction from clinical narratives, 2021, arXiv preprint arXiv:2107.03134.

[80] K. Roy, U. Lokala, V. Khandelwal, A. Sheth, "Is depression related to cannabis?": A knowledge-infused model for entity and relation extraction with limited supervision, 2021, arXiv preprint arXiv:2102.01222.

[81] X. Wei, X. Cui, N. Cheng, X. Wang, X. Zhang, S. Huang, P. Xie, J. Xu, Y. Chen, M. Zhang, et al., Zero-shot information extraction via chatting with ChatGPT, 2023, arXiv preprint arXiv:2302.10205.

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

[83] A. Akkasi, M.-F. Moens, Causal relationship extraction from biomedical text using deep neural models: A comprehensive survey, J. Biomed. Inform. 119 (2021) 103820.

[84] H.Q. Yu, Extracting and representing causal knowledge of health condition, in: BIRDS@ SIGIR, 2020, pp. 70–78.

[85] K. Lyu, Y. Tian, Y. Shang, T. Zhou, Z. Yang, Q. Liu, X. Yao, P. Zhang, J. Chen, J. Li, Causal knowledge graph construction and evaluation for clinical decision support of diabetic nephropathy, J. Biomed. Inform. 139 (2023) 104298.

[86] X. Chen, S. Jia, Y. Xiang, A review: Knowledge reasoning over knowledge graph, Expert Syst. Appl. 141 (2020) 112948.

[87] L. Yao, C. Mao, Y. Luo, KG-BERT: BERT for knowledge graph completion, 2019, arXiv preprint arXiv:1909.03193.

[88] Y. Shen, N. Ding, H.-T. Zheng, Y. Li, M. Yang, Modeling relation paths for knowledge graph completion, IEEE Trans. Knowl. Data Eng. 33 (11) (2020) 3607–3617.

[89] Q. Lin, R. Mao, J. Liu, F. Xu, E. Cambria, Fusing topology contexts and logical rules in language models for knowledge graph completion, Inf. Fusion 90 (2023) 253–264.

[90] M. Gao, J. Lu, F. Chen, Medical knowledge graph completion based on word embeddings, Information 13 (4) (2022) 205.

[91] Z. Lin, L. Hong, X. Cai, S. Chen, Z. Shao, Y. Huang, C. Yang, L. Chen, Risk detection of clinical medication based on knowledge graph reasoning, CCF Trans. Pervasive Comput. Interact. 5 (1) (2023) 82–97.

[92] W.V. WOENSEL, C. Armstrong, M. Rajaratnam, V. Gupta, S.S.R. Abidi, Using knowledge graphs to plausibly infer missing associations in EMR data, in: Public Health and Informatics: Proceedings of MIE 2021, Vol. 281, IOS Press, 2021, p. 417.

[93] V. Murali, Y.P. Muralidhar, C. Königs, M. Nair, S. Madhu, P. Nedungadi, G. Srinivasa, P. Athri, Predicting clinical trial outcomes using drug bioactivities through graph database integration and machine learning, Chem. Biol. Drug Des. 100 (2) (2022) 169–184.

[94] Y. Lan, S. He, K. Liu, X. Zeng, S. Liu, J. Zhao, Path-based knowledge reasoning with textual semantic information for medical knowledge graph completion, BMC Med. Inform. Decis. Mak. 21 (2021) 1–12.

[95] J. Jiang, T. Wang, B. Wang, L. Ma, Y. Guan, Gated tree-based graph attention network (GTGAT) for medical knowledge graph reasoning, Artif. Intell. Med. 130 (2022) 102329.

[96] T. Wu, A. Khan, M. Yong, G. Qi, M. Wang, Efficiently embedding dynamic knowledge graphs, Knowl.-Based Syst. 250 (2022) 109124.

[97] A. Vretinaris, C. Lei, V. Efthymiou, X. Qin, F. Özcan, Medical entity disambiguation using graph neural networks, in: Proceedings of the 2021 International Conference on Management of Data, 2021, pp. 2310–2318.

[98] M.W. Chekol, G. Pirro, J. Schoenfisch, H. Stuckenschmidt, TeCoRe: Temporal conflict resolution in knowledge graphs, Proc. VLDB Endow. 10 (2017) Iss–12.

[99] S. Liu, L. Wang, V. Chaudhary, H. Liu, Attention neural model for temporal relation extraction, in: Proceedings of the 2nd Clinical Natural Language Processing Workshop, 2019, pp. 134–139.

[100] H. Advaith, S. Adarsh, G. Akshay, P. Sreeram, G. Sajeev, A proximity based community detection in temporal graphs, in: 2020 IEEE International Conference on Electronics, Computing and Communication Technologies, CONECCT, IEEE, 2020, pp. 1–6.

[101] Y. Wang, Y. Qiao, J. Ma, G. Hu, C. Zhang, A.K. Sangaiah, H. Zhang, K. Ren, A novel time constraint-based approach for knowledge graph conflict resolution, Appl. Sci. 9 (20) (2019) 4399.

[102] J. Ma, C. Zhou, Y. Chen, Y. Wang, G. Hu, Y. Qiao, TeCre: A novel temporal conflict resolution method based on temporal knowledge graph embedding, Information 14 (3) (2023) 155.

[103] I.Y. Chen, M. Agrawal, S. Horng, D. Sontag, Robustly extracting medical knowledge from EHRs: A case study of learning a health knowledge graph, in: Pacific Symposium on Biocomputing 2020, World Scientific, 2019, pp. 19–30.

[104] Y. Shang, Y. Tian, M. Zhou, T. Zhou, K. Lyu, Z. Wang, R. Xin, T. Liang, S. Zhu, J. Li, EHR-oriented knowledge graph system: Toward efficient utilization of non-used information buried in routine clinical practice, IEEE J. Biomed. Health Inf. 25 (7) (2021) 2463–2475.

[105] C. Hong, E. Rush, M. Liu, D. Zhou, J. Sun, A. Sonabend, V.M. Castro, P. Schubert, V.A. Panickan, T. Cai, et al., Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data, NPJ Digit. Med. 4 (1) (2021) 151.

[106] D. Zhou, Z. Gan, X. Shi, A. Patwari, E. Rush, C.-L. Bonzel, V.A. Panickan, C. Hong, Y.-L. Ho, T. Cai, et al., Multiview incomplete knowledge graph integration with application to cross-institutional EHR data harmonization, J. Biomed. Inform. 133 (2022) 104147.

[107] X. Liu, L.-Q. Xu, Knowledge graph building from real-world multisource "dirty" clinical electronic medical records for intelligent consultation applications, in: 2021 IEEE International Conference on Digital Health, ICDH, 2021, pp. 260–265, http://dx.doi.org/10.1109/ICDH52753.2021.00049.

[108] S. Kumar, A. Nanelia, R. Mariappan, A. Rajagopal, V. Rajan, et al., Patient representation learning from heterogeneous data sources and knowledge graphs using deep collective matrix factorization: Evaluation study, JMIR Med. Inform. 10 (1) (2022) e28842.

[109] S.G. Finlayson, P. LePendu, N.H. Shah, Building the graph of medicine from millions of clinical narratives, Sci. Data 1 (1) (2014) 1–9.

[110] A. Harnoune, M. Rhanoui, M. Mikram, S. Yousfi, Z. Elkaimbillah, B. El Asri, BERT based clinical knowledge extraction for biomedical knowledge graph construction and analysis, Comput. Methods Programs Biomed. Update 1 (2021) 100042.

[111] Y. Li, B. Qian, X. Zhang, H. Liu, Graph neural network-based diagnosis prediction, Big Data 8 (5) (2020) 379–390.

[112] Y. Si, Jingcheng Du, Zhao Li, Xiaoqian Jiang, Timothy Miller, Fei Wang, W. Jim Zheng, and Kirk Roberts. 2021."Deep representation learning of patient data from electronic health records (EHR): A systematic review", J. Biomed. Inform. 115, 103671.

[113] H. Jin, Y. Luo, P. Li, J. Mathew, A review of secure and privacy-preserving medical data sharing, IEEE Access 7 (2019) 61656–61669.

[114] G.A. Kaissis, M.R. Makowski, D. Rückert, R.F. Braren, Secure, privacy-preserving and federated machine learning in medical imaging, Nat. Mach. Intell. 2 (6) (2020) 305–311.

[115] R. Torkzadehmahani, R. Nasirigerdeh, D.B. Blumenthal, T. Kacprowski, M. List, J. Matschinske, J. Spaeth, N.K. Wenke, J. Baumbach, Privacy-preserving artificial intelligence techniques in biomedicine, Methods Inf. Med. 61 (S 01) (2022) e12–e27.

[116] J.-S. Lee, C.-J. Chew, J.-Y. Liu, Y.-C. Chen, K.-Y. Tsai, Medical blockchain: Data sharing and privacy preserving of EHR based on smart contract, J. Inform. Secur. Appl. 65 (2022) 103117.

[117] F. Gong, M. Wang, H. Wang, S. Wang, M. Liu, SMR: Medical knowledge graph embedding for safe medicine recommendation, Big Data Res. 23 (2021) 100174.

[118] S. Wang, M. Lin, T. Ghosal, Y. Ding, Y. Peng, Knowledge graph applications in medical imaging analysis: A scoping review, Health Data Sci. 2022 (2022).

[119] R. Li, F. Ma, J. Gao, Integrating multimodal electronic health records for diagnosis prediction, in: AMIA Annual Symposium Proceedings, Vol. 2021, American Medical Informatics Association, 2021, p. 726.

[120] D.M. Bean, H. Wu, E. Iqbal, O. Dzahini, Z.M. Ibrahim, M. Broadbent, R. Stewart, R.J. Dobson, Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records, Sci. Rep. 7 (1) (2017) 1–11.

[121] P. Nedungadi, R. Raman, The medical virtual patient simulator (medvps) platform, in: Intelligent Systems Technologies and Applications: Vol. 1, Springer, 2016, pp. 59–67.

[122] S. Ansong, K.F. Eteffa, C. Li, M. Sheng, Y. Zhang, C. Xing, How to empower disease diagnosis in a medical education system using knowledge graph, in: Web Information Systems and Applications: 16th International Conference, WISA 2019, Qingdao, China, September 20-22, 2019, Proceedings 16, Springer, 2019, pp. 518–523.

[123] B. Cope, M. Kalantzis, C. Zhai, A. Krussel, D. Searsmith, D. Ferguson, R. Tapping, Y. Berrocal, Maps of medical reason: Applying knowledge graphs and artificial intelligence in medical education and practice, in: Bioinformational Philosophy and Postdigital Knowledge Ecologies, Springer, 2022, pp. 133–159.

[124] R. Miotto, L. Li, B.A. Kidd, J.T. Dudley, Deep patient: An unsupervised representation to predict the future of patients from the electronic health records, Sci. Rep. 6 (1) (2016) 1–10.

[125] S. Dubois, N. Romano, D.C. Kale, N. Shah, K. Jung, Learning effective representations from clinical notes, Stat 1050 (2017) 15.

[126] Z. Zhu, C. Yin, B. Qian, Y. Cheng, J. Wei, F. Wang, Measuring patient similarities via a deep architecture with medical concept embedding, in: 2016 IEEE 16th International Conference on Data Mining, ICDM, IEEE, 2016, pp. 749–758.

[127] V.S. Bharadhwaj, M. Ali, C. Birkenbihl, S. Mubeen, J. Lehmann, M. Hofmann-Apitius, C.T. Hoyt, D. Domingo-Fernández, CLEP: A hybrid data-and knowledge-driven framework for generating patient representations, Bioinformatics 37 (19) (2021) 3311–3318.

[128] K. Soman, C.A. Nelson, G. Cerono, S.E. Baranzini, Time-aware embeddings of clinical data using a knowledge graph, in: Pacific Symposium on Biocomputing 2023: Kohala Coast, Hawaii, USA, 3–7 January 2023, World Scientific, 2022, pp. 97–108.

[129] Z. Zhang, F. Zhuang, M. Qu, F. Lin, Q. He, Knowledge graph embedding with hierarchical relation structure, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 3198–3207.

[130] P. Chandak, K. Huang, M. Zitnik, Building a knowledge graph to enable precision medicine, Sci. Data 10 (1) (2023) 67.

[131] X. Yuan, Interactive machine comprehension with dynamic knowledge graphs, 2021, arXiv preprint arXiv:2109.00077.

[132] C. Wewer, F. Lemmerich, M. Cochez, Updating embeddings for dynamic knowledge graphs, 2021, arXiv preprint arXiv:2109.10896.