Clinical Therapeutics 46 (2024) 544–554

Contents lists available at ScienceDirect

Clinical Therapeutics

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

Review

Knowledge Graphs in Pharmacovigilance: A Scoping Review

Manfred Hauben, MD, MPH[a,b], Mazin Rafi, MSc[c,*], Ibrahim Abdelaziz, PhD[d], Oktie Hassanzadeh, PhD[d]

[a] Department of Family and Community Medicine, New York Medical College, Valhalla, New York
[b] Truliant Consulting, Baltimore, Maryland
[c] Department of Statistics, Rutgers University, Piscataway, New Jersey
[d] IBM Research - Yorktown Heights, Yorktown Heights, New York

ARTICLE INFO

Key words:
Adverse drug reactions
Drug safety
Graph machine learning
Knowledge graphs
Pharmacovigilance
Scoping review

ABSTRACT

Purpose: To critically assess the role and added value of knowledge graphs in pharmacovigilance, focusing on their ability to predict adverse drug reactions.
Methods: A systematic scoping review was conducted in which detailed information, including objectives, technology, data sources, methodology, and performance metrics, were extracted from a set of peer-reviewed publications reporting the use of knowledge graphs to support pharmacovigilance signal detection.
Findings: The review, which included 47 peer-reviewed articles, found knowledge graphs were utilized for detecting/predicting single-drug adverse reactions and drug-drug interactions, with variable reported performance and sparse comparisons to legacy methods.
Implications: Research to date suggests that knowledge graphs have the potential to augment predictive signal detection in pharmacovigilance, but further research using more reliable reference sets of adverse drug reactions and comparison with legacy pharmacovigilance methods are needed to more clearly define best practices and to establish their place in holistic pharmacovigilance systems.

Introduction

The principal objective of pharmacovigilance (PV) is the timely identification of novel adverse drug reactions. This could involve drugs in development (for which there is relatively limited information) and approved drugs (for which more data may exist.) Adverse drug reaction profiles are largely unknown, even by the time the drug is authorized for clinical use, and continuously unfold over time with the application of pharmacovigilance signal detection methods.

Safety signal detection and prediction are challenging due to the interwoven complexity of a drug's pharmacological actions, underlying comorbid illnesses, and increasing rates of polypharmacy. Formulating adverse drug reactions as occurrences within complex biological systems brings knowledge graphs (KG) and graph machine learning into focus in pharmacovigilance. KGs are arguably inspired by scientific determinism. Expressed in a pharmacovigilance context, the occurrence of a previously unknown adverse drug reaction (ADR) is a predictable outcome of a set of potentially specifiable biological conditions. Given the interwoven complexity and emergent properties of biological systems in health and disease, the full ensemble of necessary conditions may be elusive. But in the era of big data, biological KGs hold the potential to compile a sufficiently complete ensemble to improve prediction of adverse drug reactions.

Despite the abundance of data sets vying for resources in pharmacovigilance (PV), the specific added value of drug safety knowledge graphs still needs to be determined, especially when compared to the plethora of other methodologies available. To address this ambiguity, we conducted a systematic scoping review. This review aims to evaluate the performance and potential of these graphs and their incremental utility and to identify areas requiring further exploration. Our goal is to systematically map the research landscape in this domain and pinpoint any gaps in knowledge or communication. We recommend reading the companion article in this issue titled "Knowledge Graphs in Pharmacovigilance: A Step-By-Step Guide," as preparation for reading this scoping review.

Methodology

Definitions

A presentation by Ehrlinger and Wöß highlighted the variability in existing definitions of knowledge graphs (SEMANTICS 2016: Posters and

* Address correspondence to: Mazin Rafi, MSc, Department of Statistics, Rutgers University, Piscataway, NJ, USA.
E-mail address: mmr257@scarletmail.rutgers.edu (M. Rafi).

https://doi.org/10.1016/j.clinthera.2024.06.003
Accepted 5 June 2024

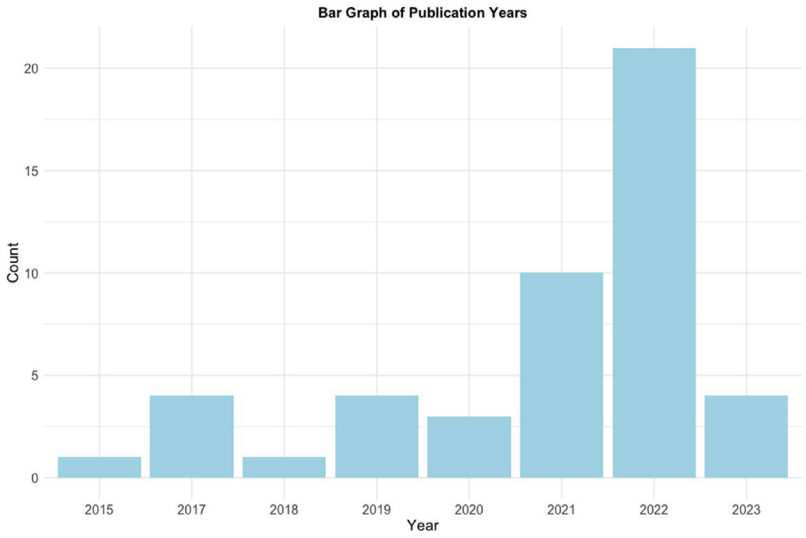

Figure 1. Frequency of included knowledge graph publication by year through March 2023.

Demos Track September 13 to 14, 2016, Leipzig, Germany). Our working definition of a KG is a heterogeneous knowledge base consisting of triples of object pairs and connecting relationships modeled through graphs and ontologies (a standardized semantic framework for representing all objects and relationships in a domain), which extract new insights from existing data sets via their integration.

Rationale

Literature Search and Data Extraction

We searched the Web of Science through March 2nd, 2023, using a structured query (Supplement 1). Titles, abstracts, and/or the full text were reviewed as necessary to adjudicate relevance. We supplemented the corpus with additional references based on personal knowledge or incidentally discovered during our review. While preprints may provide a more comprehensive picture of ongoing research activity, they were excluded due to their lack of rigorous peer review.

We extracted the following information: (1) Pharmacovigilance task (adverse drug reaction prediction including "cold start," signal detection, signal refinement, or pharmacovigilance in the clinic). (2) Whether the knowledge graph was used to predict adverse reactions with a single drug (ADRs), drug-drug interactions (DDIs) or both (3) Whether the study involved de novo KG construction from component data sources or a prefabricated KG or linked open data source (LOD). (4) Data sources, including if federated in a LOD. (5) The numbers and types of KG entities and relationships in the KGs. (6) Embedding methods. (7) Machine learning methods applied to the embedded KGs. (8) Best reported performance metrics compared to non-network similarity methods or traditional pharmacovigilance approaches. (9) Journals in which the articles were published. (10) Geographic location of the lead authors/affiliations. (11) Funding source.

A publicly available GitHub page provides a table of a subset of the columns on this sheet and a list of all the citations.

Results

General

Of the 416 articles identified with our Web of Science search, 42 met our inclusion criteria.[1–42] An additional five relevant articles were incidentally discovered through a bibliographic review of the original 42 included articles, resulting in a total of 47 included articles.[1–47] The other 374 articles identified through our Web of Science search were excluded for the following reasons: They did not involve the safety of medical products (e.g., non-medicinal chemical safety), did not involve use of KGs as we have defined them (e.g. involved homogeneous networks), involved only drug-target prediction or drug repurposing, were editorial or review articles (despite exclusion of those type of articles in our search strategy), or a corrigendum.[48–56]

Publications Over Time

The year of publication of the included articles was left-skewed with a generally increasing trend over time (Figure 1).

Journals

The 47 included articles were published in 28 different journals. Figure 2 displays the number of included articles by journal. A substantial majority of the articles included were published in bioinformatics and medical/health informatics journals, with none published in drug safety, pharmacovigilance, or pharmacoepidemiology journals.

Figure 3 displays the impact factors and citation numbers by year of publication. The number of citations per article, as measured by the median, generally declined over time. Notable is the outlier with 891 citations.[39] This paper reported the construction of a knowledge graph consisting of drug-drug and drug-protein interactions for polypharmacy side effects. One other paper[44] had over 100 citations (221), which involved drug-drug interaction prediction. The journals were notable for moderately high impact factors (median IF being just under 5) without a clear trend over time.

Funding Sources and First Author Location

The source of funding of the included articles was academia in 44 papers[1-8,10-37,39,40,42-47] and software industry in 3 papers.[9,38,41] The first authors of the included articles represented 17 different countries (Figure 4). The Republic of China was by far the most frequent (24).

Information Technology

Most papers did not report the underlying data model used for the KG. Among the six that did, five reported using resource description framework (RDF),[3,22,27,42,43] while one reported labeling the property graph as the schema.[11] Most studies did not specify whether/which GDBMS or graph store was used. When specified, Apache Spark, neo4j, graphDB, Virtuoso, SDM-RDFizer, and DeTrusty were used.[3,8,27,41] Most studies did not specify or did not use a graph

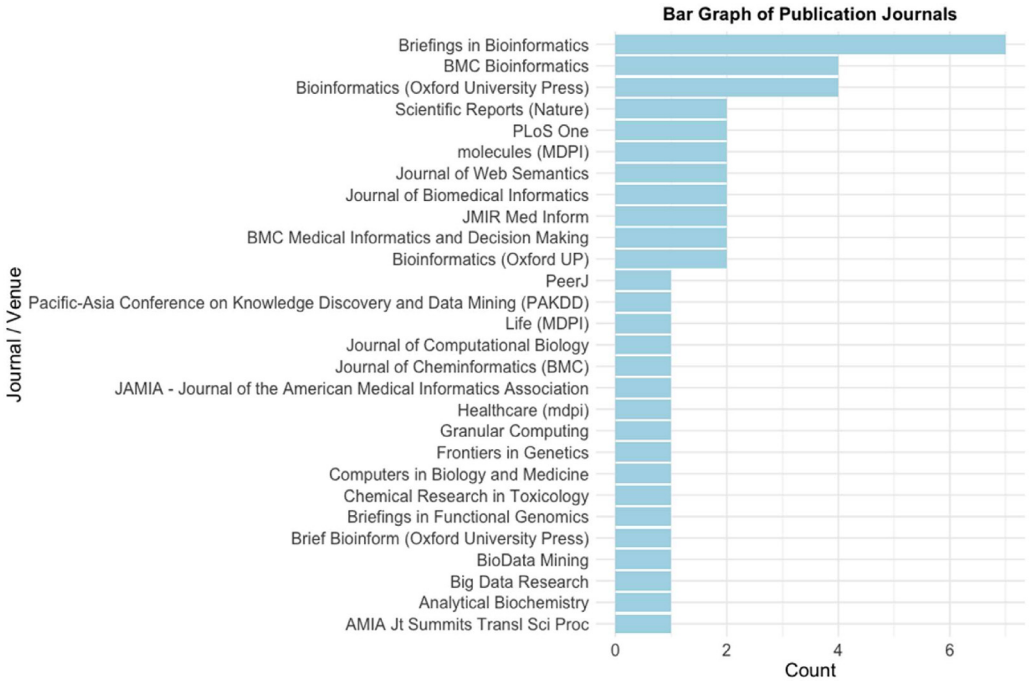

Figure 2. Journals publishing included KG publications.

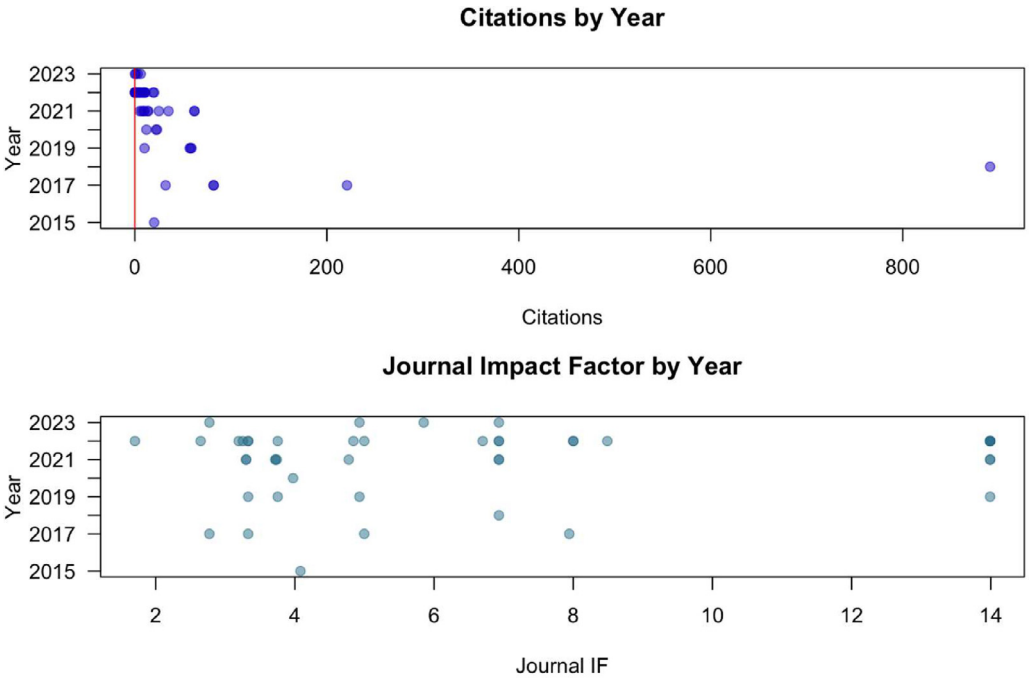

Figure 3. Number of citations of each included publication and impact factors of journals.

query language. Of the six papers that report a graph query language, five used SPARQL,[3,22,27,42,43] and one used Cypher.[8] Twenty-seven of the included articles provided open-source code with a GitHub link.[2-6,9,12-16,19-22,25,26,29,30,34,36-40,44,45] No studies provided details of their entity resolution/node deduplication protocol.

AE Prediction Types

All but 3 papers reported the use of KG for classical PV activities of cold or warm start adverse event prediction or signal detection or refinement. Three papers reported the use of KGs for clinical PV, specifically for safe medicine predictions in the clinic.

Twenty-seven studies involved only adverse DDIs[1,2,4-7,11,12,14-18,21-23,26,27,29,31,34,36,37,41-44], 16 involved only simple ADRs[8,10,13,19,20,25,28,30,32,33,35,38,40,45-47], and four involved both adverse DDIs and simple ADRs[3,9,24,39]. All but one paper involved adverse drug reactions/interactions. One involved adverse vaccine reactions, specifically with COVID-19 vaccine.[8] The stated objective of three studies included using KGs to provide mechanistic insights into predicted ADRs or DDIs (signal refinement use task)[28,33,43] The latter objective corresponds to

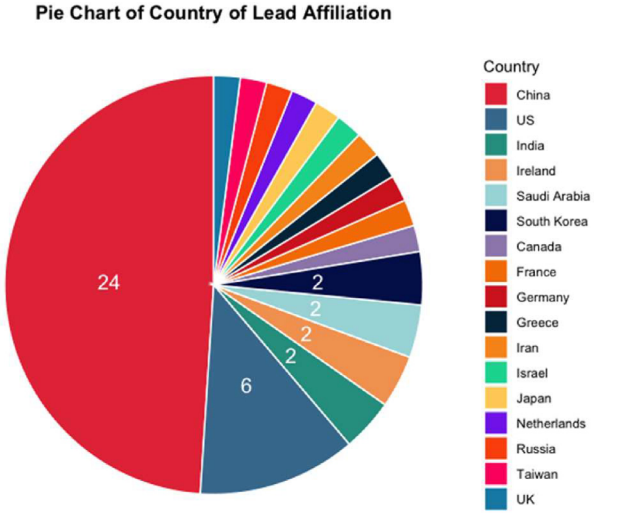

Figure 4. Geographic location of first authors of included KG publications.

the pharmacovigilance use case of signal refinement, in that the mechanistic insights could be used to strengthen/weaken a signal detected from other sources.[48]

Six papers studied circumscribed subsets of drugs or ADRs: COVID-19 drugs,[3] vancomycin, amlodipine, cisplatin, and glimepiride,[4] COVID-19 vaccine,[8] a set of adverse events from the Understanding Adverse Drug Reactions and the Observational Medical Outcomes Partnership (acute kidney insufficiency, acute myocardial infarction, anaphylaxis, aplastic anemia, acute liver failure, erythema multiforme, gastrointestinal hemorrhage, mitral valve disease, gastrointestinal hemorrhage, neutropenia, and rhabdomyolysis),[25] drug-induced liver injury (DILI) and severe cutaneous adverse drug reactions (SCARs)[28] and oncology drugs.[33]

Eight studies involved cold start prediction.[7,9,16,19,21,24,31,36] The notion of "cold start" was used in two slightly different ways across the included papers: Predictions with only limited drug information because the drug is in early stages of development versus a drug in any stage of development that is new to the KG. The latter usage is also known as "inductive learning." Five cold start papers[7,16,26,31,36] involved drug-drug interaction (DDI) prediction, and three[9,19,24] involved serious adverse drug reaction (sADR) prediction, with two studies focusing on safe medicines recommendations for patients, particularly in the cold start scenario of new patients.[19,24]

Approaches to cold start predictions included: (1) Embedding based on extensive chemical structure data/ embedding dimensions because significant chemical information should be available even for drugs in early development. This information included circular fingerprints, vectors of global molecular features, embedding from pre-trained hyperbolic VAEs, weighted chemical similarity edges, and molecular graphs.[9] (2) Using additional properties known for both old and new drugs, such as carriers, transporters, enzymes, and targets (CTET.)[21] This would be expected to be limited to the second notion of a cold start above. (3) Implement a train, validate, and test split in which so-called "weak nodes" with limited connections to the network are enriched in the test set[9]. (4) Removing known drugs from the KG and treating them as new drugs.

Data Sources

We identified 54 data sources that were used across the 47 included articles. Some report a single source (a KG), while others rely on several sources that are either integrated or used separately. Seventeen[2,3,6,8,10,13,14,22,26,32,33,35,39-41,46,47], studies constructed their own KG while thirty[1,4,5,7,9,11,12,15-21,23-25,27-31,34,36-38,42-45] used a pre-existing open source heterogeneous KG.

Overall, DrugBank was the most frequently used data source in 31 of the 47 (66%) studies. It was the most frequent source for both entities[1-3,4,7,11,13,14,16,18,20-22,23,26,27,29,32,35,36,37,40,41,43] and auxiliary information[10,19,24,26] The frequent use of DrugBank relates to the variety of its integrated information sources. SIDER was the second most frequent, used in 17 of 47 studies, followed by KEGG, used in 14 studies. Spontaneous reporting system data was a source of entities in multiple studies. Entities in various studies drew from the spontaneous reporting system data, either directly utilizing the FDA's Adverse Event Reporting System (FAERS),[42] the Vaccine Adverse Event Reporting System (VAERS),[8] or an AERS-derived (via machine learning) dataset such as OFFSIDES[34,39,44,46] and TWOSIDES.[20,26,29,34,39,44] OFFSIDES is a data set that claims to provide scientifically supported side effects that were not listed in FDA-approved labeling at the time of entry.[49] TWOSIDES is a data set of drug pair-side effect combinations that the creators of the resource claim are scientifically supported drug-drug interactions.[49] MIMIC-III was an electronic health record system used in two studies.[24,31] Figure 5 displays all the data sources and their frequency of use.

Reflecting on this trend, it is noteworthy that most of the papers introducing new models do not actually construct their own knowledge graphs but rather leverage available open-access knowledge graphs to extract relevant data and features for their studies. The new "models" often involve a reworking of a prefabricated KG or a different method/approach to a current dataset. For instance, LaGAT demonstrates a new method instead of building a new KG, and SimVec KG, initially perceived as a novel KG, heavily draws inspiration from the TriVec and Decagon papers. This observation leads to a reconsideration of what constitutes a "new KG" – whether it is one modified from a pre-existing one, or one that is entirely original, such as 3DGT-DDI.

KG Characteristics

Publications were inconsistent in providing the size of the KGs-i.e., clearly specifying the number of nodes, edges, and additional feature labels. In the reports providing relevant information, the number of drugs reported to be included in the KG ranged from 873 to 63,485,[28] with a majority reporting between 1000-12000 drugs (Figure 6). The outlier of 63,485 from a study using PGxLOD[28] seems to exceed the expected number of unique drugs substantially. This reflects the fact that a given drug can currently be represented multiple times in the PGxLOD, so the number of unique drugs is substantially less. (Coulet A, Monnin P. Personal communication, 11/24/23).

The reported number of side effects ranged from 86[20] to 4,600,000[34] with a median of 5589. However, the larger numbers, such as 4,600,000, often represent non-unique side effect counts (e.g., a given adverse event associated with 1000 different drug-drug combinations might be counted 1000 times).

Among reports providing relevant information, the median number of triples (facts) in the KG was 5,874,261, ranging from 7,146[27] to 96,000,000[25] (Figure 7).

The median number of types of entities per KG was 4, with a range of 2-67. As shown in Figure 8, the number of entities tended to be somewhat higher for KGs used for DDI prediction than sADRs. This was due to the inclusion of additional entities, such as CTET, that are directly relevant to DDIs but not necessarily to ADRs involving a single drug ADRs. The outlier with 67 entity types presented a data ecosystem to predict sADRs and adverse DDIs with COVID-19 drug treatments, using well over half a million publications from three data sources: PubMed, PMC, and CORD-19.[3] Two other studies involved 12 and 13 entities, respectively.[7,18] The study with 12 entities used DrugBank and Hetionet to extract 12 entities (anatomy, cellular components, molecular function, biological processes, biological pathways, side effects) to predict DDIs and sADRs of drugs used to treat COVID-19. The study with 13 entities predicted DDIs using a pre-existing open-source biomedical knowledge graph.[18]

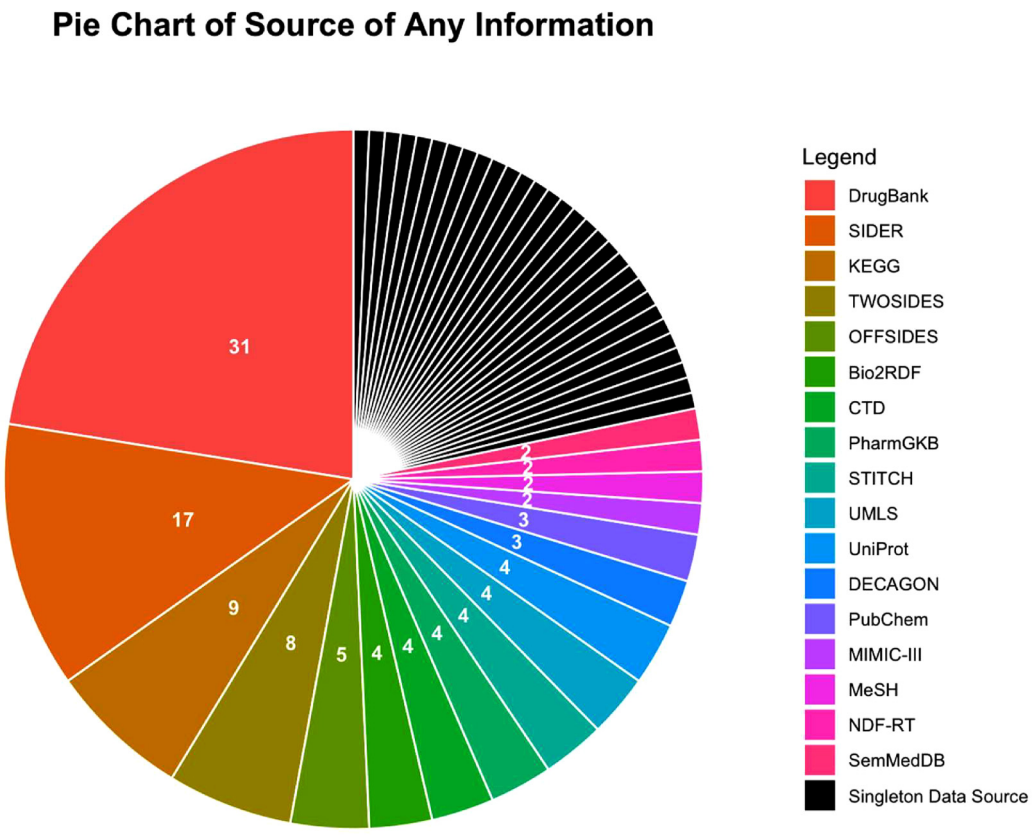

Figure 5. Data sources used for KGs from the included papers.

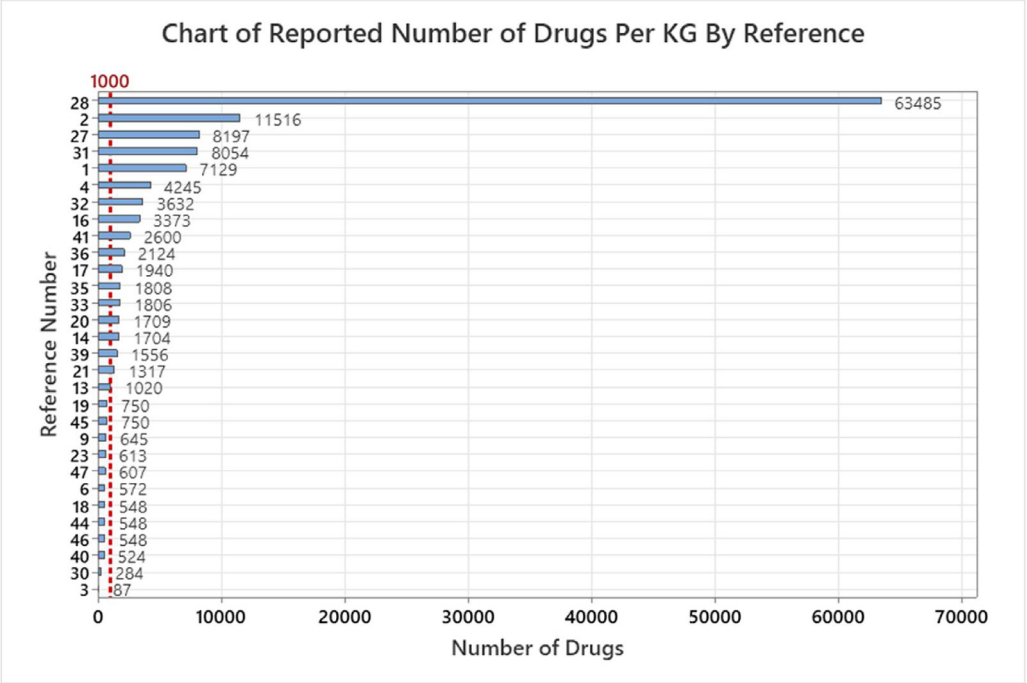

Figure 6. Number of drugs in KGs from the included papers providing this information.

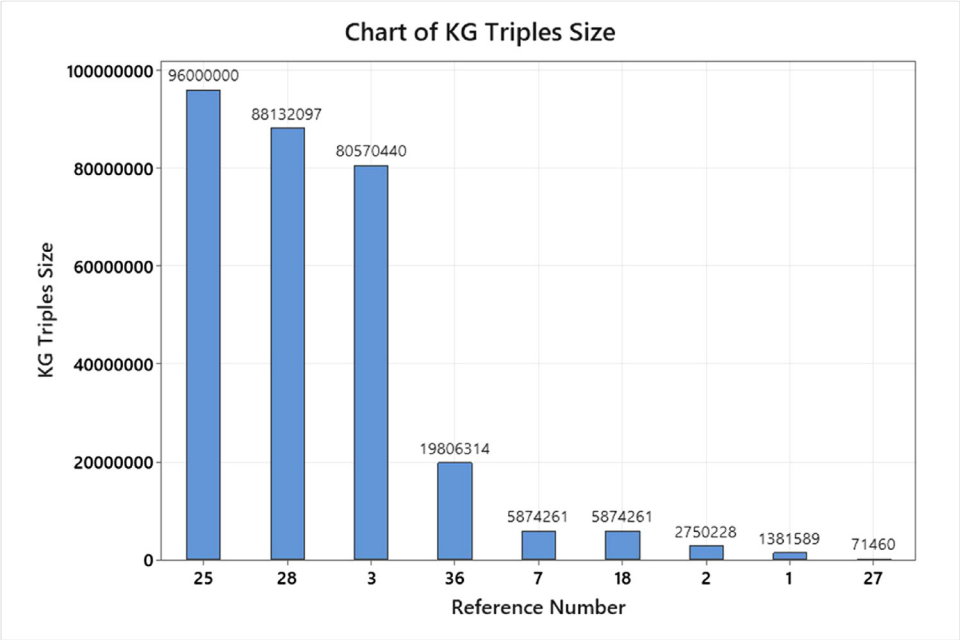

Figure 7. Number of triples in KGs from the included papers providing relevant information.

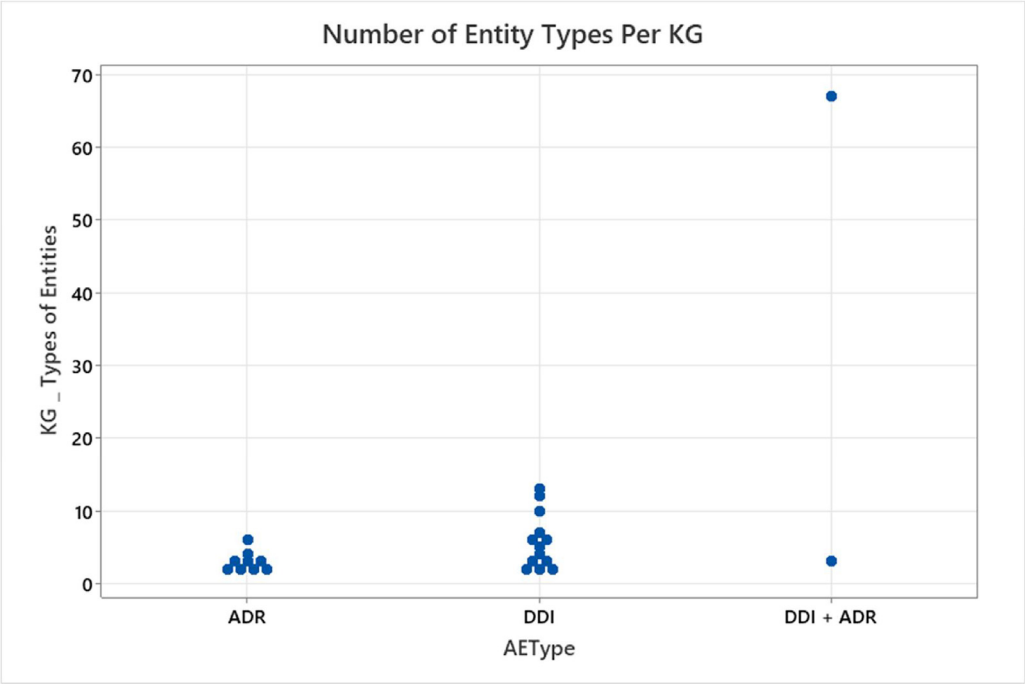

Figure 8. Number of types of entity in KGs in included papers providing relevant information.

Three articles reported KGs containing individual patients as nodes.[8,24,31] One used structured and unstructured data in VAERS data to create a KG network with individual patient nodes, vaccines, medical history (UMLS), and symptoms for querying, similarity analysis, and prediction of hospitalization with COVID-19 vaccine sADRs.[8] In the remaining two papers, electronic medical records, specifically MIMIC-III, were used to create a heterogeneous medical knowledge graph of patients, diseases (ICD-9) and drugs (DrugBank) in order to make safe medicines recommendations.[24,31] MIMIC-III is an extensive, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital containing vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, and more. In one excluded study, MIMIC-III was used in conjunction with two other graphs: Decagon and CANCER. Decagon is a knowledge graph model for modeling polypharmacy side-effects by constructing a multimodal graph that includes adverse drug-drug interactions, as alleged in TwoSides, a database of DDIs constructed from a subset of spontaneous adverse event reports in the FAERS database. Meanwhile, the CANCER dataset EHRs provide comprehensive patient history and treatment outcomes, which, when integrated with knowledge graphs like Decagon, can offer advanced predictive analytics for patient care management. This holistic approach facilitates informed decision-making in clinical settings, aiming to enhance the safety and effectiveness of treatments. The investigators integrated an electronic record of health data with a drug-drug

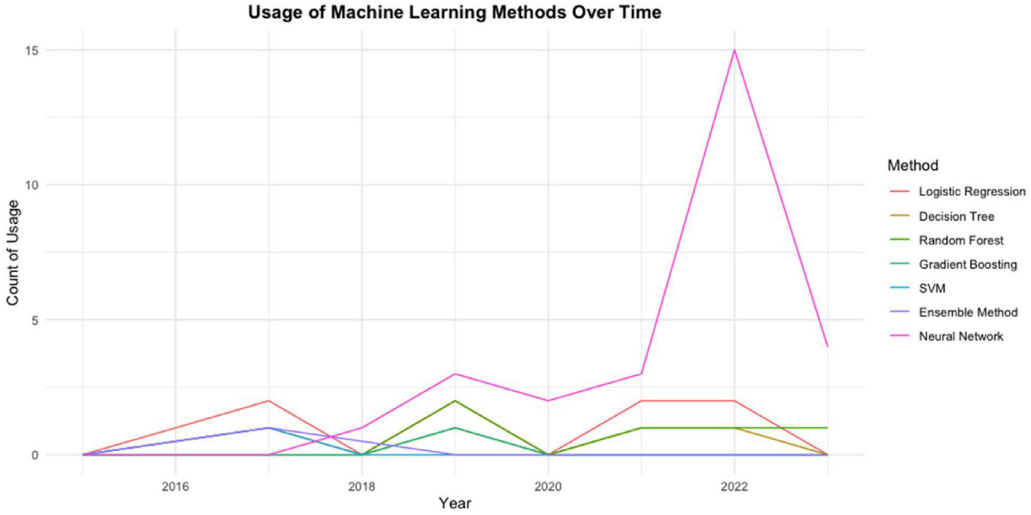

Figure 9. This figure shows Machine Learning Algorithm trends. Neural networks seem to be used more over time.

interaction knowledge graph to make safe and effective treatment recommendations. The integration of patient health care data from MIMIC-III and the drug-drug interaction insights from Decagon and EHRs, was employed to make safe and effective treatment recommendations; this encompasses both safety and efficacy objectives by predicting adverse and synergistic/therapeutic interactions.[56]

Embedding Methods

A majority of studies (89.6%) reported embedding methods. The embedding methods were variable: neural network-based (13), translational/rotational-based (11), semantic matching models (7), random walk-based (7), word embedding (6), and tensor factorization (2). The neural network architectures were predominantly Graph Neural Networks (GNNs), including variations like Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), and Graph Convolutional Networks with Multi-Kernel (GCNMK). The random walk-based methods were primarily node2vec and Sim2Vec. Twelve papers reported using multiple embedding methods: TransE, DistMult, ComplEx,[2] TransE, DistMult, HolE, RESCAL,[3] word embedding, GNN,[14] Mol2Context-vec, ComplEx, DURA,[17] TransE, DeepWalk,[25] ComplEx, RotateE, SimplE,[26] TransE, TransR,[27] GFAN, GAT,[30] TransR, LINE,[31] TransE, TransD, RDF2Vec,[36] TransH, HolE,[41] and skip-gram, GNT.[47] Multiple embedding methods were used for comparative performance assessment or because different types of relationships were embedded, some of which are not effectively embedded by specific methods.

Machine Learning Algorithms

Random forests and logistic regression were the most frequently utilized Machine Learning Algorithms (MLAs) (42 each), followed by neural networks and decision trees (41 each), support vector machines and gradient boosting (40 each), and ensemble methods (30). While the overall frequency (summed over time) favors more traditional MLA, such as random forests and logistic regression over time, neural networks have been increasingly used and are currently the most common MLA (Figure 9).

A variety of neural network architectures were used including MLP multi-layer perceptron,[4,5,12,18,36,37] graph convolutional networks,[1,4,12,16,19,23,47] neural factorization machines,[1] bidirectional encoder representations from transformers,[2,14] and graph attention networks.[5,19,30,45,47] Comparison methods of MLAs often included a comparison to other baseline Neural Network classifiers such as GNNs.

Ground Truth

The benchmark adverse drug event data sets used as ground truth were related to product labels contained in the component data sets such as DrugBank or SIDER or spontaneous reporting systems (FAERS, VAERS) or derived from spontaneous reporting systems (e.g., TWOSIDES)-derived data. Only one included article used evidence-based reference sets of drug-event associations that have been adjudicated in a structured manner: Exploring and Understanding Adverse Drug Reactions (EU-ADR) and the Observational Medical Outcomes Partnership (OMOP). This particular article was also distinctive in that it built a KG based on published medical literature rather than the usual bioinformatics data sources.[25]

Performance Metrics

Cross-validation (CV) was reported in 30 papers. CV was 5-fold in sixteen papers,[2,6,7,8,13,14,25,30,46] 10-fold in thirteen papers,[1,3,4,13,19,21,27,28,35,36,41,42,45] and 3-fold in one paper.[33] The following metrics were reported in at least ten studies (in descending frequency order): AUC (38), AUPR (30), F-1 score (26), precision (24), recall (18), and accuracy (13). Several studies used some form of "@k" or "at n" performance metrics such as recall@n,[19] precision@n,[19] NDCG@n,[19,24] AP@N or MAP@N.[20,24,25] None of the studies reported negative predictive value. Additional metrics reported included the Matthews correlation coefficient.

Figure 10 presents a dot plot visualization that portrays the individual values of various performance metrics as reported by the 47 studies by index. The figure includes an assortment of ranking metrics such as hits@k and precision@n. However, it is important to note that certain metrics were infrequently utilized or their reporting was ambiguous—specifically, some were depicted solely through bar graphs without precise value labels.

Figure 11 displays median performance measures by adverse event type.

The biomedical adjudication of results was only sometimes transparently presented, such as when results were presented stratified by adverse events.[30,34] One paper involving DDI prediction reported the area under the precision-recall curve (AUPRC) stratified by side effect for the best (AUCPR>C.0.9) and worst (AUCPRC<0.712) performance side effects.[34] The best were mumps, carbuncle, coccydynia, tympanic perforation, dyshidrosis, spondylosis, schizoaffective disorder, breast dysplasia, ganglion, and uterine polyp. The worst were bleeding, increased body temperature, emesis, renal disorder, leucopenia, diarrhea, icterus,

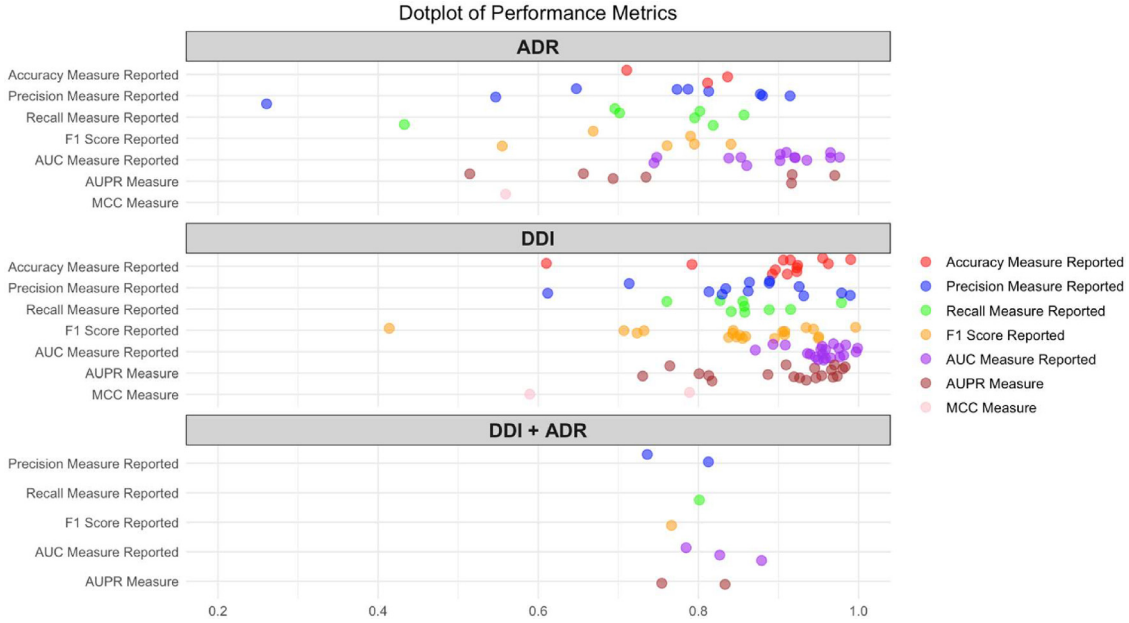

Figure 10. Reported performance tended to be higher for KGs used for DDI prediction compared to sADR prediction.

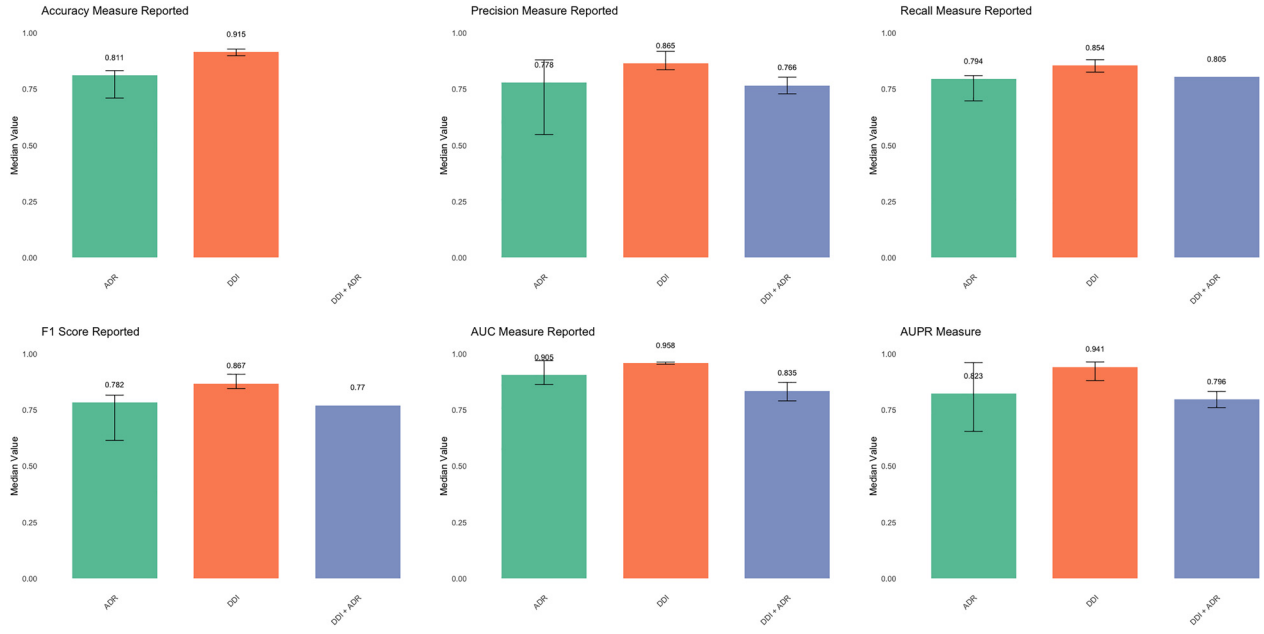

Figure 11. Median of Selected performance measure by adverse event type: single drug ADR versus drug-drug Interaction. Bars are 95% confidence intervals of the median obtained via bootstrapping with 1000 samples. Performance for DDI is higher for each measure.

nausea, itch, and anemia. The authors claim that "manual examination of the results and discussion with domain experts.." reveals that the best-performing side effects have strong "apparent molecular underpinnings" and that "side effects with the worst performance tend to be common side effects and/or have non-molecular origins with potentially important environmental and behavioral components." without precisely defining what "molecular underpinning" and "non-molecular origin" are, or how they arrived at these assessments. Another publication provided a table of ten DDI-related adverse events with the highest AUCs (>0.84) along with the number of drug airs related to these adverse events[30]: pregnancy-induced hypertension (50), high-risk pregnancy (40), macroglossia (91), renovascular hypertension (5), familial adenomatous polyposis (23), gastric stasis (36), vascular headache (7), dacryocystitis (10), arterial insufficiency (99) and pneumocystis pneumonia (114).

Given the numerous analytical choices in each study, it would be interesting to perform a multivariate regression to better understand the impact of these choices, but more information would be needed, including a larger sample of articles, and more complete information in each article, for this to be statistically meaningful. With appropriate caution, we can observe some general correlations or trends between KG characteristics and reported performance (Table 1).

Notably, none of the studies compared signal detection performance with legacy methods routinely used in pharmacovigilance, such as disproportionality analysis. Disproportionality analysis is a widely deployed signal detection technique that is typically used to screen large

Table 1
Univariate observations related to omnibus performance measures (AUC, F1-score).

Factor	Observations
ADR Type	KGs for DDIs tended to return higher performance.
Cold versus Warm start	Cold start scenarios are rare and associated with lower performance.
Number of data sources	Most studies integrated data from multiple sources, such as DrugBank, KEGG, HIPPIE, CTD, UniProt, and others, which contributed to improved predictive performance.
Use of prefabricated KG	The use of prefabricated KGs often resulted in better performance.
Number of features	Performance improved with more features.
Number of entities	Studies generally reported lower predictive performance when dealing with datasets containing more than six distinct entity types.
Number of triples	While detailed numbers for triple sizes were not consistently reported, larger KG triple sizes, when available, tended to be associated with better predictive performance.
Embedding dimensions	Detailed information about the embedding dimensions used in the studies was typically not reported.
Embedding method	Lower performance reported with word2vec, node2vec, SimVec. Higher performance with translational embeddings; highest performance with GNN and neural factorization machine.
Neighborhood aggregation function	No specific trend observed with the choice of neighborhood aggregation function.
Post-embedding machine learning method	Studies that employed post-embedding machine learning methods, particularly complex techniques such as deep neural networks, generally achieved higher predictive performance.
Type of ground truth	Studies using integrated ground truth sources to validate DDI predictions generally reported improved performance.

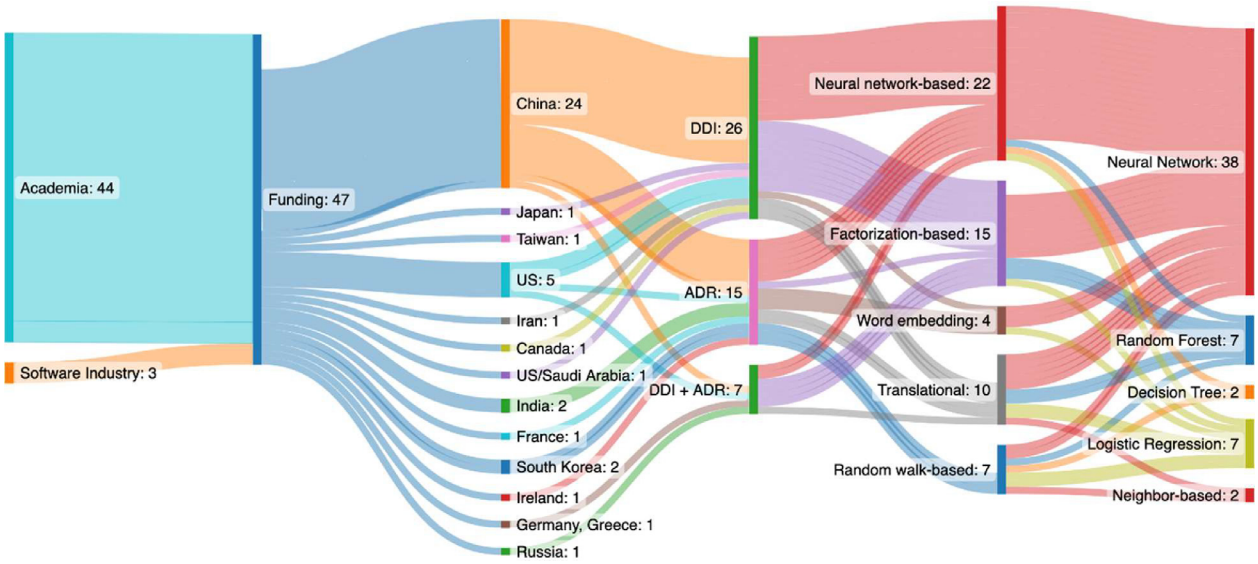

Figure 12. Sankey diagram of key features of the included studies.

spontaneous reporting databases by calculating observed-to-expected reporting frequencies.[57] An adverse event reported significantly above expected frequencies, a so-called signal of disproportionate reporting (SDR), is a common starting point for signal detection. When a comparison with some form of a legacy analysis was performed, it was a legacy methodology employing similar biological and chemical knowledge bases in a machine learning application. One study assessed the detection capabilities by assessing predicted new DDIs from an earlier version of DrugBank.

Figure 12 provides an aerial view of the included studies.

Discussion

We extracted a robust body of research on applying KGs in PV. This involved sADRS, DDIS, and safe medicine recommendations for individual patients. Despite the proliferation of such brilliant and elegant research, our scoping review identified gaps in the corpus of selected papers that might hamper a fuller realization of the potential of KGs in this area. These include: (1) A lack of publications in the drug safety/pharmacovigilance/pharmacoepidemiology literature. This is especially concerning because it may deprive the research of the requisite clinical and scientific judgment necessary to adjudicate study results, with such adjudications being questionable in some instances, as described above under performance metrics. We submit that some clinical adjudications of results would have failed to pass peer review in a pharmacovigilance, drug safety, or pharmacoepidemiology journal. (2) Paucity of comparisons of KG-based approaches to PV signal detection and refinement using legacy methods such as disproportionality analysis of spontaneous reports, with a corresponding discussion of how warm start KG-based signal detection pipelines might be practically implemented in a pharmacovigilance system. (3) Under-representation of ground truth reference stronger than product labeling. (4) Minimal work on vaccines. (5) Opaque writing.

We identified univariate trends in performance as a function of individual design choices, and we stress that every study involves numerous choices. These choices include data sources, number of types of entities, node and edge dimensions, number of parameters, GNN layers, embedding dimensions, BFS/DFS and biases in random-walk based methods, number of information types that are aggregated, neighbor aggregation metric, and thus any observed univariate trend can be heavily confounded.

The increasing use of graph neural network-based approaches may be due to the rapid growth in research in the areas of Graph Representation Learning and Graph Neural Networks in general by a large community of researchers affiliated with Chinese academic institutions (as indicated by a recently published book in Chinese.) We see the majority of recent work in the area of drug safety also from the same affiliations.

An obvious limitation in any scoping review is the capture of relevant articles. In our case, we conducted a rigorous WOS search to

identify peer-reviewed publications describing KG used for PV tasks. We excluded all articles on preprint servers, conference abstracts and posters, and peer-reviewed publications that report just the construction of knowledge graphs. Some exclusions involved papers describing KGs suited for PV tasks but not tested for those purposes. A final limitation is that additional publications may have appeared during the delay between submission and publication of our paper that may have altered our results and conclusions had they been included.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: At the time of manuscript submission Manfred Hauben was a full time employee of Pfizer Inc., New York, New York, USA; he has also won stocks in pharmaceutical companies that may manufacture or market drugs mentioned in the article. However, he reports no other conflicts of interest that are directly relevant to the content of this manuscript.

Mazin Rafi, a former Summer Associate of Pfizer Incorporated, is currently pursuing an MSc in Data Science through Rutgers University, New Brunswick, USA. He has no other conflicts of interest that are relevant to the content of this manuscript.

Ibrahim Abdelaziz is affiliated with IBM Research - Yorktown Heights, Yorktown Heights, NY, USA. He reports no conflicts of interest that are relevant to the content of this manuscript.

Oktie Hassanzadeh is also affiliated with IBM Research - Yorktown Heights, Yorktown Heights, NY, USA. He reports no conflicts of interest that are relevant to the content of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.clinthera.2024.06.003.

References

1. Zhang J, Chen M, Liu J, et al. A knowledge-graph-based multimodal deep learning framework for identifying drug–drug interactions. *Molecules*. 2023;28:1490. doi:10.3390/molecules28031490.
2. Asada M, Miwa M, Sasaki Y. Integrating heterogeneous knowledge graphs into drug–drug interaction extraction from the literature. *Bioinformatics*. 2023;39:btac754. doi:10.1093/bioinformatics/btac754.
3. Sakor A, Jozashoori S, Niazmand E, et al. Knowledge4COVID-19: a semantic-based approach for constructing a COVID-19 related knowledge graph from various sources and analyzing treatments' toxicities. *J Web Semantics*. 2023;75:100760. doi:10.1016/j.websem.2022.100760.
4. Chen YH, Shih YT, Chien CS, Tsai CS. Predicting adverse drug effects: a heterogeneous graph convolution network with a multi-layer perceptron approach. *PLOS ONE*. 2022;17:e0266435. doi:10.1371/journal.pone.0266435.
5. Hong Y, Luo P, Jin S, Liu X. LaGAT: link-aware graph attention network for drug–drug interaction prediction. *Bioinformatics*. 2022;38:5406–5412. doi:10.1093/bioinformatics/btac682.
6. Al-Rabeah MH, Lakizadeh A. Prediction of drug-drug interaction events using graph neural networks based feature extraction. *Sci Rep*. 2022;12:15590. doi:10.1038/s41598-022-19999-4.
7. Ren ZH, You ZH, Yu CQ, et al. A biomedical knowledge graph-based method for drug–drug interactions prediction through combining local and global features with deep neural networks. *Briefings Bioinform*. 2022;23:bbac363. doi:10.1093/bib/bbac363.
8. Liu Z, Gao X, Li C. Modeling COVID-19 vaccine adverse effects with a visualized knowledge graph database. *Healthcare*. 2022;10:1419. doi:10.3390/healthcare10081419.
9. Lukashina N, Kartysheva E, Spjuth O, Virko E, Shpilman A. SimVec: predicting polypharmacy side effects for new drugs. *J Cheminform*. 2022;14:49. doi:10.1186/s13321-022-00632-5.
10. Joshi P, V M, Mukherjee A. A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network. *J Biomed Inform*. 2022;132:104122. doi:10.1016/j.jbi.2022.104122.
11. Chen M, Jiang W, Pan Y, Dai J, Lei Y, Ji C. SGFNNs: signed graph filtering-based neural networks for predicting drug-drug interactions. *J Comput Biol*. 2022;29:1104–1116. doi:10.1089/cmb.2022.0113.
12. He C, Liu Y, Li H, et al. Multi-type feature fusion based on graph neural network for drug-drug interaction prediction. *BMC Bioinform*. 2022;23:224. doi:10.1186/s12859-022-04763-2.
13. Yu L, Cheng M, Qiu W, Xiao X, Lin W. idse-HE: Hybrid embedding graph neural network for drug side effects prediction. *J Biomed Inform*. 2022;131:104098. doi:10.1016/j.jbi.2022.104098.
14. He H, Chen G, Yu-Chian Chen C. 3DGT-DDI: 3D graph and text based neural network for drug–drug interaction prediction. *Briefings Bioinform*. 2022;23:bbac134. doi:10.1093/bib/bbac134.
15. Su X, Hu L, You Z, Hu P, Zhao B. Attention-based knowledge graph representation learning for predicting drug-drug interactions. *Briefings Bioinform*. 2022;23:bbac140. doi:10.1093/bib/bbac140.
16. Feng YH, Zhang SW, Zhang QQ, Zhang CH, Shi JY. deepMDDI: A deep graph convolutional network framework for multi-label prediction of drug-drug interactions. *Analyt Biochem*. 2022;646:114631. doi:10.1016/j.ab.2022.114631.
17. Ren ZH, Yu CQ, Li LP, et al. BioDKG–DDI: predicting drug–drug interactions based on drug knowledge graph fusing biochemical information. *Briefings Functional Genomics*. 2022;21:216–229. doi:10.1093/bfgp/elac004.
18. Hao X, Chen Q, Pan H, et al. Enhancing drug–drug interaction prediction by three-way decision and knowledge graph embedding. *Granul Comput*. 2023;8:67–76. doi:10.1007/s41066-022-00315-4.
19. Xu X, Yue L, Li B, et al. DSGAT: predicting frequencies of drug side effects by graph attention networks. *Briefings Bioinform*. 2022;23:bbab586. doi:10.1093/bib/bbab586.
20. Yao J, Sun W, Jian Z, Wu Q, Wang X. Effective knowledge graph embeddings based on multidirectional semantics relations for polypharmacy side effects prediction. *Bioinformatics*. 2022;38:2315–2322. doi:10.1093/bioinformatics/btac094.
21. Liu Z, Wang XN, Yu H, Shi JY, Dong WM. Predict multi-type drug–drug interactions in cold start scenario. *BMC Bioinformatics*. 2022;23:75. doi:10.1186/s12859-022-04610-4.
22. Han X, Xie R, Li X, Li J. SmileGNN: drug–drug interaction prediction based on the SMILES and Graph Neural Network. *Life*. 2022;12:319. doi:10.3390/life12020319.
23. Wang F, Lei X, Liao B, Wu FX. Predicting drug–drug interactions by graph convolutional network with multi-kernel. *Briefings Bioinform*. 2022;23:bbab511. doi:10.1093/bib/bbab511.
24. Wang N, Cai X, Yang L, Mei X. Safe medicine recommendation via star interactive enhanced-based transformer model. *Comp Biol Med*. 2022;141:105159. doi:10.1016/j.compbiomed.2021.105159.
25. Dasgupta S, Jayagopal A, Hong ALJ, Mariappan R, Rajan V. Adverse drug event prediction using noisy literature-derived knowledge graphs: algorithm development and validation. *JMIR Med Inform*. 2021;9:e32730. doi:10.2196/32730.
26. Dai Y, Guo C, Guo W, Eickhoff C. Drug–drug interaction prediction with Wasserstein Adversarial Autoencoder-based knowledge graph embeddings. *Briefings Bioinform*. 2021;22:bbaa256. doi:10.1093/bib/bbaa256.
27. Wang M, Wang H, Liu X, Ma X, Wang B. Drug-drug interaction predictions via knowledge graph and text embedding: instrument validation study. *JMIR Med Inform*. 2021;9:e28277. doi:10.2196/28277.
28. Bresso E, Monnin P, Bousquet C, et al. Investigating ADR mechanisms with explainable AI: a feasibility study with knowledge graph mining. *BMC Med Inform Decision Making*. 2021;21:171. doi:10.1186/s12911-021-01518-6.
29. Yu Y, Huang K, Zhang C, Glass LM, Sun J, Xiao C. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics*. 2021;37:2988–2995. doi:10.1093/bioinformatics/btab207.
30. Bang S, Jhee JH, Shin H. Polypharmacy side-effect prediction with enhanced interpretability based on graph feature attention network. *Bioinformatics*. 2021;37:2955–2962. doi:10.1093/bioinformatics/btab174.
31. Gong F, Wang M, Wang H, Wang S, Liu M. SMR: medical knowledge graph embedding for safe medicine recommendation. *Big Data Res*. 2021;23:100174. doi:10.1016/j.bdr.2020.100174.
32. Zhang F, Sun B, Diao X, Zhao W, Shu T. Prediction of adverse drug reactions based on knowledge graph embedding. *BMC Medical Inform Decision Making*. 2021;21:38. doi:10.1186/s12911-021-01402-3.
33. Wang M, Ma X, Si J, et al. Adverse drug reaction discovery using a tumor-biomarker knowledge graph. *Front Genetics*. 2021;11:1–11. doi:10.3389/fgene.2020.625659.
34. Nováček V, Mohamed SK. Predicting polypharmacy side-effects using knowledge graph embeddings. *AMIA Jt Summits Transl Sci Proc*. 2020;2020:449–458. Accessed May 12, 2023. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233093/ .
35. Xue R, Liao J, Shao X, et al. Prediction of adverse drug reactions by combining biomedical tripartite network and graph representation model. *Chem Res Toxicol*. 2020;33:202–210. doi:10.1021/acs.chemrestox.9b00238.
36. Celebi R, Uyar H, Yasar E, Gumus O, Dikenelli O, Dumontier M. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. *BMC Bioinform*. 2019;20:726. doi:10.1186/s12859-019-3284-5.
37. Shtar G, Rokach L, Shapira B. Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures. *PLOS ONE*. 2019;14:e0219796. doi:10.1371/journal.pone.0219796.
38. Muñoz E, Nováček V, Vandenbussche PY. Facilitating prediction of adverse drug reactions by using knowledge graphs and multi-label learning models. *Briefings Bioinform*. 2019;20:190–202. doi:10.1093/bib/bbx099.
39. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34:i457–i466. doi:10.1093/bioinformatics/bty294.
40. Bean DM, Wu H, Iqbal E, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Sci Rep*. 2017;7:16416. doi:10.1038/s41598-017-16674-x.
41. Abdelaziz I, Fokoue A, Hassanzadeh O, Zhang P, Sadoghi M. Large-scale structural and textual similarity-based mining of knowledge graph to predict drug–drug interactions. *J Web Seman*. 2017;44:104–117. doi:10.1016/j.websem.2017.06.002.

42. Jiang G, Liu H, Solbrig HR, Chute CG. Mining severe drug-drug interaction adverse events using Semantic Web technologies: a case study. *BioData Min*. 2015;8:12. doi:10.1186/s13040-015-0044-6.
43. Noor A, Assiri A, Ayvaz S, Clark C, Dumontier M. Drug-drug interaction discovery and demystification using Semantic Web technologies. *J Am Med Inform Assoc*. 2017;24:556–564. doi:10.1093/jamia/ocw128.
44. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinform*. 2017;18:18. doi:10.1186/s12859-016-1415-9.
45. Zhao H, Zheng K, Li Y, Wang J. A novel graph attention model for predicting frequencies of drug–side effects from multi-view data. *Briefings Bioinform*. 2021;22:bbab239. doi:10.1093/bib/bbab239.
46. Hu B, Wang H, Yu Z. Drug side-effect prediction via random walk on the signed heterogeneous drug network. *Molecules*. 2019;24:3668. doi:10.3390/molecules24203668.
47. Kwak H, Lee M, Yoon S, Chang J, Park S, Jung K. Drug-disease graph: predicting adverse drug reaction signals via graph neural network with clinical data. *Adv Knowledge Discov Data Mining*. 2020;12085:633–644. doi:10.1007/978-3-030-47436-2_48.
48. Ji S, Gao Y, Marttinen P. Knowledge-augmented graph neural networks with concept-aware attention for adverse drug event detection. *Computation and Language*. 2023;2301:1–13. doi:10.48550/arXiv.2301.10451.
49. Galeano D, Li S, Gerstein M, Paccanaro A. Predicting the frequencies of drug side effects. *Nat Commun*. 2020;11:4575. doi:10.1038/s41467-020-18305-y.
50. Ye Q, Hsieh CY, Yang Z, et al. A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. *Nat Commun*. 2021;12:6775. doi:10.1038/s41467-021-27137-3.
51. Xu X, Meng F, Sun L. Knowledge mining of interactions between drugs from the extensive literature with a novel graph-convolutional-network-based method. *Electronics*. 2023;12:311. doi:10.3390/electronics12020311.
52. Li J, Yang X, Guan Y, Pan Z. Prediction of drug-target interaction using dual-network integrated logistic matrix factorization and knowledge graph embedding. *Molecules*. 2022;27:5131. doi:10.3390/molecules27165131.
53. Alshahrani M, Almansour A, Alkhaldi A, et al. Combining biomedical knowledge graphs and text to improve predictions for drug-target interactions and drug-indications. *Peer J*. 2022;10:e13061. doi:10.7717/peerj.13061.
54. Huang L, Fernandes H, Zia H, et al. The cancer precision medicine knowledge base for structured clinical-grade mutations and interpretations. *J Am Med Inform Assoc*. 2017;24:513–519. doi:10.1093/jamia/ocw148.
55. Shen C, Li Z, Chu Y, Zhao Z. GAR: Graph adversarial representation for adverse drug event detection on Twitter. *Appl Soft Computing*. 2021;106:107324. doi:10.1016/j.asoc.2021.107324.
56. Symeonidis P, Chairistanidis S, Zanker M. Safe, effective and explainable drug recommendation based on medical data integration. *User Modeling and User-Adapted Interaction*. 2022;32:999–1018. doi:10.1007/s11257-022-09342-x.
57. Montastruc JL, Sommet A, Bagheri H, et al. Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *Br J Clin Pharmacol*. 2011;72:905–908].